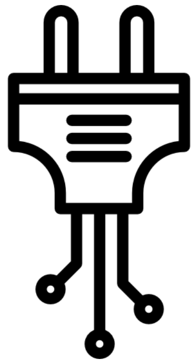

Smart Home Energy Efficiency

Application of machine learning on smart meter power data to improve efficient electric energy usage in households

Milo Boers
Delft University of Technology

A thesis submitted in partial fulfilment for the degree of Master in Sustainable Energy Technology. Commissioned by Toon, part of Eneco.

July, 2022

Master Thesis

Smart Home Energy Efficiency

Application of machine learning on smart meter power data to improve efficient electric energy usage in households

Delft University of Technology

Faculty of Electrical Engineering,
Mathematics and Computer Science
Master Sustainable Energy Technology
Mekelweg 4
2628 CD Delft
The Netherlands

Eneco Group

Toon Energy (formerly Quby)
Department of Smart Energy
Joan Muyskenweg 22
1096 CJ Amsterdam
The Netherlands

Milo Boers
Student number 4007565
July 4, 2022

First assessor:
Prof.dr. Kornelis Blok (Delft University of Technology)

Second assessor:
Dr. Linda Kamp (Delft University of Technology)

Third assessor:
Prof.dr. Laure Itard (Delft University of Technology)

Internship mentor:
Dr. Kaustav Basu (Quby)

Internship supervisor:
Dr. Stephen Galsworthy (Quby)

Foreword

This thesis is the result of my graduation project for the Master in Sustainable Energy Technology, at Delft University of Technology. In partial fulfilment of this project a nine months research internship was conducted at smart energy company Toon (formerly Quby) and part of the Dutch utility company Eneco Group.

This paper is an elaboration on my research about the utilisation of smart meter power data to improve efficient electric energy usage in households. In my eyes the application of machine learning in the energy sector has much untapped potential to monitor and recommend and potentially even manage energy demand in a more efficient manner.

Several parts of the developed methodology and findings have been used for the development of the Toon Waste Checker, a patented technology that advises customers on their energy usage and supports them in reducing energy inefficiency. [1] [2] As the team has further developed the technology, the results of this study do not reflect the technology used within the Waste Checker to date.

Acknowledgements
This thesis marks the end of a long and cumbersome journey that accompanied me for the last 5 years. Having put the project almost fully on hold end of 2017 and throughout 2018 and starting my first full-time job shortly after, trust in myself to finish this thesis was on an all time low. A nagging voice in my head and the amazing support I received kept me going while working on the report on numerous vacation days and endless weekends and finally I decided to take the opportunity between two jobs for a last push. I am incredible grateful to finally having had the chance to bring this project to closure and would like to deeply thank all my supporters who have helped me though this time.

First I would like to thank my supervisor Kornelis Blok for the opportunity to conduct my thesis research under his guidance. Thank you for bringing me in contact with Quby to conduct my research internship there. Thank you for all the help with ideating about potential research pathways and all the feedback that I received from you. Most of all, thank you for remaining my supervisor for all these years, for all the support, kindness and understanding you have given me. And thank you for being an incredibly inspiring and knowledgeable force for good within the energy transition.

I would also like to express my great gratitude towards Linda Kamp, who joined last year to guide me with the planning and process of my thesis finalization and helped step by step improving my storyline. This has been of great importance for me and helped me make my decision to take a break from work to finalise this research.

Furthermore, I would like to thank Laure Itard. Laure joined the thesis as third supervisor and her feedback has been a great angle for improvement of the theory and methodology on smart monitoring of household energy usage.

Next to my academic supporters, I would like to thank my internship supporters. I would like to thank Tako in 't Veld, head of Quby Smart Energy and Stephen Galsworthy, who gave me the opportunity to conduct this research as my direct manager.

I would also very much like to thank Kaustav Basu, who mentored me during my journey of learning how to code within a data analytics environment. It was his guidance that helped me further develop this skill that has been of great use for me eversince.

In addition, I would like to thank professor Stamminger from the Universität Bonn, who has been a great source of knowledge. His collaboration has given substantial direction to the analysis and potential literature to consider.

Furthermore, I would like to thank several of the co-workers at Quby, who I had the privilege to work with and learn from, including Tudor Toma, Telmo Oliveira, Ernie Durdevic and Bert Jan Katsman.

Not to forget all the great support I have received from family and friends. In particular a special thank you to my parents Michel and Marisca and my two younger brothers Maurits and Mathis. Their love and support helped me restart the project and their encouragement has been a strong driver for me.

Besides many of my other dear family members, I would like to give a special thank you to my aunt Ellen, who has been an advisor and support throughout my study time, was and still is a great guidance for me throughout my life.

I would also like to thank several of my close friends. In particular Kai, Ruben, Cees, Simon, Mirko and Stefan who always found the understanding and words of encouragement or a good joke. I would also like to thank Lara. I am grateful for all her help throughout the initial research phase, the great years we spent together and the good friends we remained.

Last but not least, I would like to give a special thank you to my partner Maike. Thank you for sticking through this with me. You have been an invaluable support. I will always be thankful for striking the balance between patience and determination and I have learned and grown so much from being together with you. I am really sorry for all those holidays and weekend activities we have missed out together in the last years, but I promise for more to come.

Abstract

Nonintrusive Load Monitoring (NILM) can be used to disaggregate household energy usage collected from a central meter to the level of individual appliances, but has so far mostly been applied in controlled, small-scale settings. Further potential such as the application for energy efficiency classification and management have remained largely untapped. In this research a machine learning model was developed and deployed to determine energy consumption, usage pattern and energy efficiency characteristics of real-life dishwasher usage based on smart meter data. The developed NILM system was deployed on a full year of smart meter data for nearly 130.000 households in the Netherlands. The analysis was accompanied by a survey to gain additional information on the households usage behaviour. The average energy consumption for all households was found to be 1.18kWh per wash, with dishwashers used 240 times per year on average. Dependencies are shown for household size, washing temperature, machine efficiency label as well as time of day, week and year. It was estimated that 9 in 10 households could reduce their dishwasher energy consumption, with an average saving potential of more than 30% per year. The developed method showed to be suitable to gain insight into average electricity consumption and usage patterns on an appliance level, non-intrusively and at large-scale. Additional survey data was shown to provide comparative insight between different user groups. The developed framework can be easily expanded for other major appliances and could be used to drive tailored consumer feedback on energy efficiency improvements within households.

Keywords: Smart Meter Load Monitoring; Load Disaggregation; Pattern Recognition; Energy Analysis; Energy Efficiency

Summary

Decreasing household energy demand by optimising the efficiency of its use would reduce energy needed. Providing insight into consumers energy usage has been shown to support households in reducing their energy consumption. Non-Intrusive Load Monitoring (NILM) could be used as scalable and continuous approach for energy assessments disaggregating electricity data without much intervention of the user.

Most NILM research so far has focused on the detection of devices in one or only several (controlled) households. No research to date has connected energy consumption patterns and efficiency assessments in a real world, consumer facing context and analysed the results at large scale. Aim of this research is to develop a NILM system based on smart meter data and to use this system to determine energy consumption, usage pattern and efficiency characteristics of real-life appliance usage in households on a large scale. Due to the many analytical steps involved in setting up this system and analysing results, it was deemed best to focus on one appliance in depth. Dishwashers are a recognisable, widely used appliance, with relatively high energy consumption and variability in real-life usage between households. Therefore, this study has been focused on dishwashers as a use case.

To gain insight into the characteristics of dishwasher usage in real life an analysis was carried out for a 3 months period for 100 households with smart plugs connected to the dishwasher and accompanied with a survey. The smart plug data was used as reference data to develop and validate a machine learning based NILM model, extending a NILM appliance disaggregation methodology developed by K. Basu (2017). The model was extended by the estimation of energy consumption per wash, weekly usage frequency and efficiency classification. The developed algorithms were applied on smart meter data gathered for a full year for more than 130.000 households to investigate energy consumption, usage pattern and efficiency characteristics non-intrusively at large scale. A survey was conducted under nearly 11.000 users to contextualise the results.

The application of the developed method in large-scale smart meter data results in an average energy consumption of 1.18kWh per wash equalling an A label machine and the dishwasher is used 240 times per year, 40 times less than assumed for the EU efficiency label. Dependencies were shown for household size, washing temperature, machine efficiency label as well as time of day, week and year.

Energy per wash and frequency of usage show a seasonal dependency with peaks in winter and lows in summer. Energy consumption per wash changes in accordance with outside temperature, deviating by 0.23 kWh between the maximum and the minimum

throughout the year on average. A relation between the frequency of dishwasher usage and the occurrence of events and holidays can be drawn. Within the weekly pattern, dishwasher usage differs per weekday with least usage on Friday. Main usage was identified to be directly after dinner time and just before people go to bed.

The energy consumption of dishwashers depends on household and machine characteristics. Number of people within the household appeared to be one of the main factors for weekly usage with 2.8 weekly washes for single households, 4.4 for two people household and increasing by 0.6 for each additional household member. The temperature of the chosen washing program impacts the energy consumption more than the efficiency label. High efficiency label and low temperature result in an average energy consumption of 1.13 kWh (for A+++, <30°C). Low efficiency label with high temperature (label A and below, <75°C) result in an average energy consumption of 1.35 kWh. However, it has to be noted that the reduced accuracy of the model further away from the average likely causes an underestimation for these differences.

The plug data was used to assess how well the developed model was able to estimate energy consumption and usage frequency from the smart meter data. The average estimation error (RMSE) was found to be 0.10kWh for energy per wash and 1.4 days/week for the usage frequency. This translates into a normalized estimation error (NRMSE) of 8.8% for energy per wash and 27.2% for the usage frequency. The average value can be estimated with much higher accuracy as a result of the sample size of 100 households for the training set, resulting in a ten-fold lower error of the estimated mean.

For the efficiency classification a high classification accuracy (F1) of 91% for energy per wash and 89% for weekly usage was found. The analysis of the frequency and energy efficiency showed that 84% of households consume more energy per wash than needed, 62% use their machine more often than needed. It was estimated that 9 in 10 households could reduce their dishwasher energy consumption, with an average energy saving potential of more than 30% per year or 94 kWh.

This research shows that a NILM system, based on data collected from the smart meter, can deliver insights into the real-life usage of dishwashers without much intervention of the user. Due to the large-scale applicability, aggregated results are more generalisable than smart plug research and provide opportunity to analyse consumption patterns for different user groups, beyond what has been possible with surveys. However, additional survey information does provide the opportunity to gain deeper insight into individual household characteristics and can aid to form a complete picture.

The findings of this research show considerable untapped potential for residential energy saving. The developed framework could be expanded for other major appliances and used to drive tailored feedback on energy efficiency improvements within households.

Contents

1

Introduction

1.1 Problem Outline

1.1.1 Motivation

The residential sector forms a major sink of energy and source of greenhouse gas (GHG) emissions [3][4]. The IPCC identified the residential sector as a high potential sector for climate change mitigation measures, with a large share of cost-effective opportunities for GHG reduction [5]. Decreasing household energy demand by optimising the efficiency of use would henceforth reduce how much energy has to be generated in the first place.

In the past decade, energy efficiency has become one of the major priorities in EU policies addressing issues that are related to energy [6][7][8]. Environmental awareness and legislation, have driven technical optimisation of household appliances towards higher energy efficiency. While these changes have significantly reduced the energy demand, consumer behaviour shifting toward a more environmentally conscious use of household appliances would have high potential for further energy conservation gains [9].

Despite a growing awareness of efficiency, research has shown that a discrepancy between potential and actual adoption of energy efficiency measures occurs, contributing to a so called 'Energy Efficiency Gap' [10]. Consumers seem to have difficulty achieving energy saving behaviour by intention alone [11] and they are hardly aware of their real-life energy consumption [12]. Research in behavioural science suggests, that the amount of energy consumption is strongly related to socio-demographic variables, while attitudinal variables more impact changes in energy consumption [13][14]. It therefore appears that good understanding of consumers' socio-demographic and behavioural variables is vital to offer expedient guidance to improve their energy consumption [15].Consumer engagement and knowledge provision have shown to be effective measures to achieve energy conserving behaviour [16] in particular when information and feedback is tailored to individual consumer level [17].

Another potential component for energy efficiency in households is automated building control [18]. This can be used to guide consumers and policy makers in decision-making about energy conservation measures and support the advancement of flexible grid operation [19]. The increasing share of weather dependent sustainable energy sources in the electricity generation mix creates higher grid volatility, making grid operators increasingly interested in optimising the communication of local power distribution networks [20]. Henceforth, methods are needed to design and integrate sensors and actuator networks that can be maintained at large scale, enabling grid operators to access holistic data-sets on consumption, usage patterns [21] and to integrate feedback mechanisms that convince customers to make their houses more grid-friendly. [22]

1.1.2 Knowledge gap

Research on household energy usage and efficiency relies on both top-down and bottom up approaches. Traditionally top-down studies gather data on generic level e.g. nation wide economic analysis, using averages and assumptions for end use estimation [23, 24, 25, 26, 27]. Bottom-up research traditionally focuses primarily on retrieving data on detailed level from few households e.g. by testing of individual appliances under standardised lab setting or by placing sensors in research households [28, 29, 30, 31, 32]. In many cases both approaches are accompanied by surveys [30, 24, 26] or survey results are used as assessment basis [33].

Consumer surveys can provide insight on a broad range of usage variables, including behavioural aspects and can be applied to large and varied groups. However, care must be taken that the sample of respondents is representative. Answers can furthermore depend on the way the questions are asked and respondents may possibly be biased or unaware of their actual preferences and behaviour [34].

When testing under controlled lab settings the energy consumption for different type of programs can be measured very accurately and in a repeatable manner. It is used among others for the EU labelling to compare different models that come to market under the same conditions. However, it does not test for real life usage and has to make assumptions on factors such as average annual washes. [35]

Placing sensors in research households, can provide insight in the interplay between multiple devices during actual usage [36]. Studies however often are constrained in sample size due to high cost and invasiveness of placing multiple sensors across the house. Furthermore, due to the invasiveness of sensor placement and constant measurement, the research may suffer selection bias (people in the household are researchers) and/or participants may (unconsciously) behave differently because they are aware of being monitored (Hawthorne effect). [37]

Non-Intrusive Load Monitoring (NILM), an interdisciplinary research field between computer science and electrical engineering, focuses on disaggregation of electricity data to appliance or even component level, based on data from the central meter and without much intervention of the user [38]. With the introduction of smart meters in many households around the globe, the usage of NILM-based energy usage disaggregation has been touted as a potential holy grail to inform efficiency behaviour by some scholars [39]. NILM has picked up over recent years and shows promising results to deploy in households [40], with increasingly higher accuracy, using increasingly sophisticated methods [41] [42] [41].

Despite the aim of application in the form of Home Energy Management Systems

(HEMS), most NILM research focuses solely on the detection of devices [43]. Research available in the field of energy efficiency, on the other hand, has so far been focused on in-depth analysis of efficient usage on small-scale, considering only one or a small set of machine, ambient or behavioural characteristics for only one or few type(s) of household appliance(s). [30, 31, 24].

To provide consumers, policymakers and grid operators with actionable insights for their respective purpose, NILM-based appliance detection has to be combined with detailed real-life analysis and information from the field of efficiency research.

To date little research is available that combines these two research fields [44]. Only recently publications start connecting NILM-detected appliances to efficiency labels [45], anomalous behaviour [46] or formulation of household efficiency indicators and feedback to consumers [47]. However, no research has yet applied NILM-based household appliance detection to analyse energy consumption, usage pattern and efficiency characteristics of household appliances in a real-life, consumer facing context at large scale.

To close this research gap a methodology needs to be developed, which allows for a combination of these components. To investigate the potential of this linkage in more detail, an in-depth analysis of a household appliance with a large data set is needed. Dishwashers are a recognisable, widely used appliance, with relatively high energy consumption and variability in real-life usage between households and are categorised as high potential appliances to adapt usage and implement efficiency measures [24]. Dishwashers could henceforth serve as an expedient case-study appliance to show the application of NILM in energy analysis research and utilisation of load disaggregation of smart meter data in real life.

1.2 Research Objective

1.2.1 Contribution

As described in the knowledge gap, no research has yet applied NILM-based household appliance detection for energy and efficiency analysis of real-life usage. Aim of this research is to utilise a smart-meter based NILM system and to further develop this to non-intrusively determine energy consumption, usage pattern and efficiency characteristics of appliance usage in households on a large scale.

In order to do this, load disaggregation research is combined with conventional energy efficiency research for the use-case of dishwashers. The first part of the research extends an existing smart-meter based NILM detection system with energy and frequency of usage estimation as well as efficiency classification and analyses its performance against a benchmark group of households. For the second part of the research the developed system is then deployed to analyse a full year of real-life household load profiles at large scale and compare energy usage, usage frequency and efficiency for dependencies on time, household and machine characteristics.

1.2.2 Research question

From this objective the following research question and sub-questions were derived:

How can energy consumption, usage pattern and efficiency characteristics of real-life dishwasher usage in households be detected and analysed with a smart meter based NILM system?

This is answered by the following five sub-questions:

1. How is electric energy usage and efficiency for dishwashers characterised based on traditional analysis?
2. How can a smart meter based NILM system be developed to detect this?
3. How does the newly developed system perform in detecting these characteristics?
4. How much energy do dishwashers in households consume and how often are they used, depending on time, household and machine characteristics?
5. How efficiently are dishwashers used in households?

1.3 Report Structure

The research outlined in this report deals both with the development of a model to detect energy consumption, usage patterns and efficiency of dishwashers from the smart meter signal and the analysis of applying this methodology at large scale. The report follows the typical literature, methodology, results, discussion and conclusion structure. The first part for each chapter focuses on the model and the second part on its application for the energy and efficiency analysis.

The report starts with a literature review of load disaggregation, followed by more detail on dishwasher energy consumption and efficiency. The methodology section describes the developed model and how this was applied to analyse dishwasher usage in households. The results first show the performance of the developed model and then the analysis resulting from the deployment of the model at large scale. Finally discussion and conclusion deal with the interpretation of results and the validity of the model performance and deployment and how these can be used and developed further.

Each chapter starts with a brief introduction to what will be presented in that chapter and how this fits within answering the main research question. At the end of each chapter a preliminary conclusion is provided, summarising the most important findings from that chapter and how this potentially already answers (part) of the research question.

2

Literature Review

2.1 Literature introduction

The literature chapter answers the first and partially answers the second research sub-questions by looking at what information is already available on Non-Intrusive Load Monitoring (NILM) at one hand and on dishwashers energy consumption and efficiency at the other hand. While the literature focuses on what has already been developed, the methodology section will further answer question 2 by explaining the methodology that was added to that. To get a better understanding of how a NILM system could be developed to detect dishwasher usage from the smart meter signal, 2.1 starts with describing household load patterns and developments and application in NILM research. In order to better understand how to detect and analyse results for dishwashers, what is already known about energy consumption characteristics of dishwashers is discussed in more detail in 2.2. After it is established how NILM could be used to detect appliances and how energy consumption of dishwashers is characterised and can be analysed, 2.3 progresses with how this could translate to energy efficiency, by analysing what is known about appliance efficiency.

In part answers the following sub-questions:

1. How is electric energy usage and efficiency for dishwashers characterised based on traditional analysis?
2. How can a smart meter based NILM system be developed to detect this?

2.2 Load Disaggregation

To get a better understanding of how a Non-Intrusive Load Monitoring (NILM) system can be developed to detect dishwasher usage from the smart meter signal, it is important to get a better understanding of household load profiles and developments in NILM research. The section starts with explanation of the household load profile, followed by how NILM works conceptually, comparing different approaches in NILM system application and finally a deeper dive into the chosen NILM application.

2.2.1 Household load profiles

The load profile of a household is an aggregation of the power consumption of all active household appliances at each point in time. While 'ECN Energietrends' [26] looks at the total energy consumption of the household, research such as by van Holsteijn and Kemna [23] and 'Energie Studie Centrum' [48] aim to describe the appliance level energy consumption and usage patterns constituting to the overall electric energy demand of the household.

Electric energy consumption in households is comprised of the interplay of different fixed and variable factors. While household size and energy efficiency of electricity consuming appliances can be regarded as fixed factors, consideration of behaviour account for variable factors [49]. These behavioural factors can for instance relate to demographics, presence profile in the house, time dependant shifts of appliance use and the way consumers use and interact with the technology [50].

Generally, the power demand of a household fluctuates depending on the type and amount of appliances used in a specific time. Intraday fluctuation shows higher energy usage during daytime than at night. Typically an evening peak occurs due to intensified lighting and cooking activities. In addition, changes between different weekdays can be observed particularly between workdays (Monday through Friday) and weekend days (Saturday and Sunday). Over the year seasonal differences cause changes in the energy consumption due to differences in outside temperature, hours of solar light and behaviour between seasons. Total energy use in winter is higher than in the summer months. Autumn and Spring show comparable patterns. [51]

An often cited research from Loughborough University looked into the household energy demand patterns based on plug measurements for individual appliances over the course of 2 years, to show how different appliances affect the total energy demand of the household over time. [52] While a unique and insightful data set for studies on energy demand modelling, the high cost, invasiveness of placing smart plugs and static nature of such a study does not provide much opportunity for continuous monitoring of developments over time and easy extension to more household groups.

2.2.2 NILM concept

In a household setting, each electric appliance consuming energy creates an electric load profile, which depends on its energy demand in accordance to the appliance functionality and use. These profiles can vary greatly according to brand and type of program, but to a certain extent, appliances like washing machines, dishwashers, etc. have a typical, reoccurring profile, which can be used in pattern recognition algorithms.

The focus of load disaggregation is to gain more granular insight in how the total electric energy of a unit e.g a household is used. A potential approach to monitor the load in a household would be to monitor every appliance that consumes electricity individually by placing metering devices such as smart plugs at every power socket. A smart plug measures the total power consumption on an individual power socket in watts and sends a signal to a receiver such as a smartphone or smart thermostat. [53]

Non-Intrusive Appliance Load Monitoring (NIALM or NILM) aims to disaggregate the total energy consumption of a building such as a household into the power signals of individual devices. Appliances have specific power profiles, depending on the task or subtasks they perform as part of a program (see 2.3.3). The key concept of NILM is that appliance signals can be centrally logged by metering technology such as smart meters and aggregated meter signal can then broken down into the individual appliances. [54] [38] [55]

As opposed to sub-metering each single appliance, NILM could pose a more economic and less intrusive approach to retrieve detailed information about the power consumption of individual devices and the associated costs. [56] [40] A methodology to identify appliance loads through measurement at main circuit level was developed at MIT in the 1980s, as described in Hart (1992) as a combinatorial optimisation problem [57]. As the basis of breaking down the total power to component level consumption, the power each appliance consumes needs to be identified. The superimposed power P of N appliances in a time period T, can be expressed by:

$$P\left(t\right) = \sum_{i=1}^{N} p_i\left(t\right) + P_{noise}\left(t\right), t\epsilon\left\{1, T\right\} \tag{2.1}$$

Where p_i is the power usage of each detected active appliance and P_{noise} is the unspecified power consumption, defined as noise [58]. Estimating p_i for i=1,2,. . . ,N, with only P(t) given is the most common approach to solve the problem of power disaggregation.

[58] [59] [60]

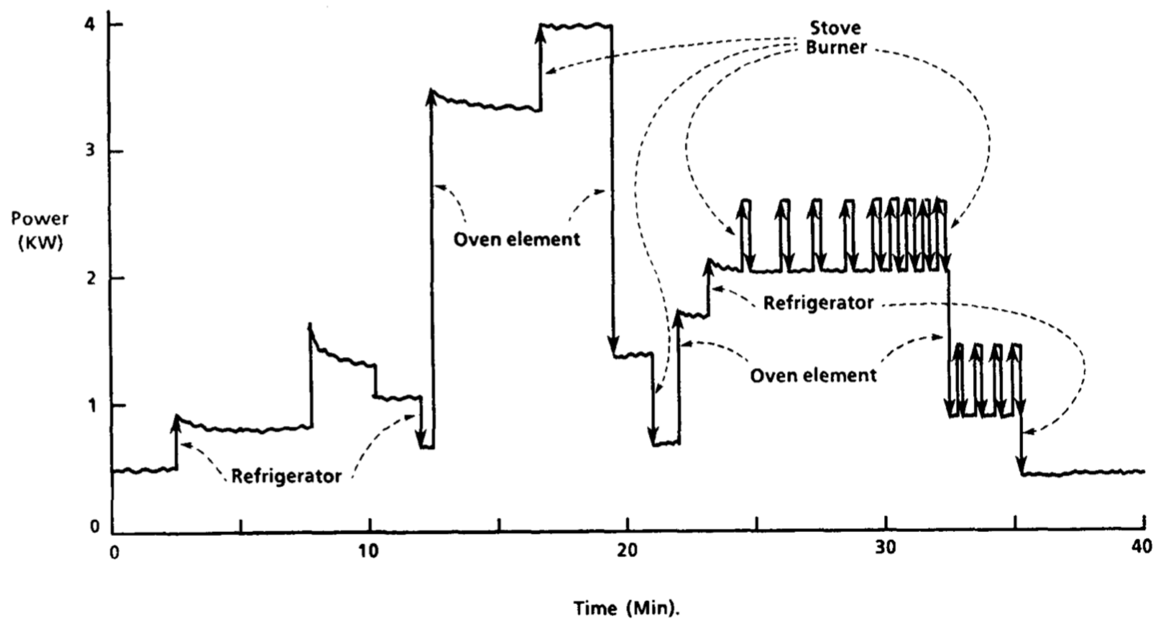

Figure 2.1: *Example disaggregated power profile from Hart 1992 [57]*

2.2.3 NILM system application

To be able to apply the NILM concept in a consumer facing environment, NILM has to be incorporated as part of an information and communication (ICT) system. While handling such a system puts manifold requirements on data gathering (incl. privacy requirements due to GDPR regulation) [61]), transfer, warehousing and processing, the focus in the following section is set on the main process steps that are part of the analytics pipeline or software, rather than the ICT hardware.

Over the years NILM systems have been developed based on varying techniques. However, they share several common principles to break down an aggregated load profile into sub-components. As shown in 2.2 two main phases can be distinguished: model development and the model deployment, which both include data acquisition and a load disaggregation step, potentially followed by a final step providing consumption insights.

In order to prepare a NILM-based model, learning algorithms are often used, that utilise training data during the development stage of the system. The training data is processed and serve as a reference data set for the model deployment. Once the model is developed and ready for deployment, this data set is no longer relevant. However, new input throughout the deployment of the model can be used to continuously update and improve the model reference over time. [62]

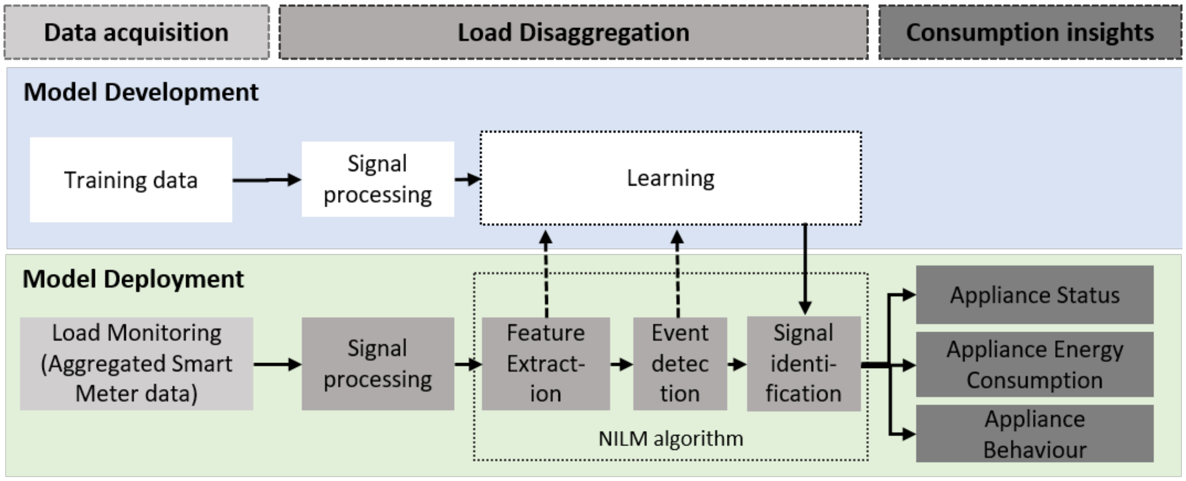

Figure 2.2: *NILM System application adapted from [62], [63]) and [64]*

The model deployment starts with data gathering via a load monitoring system e.g. by a smart meter [65]. The gathered data undergoes a signal processing step before the three main steps of the NILM algorithm, the feature extraction, event detection and signal identification are following. The feature extraction is used in order to transform the time series of continuous load consumption data into an analysable set of features. The calibrated model is then used to make the actual appliance identification on new, unseen data. [64]

The output of the load disaggregation system allows for analysis of consumption insights. This step utilises the outcomes of the NILM algorithm in order to derive useful results for the user. The step, also sometimes referred to as energy management, although technically not managing the appliance energy consumption pattern, provides information to the user to manage their energy consumption. This includes factors such as whether appliances were detected to be on or off at certain moments in time, their energy consumption, to what extent behaviour is normal or an anomaly compared to some norm, appliance degrading over ageing or inefficient/faulty appliance behaviour. [62]

The described process steps can either all be done locally within a household system or the data can be send to a service provider who processes and disaggregates the data in a central analytics system. The resulting consumption insights can then be sent back to the user to provide feedback. [65]

The following sections explain the outlined steps in more detail.

2.2.4 Data acquisition

Data is obtained by a collection unit (hardware), of which its characteristics determine the granularity, performance and type of NILM algorithm that can be applied. The disaggregation of distinctive load patterns is dependant on the following major characteristics:

- **Type of power**: Power is either measured directly or via voltage and current measurements. Measurement approaches can solely focus on true power or parameters arising from these signals such as real and apparent power or I-V trajectory [66], which may then be further used as features.

- **Sampling frequency:** Granularity is defined based on the frequency of monitoring, divided into two main ranges: High sampling rate with a frequency $\geq$ 50Hz to several kHz (i.e. time $\leq$ 0.02s) and low sampling rate with a freq. $\leq$ 1Hz (i.e. time $\geq$ 1s). Signal measurement at higher frequency generally results in a more granular pattern, which provides more opportunity to pick up on specific relevant sub-patterns, however, often cost and intrusiveness of installation scale with higher sampling precision [43].

As pointed out by Tina (2014) the quality of the deployed disaggregation algorithm is highly dependant on the quality of the meter data acquisition. This includes error threshold, accuracy for measurements of voltages and currents with high harmonic content and variable frequency as well as transmission frequency. It has been shown that measurements at sub 1 sec intervals result in more efficiently functioning NILM systems than intervals of lower frequency. [65]

Dependant on the frequency of monitoring the computational effort, analytical method and applicable tools vary. High sampling rates allow for the measurement of reactive power, active power and harmonic content and the assessment of current–voltage waveform. While allowing for a more versatile analysis of the measurements, the method demands for additional hardware and high computational requirement which are impracticable, increase cost and require additional hardware to be installed in the household.

Smart meters have a relatively low sampling rate (e.g. seconds or even minutes) and only provide active power measurements. However, electricity measurements from smart meters are available across the world. The availability of smart meters already installed in the household, makes this approach considerably less intrusive and by that more closely resembling the ideal of ”non-intrusive” load monitoring. As there is less information to disaggregate on, this makes disaggregation more dependant on the applied algorithm and only appliances with higher energy use can be detected with reasonable precision. [43]

Data sets

A number of public data sets are available for NILM analysis, which could be used as input data for model deployment. An overview is given in table 2.1. The data sets contain energy measurements for multiple appliances monitored for a specified period of time. Three main factors of variation have been taken into account when comparing these data sets: Number of households in the monitoring campaign (count), measurement period (time) and measurement sampling rate (frequency). Additionally it was investigated whether dishwashers were part of the monitoring campaign, what type of monitoring location and in what country the data was gathered.

Table 2.1: *Overview of publicly available NILM Datasets*

Dataset	Count [a]	Frequency	Time	DW [b]	Type	Country
[67] Dataport	722	1 min	3+ y	yes	Residential	USA
[68] IDEAL	39	1 s	23 m	yes	Residential	GBR
[52] Loughborough	22	1 min	1 y	yes	Residential	GBR
[69] REFIT	20	8 s	2 y	yes	Residential	GBR
[70] Tracebase	15	1 s	5,2 y	yes	Indiv. appl.	GER
[71] GREEND	9	1 s	1 y	yes	Residential	AUT, ITA
[72] REDD	5	15 kHz	19 d	yes	Residential	USA
[73] UK-DALE	5	16 kHz	2 y	yes	Residential	GBR
[74] AMPds	1	1 min	2 y	yes	Residential	CAN
[75] iAWE	1	1 s	73 d	no	Residential	IND
[76] BLUED	1	12 kHz	1 w	no	Residential	USA
[77] COMBED	1	30 s	1 m	no	Acad. build.	IND
[78] DRED	1	1 s	6 m	no	Residential	NLD
[79] PLAID	235 appl.	30 kHz	5 s	no	Indiv. appl.	USA
[80] WHITED	110 appl.	44 kHz	5 s	no	Indiv. appl.	Var
[32] COOLL	42 appl.	100 kHz	6 s	no	Indiv. appl.	FRA

[a] Number of Households [b] Dishwashers

As shown in the table above, Dataport, IDEAL, Loughborough, REFIT and Tracebase do measure for at least ten households and a time frame of one to several years. The Dataport data set is stated to be the world’s largest data base for disaggregated household energy data [52]. Similarly, the data set from Loughborough University [52] is an often cited data set in research around domestic energy demand modelling, as the data set uses actual measurements as reference to simulate the energy consumption patterns for appliance usage and the household overall. While potentially relevant for energy modelling [20], the relatively low interval of once per minute is lower than modern installed smart meters in many European households, which often measure at 10 seconds intervals, making this data set less relevant for NILM algorithms.

For most other data sets, measurements are often only taken for few households and a short period of time. As discussed earlier, high frequency measurements provide high granularity and the opportunity to measure for other energy variables than active power only. The often cited MIT data set REDD [72] was the first publicly available high-frequency data set collected for NILM research. It provides the AC waveform data (sampled at 15kHz) from which both the real and reactive power can be calculated. However, the small sample size, short time span and needed monitoring and IT hardware make it less suitable and scalable for real-world application.

The IDEAL data set [68] is the most recent published data set. Published in 2021 in Nature, the data set contains (1-second) measurements of electricity and gas usage for 255 dwellings for nearly two years in the UK. The data set contains several components of additional data, such as room temperature, ambient air, humidity, light, sociodemographics, physical characteristics of the building, appliance characteristics, knowledge, attitudes, perception and behaviors of the users. However, not all data is gathered for each household. For only 39 of the 255 households smart plugs were connected.

2.2.5 Load disaggregation

Signal processing

Before the NILM algorithm can be deployed, the gathered signal needs to be pre-processed. This step commonly includes data cleaning (outlier detection, etc.) as well as potential aggregation or reorganisation steps, depending on compatibility of the input resolution and sampling frequency with the NILM algorithm. [64] However, applying more sophisticated approaches to reduce noise in the aggregated power signal could lead to improvement of the deployed NILM system. [81]

Feature extraction

The second step of the load disaggregation step is the feature extraction. Characteristic features for appliances, or signatures/fingerprints, are mathematically described. The pre-processed time series data is further processed to search for these signatures. The type of signatures will depend on the type of device, sampling frequency and type of power sampled.[64]

Typically three types of features are distinguished: Steady state features, transient state features and non-traditional features.

Steady state features
During the steady state operation of an appliance, steady state features can be extracted. For appliance identification, changes in e.g. voltage-current trajectory, time and frequency, noise as well as power related features such as real power, reactive power and root mean square (RMS) are used. [82], [83]

Transient features
If appliances are in a similar range of power, steady state features may not enable an accurate distinction of these appliances [82], [83]. Transient features, which are extracted in the transient state, require sampling frequency in kHz range are able to supply unique appliance information that increase the accuracy of a NILM system. Features include transient power, start-up current, transient voltage noise, transient VI trajectories and higher-order harmonics. The NILM system (i.e. the hardware) must be able to process these higher frequencies accordingly. [82], [84]

Non-traditional features
Non-traditional features compose those features that go beyond electrical signals identification at the energy inlet of an appliance. These can for instance be start time, time of the day, frequency of usage. Among others, light sensing and temperature can be used to derive more appliance characteristic information. [58], [85]

Event identification

The appliance identification is the recognition of an event pattern based on the before extracted features that are typical for the appliance of interest. Two main approaches of event identification are: event-based versus state (non-event) based:

Event based
Event based methods are used to detect events of the appliance moving from one steady state into the next, e.g. switching on or off certain subcomponents [86] This can typically be done by using either of three different types of detection methods: expert heuristics, probabilistic models and matched filter. [83]

When using expert heuristic methods, thresholds are used for events that have been set by expert knowledge (a priori) (fixed) or comparing to the power samples in the steady state phase (adaptive). [87] While potentially simpler in application, the effectiveness of this method is limited, since the level of power consumption of an appliance may vary due to its state or the program. Generally, these methods are not as effective when impacted by noises in the power signal.[88]

Probabilistic methods use statistical and probabilistic features (e.g. standard deviation, variance, likelihood ratio (GLR), chi-squared goodness of fit) to estimate changes to the aggregated signal. [88], [89]

Another method are match-filter methods using likelyhood thresholds and univariate (single) features for identifying universal patterns. [90], [91] Here again single feature thresholds may not always be effective for event detection. Multivariate event detection as recently introduced by [89] appears to be an effective event detection with minimum false rate.

In addition hybrid event detection methods have been researched more commonly. Amongst those are density-based spatial clustering with noise (SBSCAN), using adjacent steady states for hybrid event detection [92] and event detection automatically collecting thresholds [93]. It has been reported that hybrid event detection may be superior to traditional event detection algorithms. [62]

State based
While many NILM approaches focus on event-based, some attempts focus on state based. State based methods, also called non-event based appliance detection utilises a sample from another aggregated signal to compare if the same appliance is on or off during the sampled period. Therefore instead of searching for state changes, a combination of states is matched that follow each other according to some probability profile. State-based methods are often combined with a Hidden Markov Model (HMM) [87] [62] The strengths of the proposed method are its very short and non-intrusive training period in simpler cases and can be an attractive option for lower sampling frequency. [94] However, expert a priori knowledge is often needed and with more complex models can quickly become more computationally expensive. [41]

Learning and inference

The third step of the load disaggregation can be split up into two approaches: optimisation and system training (or machine learning). While the initial optimisation approach as introduced by Hart in the 1980s and 1990s at MIT (described in subsection 2.2.2) has been researched by many [95] [96] [97], it has shown to be of limited capacity with regards to complex environments comprised of large numbers of devices. [58]. Modern day approaches therefore focus on machine learning techniques. The learning algorithms are used to learn (”calibrate”) the model parameters which describe the appliance of interest based on the identified events during the training (”development”) stage. [87]

Root-mean square error (RMSE), mean average error (MAE) disaggregation percentage (D), precision (P), recall (R), Accuracy (Acc) F-measure (F1) are common variables to analyse NILM algorithm performance and trigger adaptation. [98].

Since NILM was introduced, research has been conducted on supervised, unsupervised and deep learning methods. Hybrid approaches are also used that combine multiple models to improve overall performance. [99] [100] [96] [101]

Supervised learning
In a supervised learning approach, a training set of appliances to be monitored are pre-labelled, which ensure that the NILM system learns to recognise the pattern for these appliances. Either event detected power signals or separately measured appliance profiles can be used as appliance signatures.

Examples for supervised methods applied in NILM research include more basic neural networks [102] [103] [104], support vector machine (SVM) [105] [106], Bayesian approaches [107], decision tree [108] and k-nearest neighbor (kNN) [109].

In lab settings the supervised learning methods are proven to be effective for appliance identification and energy disaggregation. However, real-life application is considerably more challenging, since these algorithms are not able to adapt to changing environments such as aging, performance degradation, replacement, and hence appliance signatures need to be frequently updated.

Unsupervised learning
Unsupervised learning algorithms are able to adapt to changes in the environment. A labelling process is not necessary, since the algorithm learns the machine signature itself by distinguishing it from others. The user does nonetheless have to validate that the appliance has been detected correctly by the NILM algorithm, to reinforce the learning. The learning may require a duration of multiple days or weeks. If new appliances are introduced, the algorithm identifies the changed signal and over time captures the new signature. [58] [57] [110]

Widely used unsupervised algorithms in NILM are the hidden Markov model (HMM), k-means clustering and expectation-maximization (EM). [111], [112], [113].

Deep learning
Both the supervised and unsupervised learning approaches are depending on input of pre-defined appliance specific features for an effective NILM modelling. [62] Deep learning algorithms are increasingly researched and used in NILM for appliance identification and energy disaggregation. The algorithms are able to learn to extract features and appliance signature from aggregated power signals, without any external input or vali-

dation. [114]

State of the art developments in NILM application include autoencoder (AE), convolution neural network (CNN), and recurrent neural network (RNN) as well as hybrid deep neural networks (DNN), that have been used as a multi-class classifier for the identification of different appliances and have equally been researched as new approaches in the field of visual recognition [114], [115], [116], [117]

New approaches on transfer learning concepts have shown promising results in NILM application for cross domain transfer, meaning that a model from one domain could be applied on another domain without requiring training. D'Incecco (2019) showed that a DNN model trained on washing machines could successfully be used to detect other appliances (e.g. microwaves and fridges [118]. This would considerably reduce the data collection process, computation time and allowing for connection of different available databases as tested for REFIT, UK-DALE and REDD [118].

However, the high complexity of these models require large straining data sets and extensive development and computing power, making deep learning potentially less applicable as initial development, but rather a next frontier of technological progress.

2.2.6 Consumption insights

Despite the theoretical aim of using NILM for home energy management, research on NILM application focuses mainly on the detection of devices [43] and the algorithms' efficiency [65, 119]. Sources where NILM is used to actually generate consumer insights, to for instance analyse energy efficiency in real world application[44] (as described in figure 2.2), are scarce, but are gaining a growing interest [120]. Few publications have been found that connect the NILM-detected appliances to their efficiency label [45], anomalous behaviour [46], or formulation of household efficiency indicators and feedback to consumers [47]. Demand side recommender systems appear as one of the first approaches to connect NILM for energy usage pattern analysis to recommend energy-aware products/services [121].

2.2.7 Chosen NILM application

As the previous subsection has shown, most NILM algorithms are developed with a prepared standardised academic benchmark data set and often only focus on solving part of the problem, such as feature extraction or a new learning method. Not much applications have been developed that put the complete analytics pipeline (as shown in figure 2.2) together and focus on the potential consumption insights in a real-world consumer facing context.

Moving to real-world application involves multiple additional steps, related to the ICT and data processing as well as overcoming challenges related to data quality, measurement frequency and controllability. Basu et al [43] developed a system taking these steps into consideration in an end-to-end application for 4 real-world households (per 10s measurements) smart meter data for one month. Beyond the ICT system, which is outside the scope of this research, the NILM algorithm that has been developed by Basu et al. can be broken down into 6 stages, contributing to different building blocks to asses the input data sets [122]. The stages follow the same basic logic as the framework presented in figure 2.2 but a more detailed description is given below. The stages are step by step summarised in figure 2.3:

1. **Data base pre-processing**
   Lengthy time series are cut into time sliding window of 1 day with "jump" J (30min.). A day is defined as start time before 24:00, with an additional 4 hours (experimentally measured maximum duration for any appliance in the used data base) for activity of appliances started before 24:00, but crossing over into the next day. Short very high power spikes might occur during some transition stages, but are filtered out as noise with a median filter. Furthermore, missing data is filled in. If the missing data occurs to be over a duration threshold, the time frame is taken out of the dataset to not affect the training set.

2. **Elements generation from events detection and mapping**
   An on/off switching event gets registered if power goes over or (positive event) or under (negative event) a set threshold. As there are many other activities happening within the household and even the power draw of the machine itself is not completely constant, there are consistent minor fluctuations in the power profile. To focus on the larger changes, a threshold is set for what is seen as relevant event. Based on experimental tested typical fluctuations of 32W, the threshold was was set at an absolute value of 35W. One or multiple negative events following a positive event are matched to the respective positive event until the next positive event occurs. The matched events are stored as an element and a power level is assigned.

3. **Clustering of elements based on their power levels**
   The stored elements are then clustered into groups of similar power levels. In order to do that a X-Means [123] clustering algorithm is deployed. X-Means is an unsupervised machine learning algorithm that follows the K-Means principle, where data points are clustered into a number of K distinctive groups. The data points within each group are as similar as possible and the data between groups as distinctive as possible. In order to do that the data are clustered together

around a common average (centroid) for each of the K groups. On beforehand these centroids are unknown, but by initiating K random values and by reducing the error (distance) function of data points belonging to each group, over multiple iterations the algorithm is able to form these distinctive groups. While for the K-Means algorithm the number of groups has to be known on beforehand, the X-Means algorithm adds an additional step in order to find the optimal number of groups. In this case the X-Means is used to group the elements into several distinctive groups with similar power levels.

4. **Feature generation based on power and components clustering**
   "Part-based" models were developed for different multi-state appliances, based on their typical temporal distribution between elements. In order to disaggregate the appliance of interest, their distinctive pattern of states (i.e. elements) has to be recognised, based on defined appliance relevant features. The features are comprised of appliance specific occurrence, duration and power levels of different states, during a particular time period. Therefore for each time sliding window a set of features is generated for each element within each cluster category. The following features are generated for each element within a cluster and a time sliding window:

   (a) Mean and standard deviation of the power level of each cluster.

   (b) Mean and standard deviation of the duration of each element in each cluster.

   (c) For similar power levels and duration:

      i. Mean and standard deviation of the On duration of an element (i.e. duration between a positive event and the next negative event) within each cluster in the considered time sliding window.

      ii. Number of occurrences of elements within the time sliding window.

      iii. Average time difference between the starting time of successive generated elements in the clusters (the start time is the first occurrence of the considered element).

   Some outliers are filtered out based on set thresholds compared to the mean value.

5. **Building the "part-based" model of the classifier**

   (a) *Training instances*
   Consecutive events with power variations below a certain threshold are merged to a single event. All power readings below the set threshold are lumped as one appliance, resulting in activity between these blocks getting paired. When an undefined appliance goes through a complete cycle of activities within a single time sliding window, it is labelled as ON for that specific window.

(b) *Classification*
Classification is then used to distinguish to which appliance category the pattern of detected elements belongs. To do that the detected pattern is compared with the specified "part-based" appliance models using a Support Vector Machine (SVM) algorithm with a Radial Basis Function (RBF) as kernel. The SVM is a supervised machine learning algorithm that takes in the features of the detected appliance and classifies which appliance model this most resembles, based on a pre-trained parameter set. In order to find the closest match, the SVM divides a multi-dimensional space defined by the feature set and calibrates its parameters to draw a hyperplane (decision boundary) that maximises the distance of the hyperplane to the closest point in each class (support vector), while minimizing the parameters. The shape of the hyperplane is based on the kernel function used to transform the multidimensional data for the SVM. While a basic SVM draws linear lines to divide the space, the RBF kernel was used in order to divide the data into non-linear shapes. Finally, the trained SVM classifies the detected appliance based on in which location its feature set falls within the SVM's calibrated feature space.

6. **Testing of the built model on an unseen data base** New unseen (labelled) data is introduced to test how well the detection performs, compared with the pre-assigned labelling.

7. **Post-processing calculation (start-time, duration and energy)**
Factors such as start time (time of day or compared to other appliance), total duration or total energy level can all be used to distinguished between appliances with comparable profiles (e.g. dishwasher and washing machine or oven, tumble dryer and hairdryer)

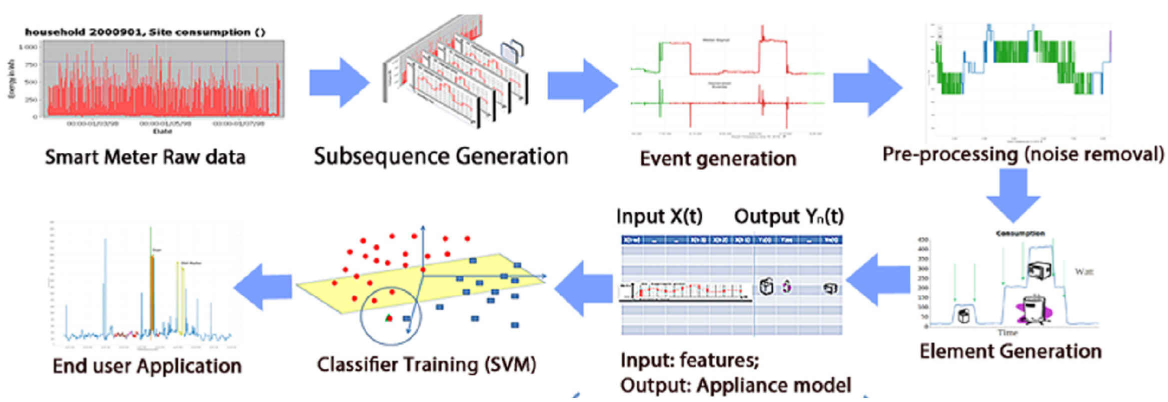

Figure 2.3: *NILM overall data-processing and analytic pipeline from Basu et al. (2017) [43]*

The results of the research indicate that smart meter data can be used to correctly identify major household appliances (with 80% accuracy) using this methodology. To compensate for the relatively small amount of subjects in the data set and to overcome the challenge of messy real-life data, and having only a very small sample, the model makes use of a hybrid approach. The model combines pattern recognition and machine learning with specification of the actual patterns to disaggregate based on knowledge about the specific profiles of the appliance(s) of interest. Additional insight on a specific appliance type (see following section on appliance consumption 2.3) could therefore be utilised to aid and further develop the detection algorithm.

Part of the developed detection method is to recognise different states that together constitute the distinctive pattern of an appliance such as e.g. the washing program of a dishwasher. These features, i.e. length (time) and height (power), provide opportunity for further statistical analysis to gain deeper insight on separate components of the power profile of an appliance and the potential to develop consumption insights from the patterns that arise by comparing this for many different households.

2.3 Appliance Energy Characteristics

To be able to detect the energy usage and interpret these results, better understanding of the energy usage of dishwashers is needed. This section starts with elaboration on traditional methods of analysing appliance energy usage and continues with more specific analysis of the dishwasher functionality, resulting in a unique load pattern that can potentially be detected by the NILM system.

2.3.1 Dishwasher energy usage

As introduced in the research gap (1.1.2), NILM research could form a modern approach to gathering appliance level household statistics. Traditionally this data is gathered by surveys, metering or literature analysis. In order to gain a better understanding for factors involved in dishwasher energy usage, findings by more traditional approaches are presented below.

An often cited research from Loughborough University looked into the household energy demand patterns based on plug measurements for individual appliances over the course of 2 years, to show how different appliances affect the total energy demand of the household over time. [52] The study uses actual appliance measurements for most household appliances from 22 households in the UK over a two year period as reference to simulate the energy consumption patterns for appliance usage and the household overall. As the study focuses more on energy demand modelling and household activity, less emphasis is placed on the energy efficiency implications.

An important meta analysis by van Holsteijn en Kemna [23] was based on a literature research to develop a model for household energy consumption in the Netherlands based on projections per appliance. The study looked at data from 1995 until 2005 and projected it until 2020. As data basis the study used BEK data [124], GEA appliance efficiency research [125], ownership estimates based on panel survey research and literature information such as the Ecodesign directives (see section 2.4.2). Usage was described by consumption components of the respective appliance such as average usage time and the number of cycles per year or hours per day in certain usage modes i.e. different washing moments. Electricity consumption per unit such as kWh per cycle or electricity usage when switched on were considered as appliance characteristics. The results of the study show that a growth in ownership of dishwashers from 39.5% in 2000 towards 60 % in 2020 was expected to take place, with only A label machines being purchased from 2014 onward (based on 13 year renewal age). From 2015 onward an A label machine with 12 couverts of 1kWh consumption per wash was assumed for the business as usual scenario and multiplied with a behavioural inefficiency factor of 1.09

(in accordance with Ecowet, Task 4), resulting in 1.09kWh per wash. Furthermore 220 washing cycles per year were assumed to stay constant over the full period, resulting in 240kWh per year.

Another prominent study, the Preparatory Studies for Eco-Design Requirements of Energy-using Products (EuP) for dishwashers, found in a questionnaire of households conducted in 10 European studies, that dishes are cleaned by dishwashers in only 30-40% of the cases (i.e. dishwasher penetration and active usage), leaving 60-70% of the dishes being hand washed. Dividing available European electricity data on dishwasher consumption of 16.2 TWh (EU-15) [126] by its 160 million households equals a stated energy consumption of roughly 240 kWh per household (with an automatic dishwasher) and year. [24]

A detailed plug based research was carried out in 2008 [36], sampling data with plug connected dataloggers from 100 french households with 10 minutes interval over one year. During one month, participants also filled out questionnaires describing the different washing/drying programs in use. Comparing the results with former data from 1995 and 2008 the average yearly energy consumption decreased from 273kWh to 171kWh (-37%), the number of annual washes from 213 to 189 and hence the energy usage per wash from 1.25kWh to 0.90kWh.

2.3.2 Dishwasher operating characteristics

To better understand the typical load pattern of a dishwasher, the dishwasher operating characteristics are described. Common household dishwashers contain one or two internal dish racks for the storage of dishes (depending on the device size). When the machine starts, a defined amount of inlet fresh water gets heated with a heating element and passes through to the sump pump, which is located at the bottom of the dishwasher. From there, the water is pumped up and sprayed over the items by moving spray arms [127].

Operating phases

Dishwashers operate at several washing programs, which differ in duration and temperature range. A washing cycle can last from 15 minutes to 3 hours. Depending on the selected program, the washing process includes several phases [128]:

- **Pre-cleaning phase:** Some programs use a pre-cleaning phase, with cold or with slightly heated water.
- **Heating phase 1**: Heating up the water to cleaning temperature. The supplied water is heated by means of a heating system (nominal power 1800-2500W) to

the temperature corresponding to the selected program. This process can be interrupted for a short time to equalize the temperature.

- **Rinse phases:** Between heating phase 1 and 2, 1+n rinsing cycles are interposed for the cleaning process
- **Heating phase 2:** For the last rinsing process, the water is heated to maximum temperature of the programme.
- **Drying phase:** The final heating phase is followed by the drying process, during which the water evaporates on the dishes.
- **Start delay or standby phase:** In addition to the active washing cycle time delay or standby functions may be used allowing for defining a particular start or end time

Energy per wash

The extent to which these different phases contribute to a total wash can vary not only for dishwashers, but other similar household appliances like washing machines. In 'Construction of a virtual washing machine' the authors modelled the water and energy consumption of a washing machine by regressing measurements of different features from tests of nine different washing machines of five different manufacturers. The authors were able to explain 92% of the variance in energy consumption per wash with a linear equation taking into account the amount of inlet water, washing temperature and total duration of the wash, where the amount of inlet water itself was a function of the amount of laundry placed in the machine and the rated capacity. [129] While a different appliance, this shows the relevance of the water heating phases in explaining the overall energy consumption per wash and the potential of describing an appliance profile by its features. The next subsection continues in more detail how the different phases together constitute to a dishwasher's load pattern.

2.3.3 Dishwasher load pattern

Information from the dishwasher's operating characteristics translate into its power profile or load pattern, that can potentially be detected by the NILM system. In the dishwasher, electrical energy is required for active and passive processes described in the previous subsection. A typical power demand profile of a dishwasher is presented in figure 2.4:

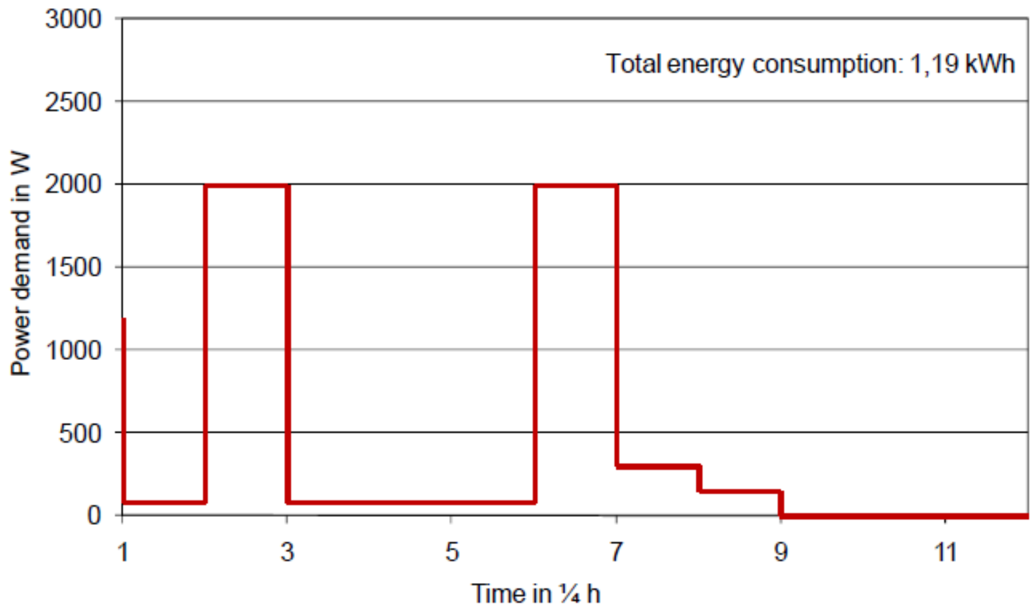

Figure 2.4: *Example load profile of dishwasher showing two distinct heating phases [128]*

The profile depicts the power demand which is automatically drawn from the power line for a washing cycle at a standard cleaning program without using a start delay function. The power demand in watts is plotted over 1/4 hour time steps. The area under the demand curve sums up to a total energy consumption of 1,19 kWh. Energy is mainly used to heat the water for the rinse cycles (two main peaks). It is also consumed by the motor of the circulation pump and other electronic components of the machine, such as the user interface (mainly in the middle). For the drying phase at the end, rinsing water is heated up to be able to store enough heat in the dishes for the water to evaporate. Additional short burst of heating can potentially be used to improve the drying process in this final rinse (final activity at the end of the profile). In the passive state, small amounts of energy are used to maintain safety functions. If start delay functions are used, additional energy is consumed for the standby activity. [130]

The following graphs of detected profiles exemplify two variations within this pattern:

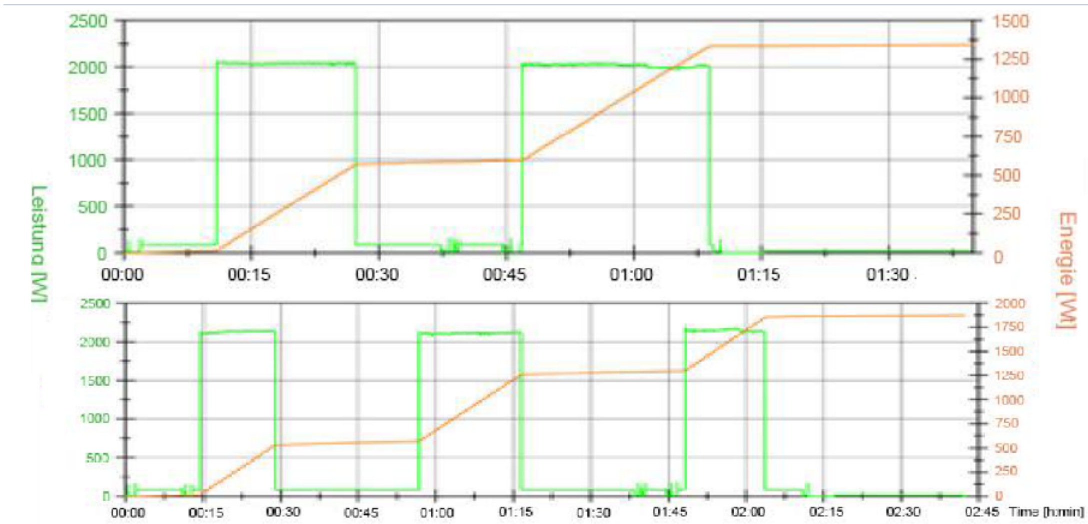

Figure 2.5: *Example load profiles detected for the "Normal" standard program (upper) and the hotter "Pots Pans" program (bottom) of dishwasher Miele G1222 [130]*

The two charts in figure 2.5 show the power profile in green on the left axis and the energy accumulated over time in orange on the right axis. It can be seen that the "Normal" standard program in the upper chart has different durations for the two heating phases. The second heating phase (over 20min.) takes considerably longer than the first (about 15min.) heating phase. The total energy usage amounts to 1.25kWh. By comparison, the two heating phases in figure 2.4 both take about the same time (about 15min.) and the total energy usage of the wash amounts to 1.19kWh, despite the additional energy usage in the drying phase. The hotter, more thorough "Pots Pans" program depicted in the chart in the bottom takes considerably longer and has three instead of two heating phases. The three phases all take about 15 tot 20 minutes, with the one in the middle the longest. Furthermore, changes can be noted on the lower power level that result from pumping and rinsing in between the water heating. Finally, some differentiation can be seen between the maximum power level (although all closely around 2000W). This program has a considerably higher energy usage of 1.75kWh. Taking these variations into account, shows there is some differentiation between profiles for the same dishwasher, based on the program chosen. Differentiation between dishwasher brands and models exist as well, so the specific program cannot be recognised from the number of heating phases. Generally a dishwasher program is characterised by several heating phases with relatively low energy usage in between for the pumping and rinsing. [130]

Day-Time variation

As shown by the research group at Loughborough University not only the profiles, but also load patterns throughout the day can vary. [52] Therefore, besides load during operation, is the chance of an appliance being in operation during different times of the day (as discussed in 2.2.1).

Another analysis, conducted on the demand shifting potential of appliances, using the EuP consumer survey transferred the outcomes into an estimated average curve for start time of dishwasher in Europe over the course of a day and the related power demand needed for operating a dishwasher per day and household. The results as shown in the Figure 2.6 show that the peak of washing activity lies between 7-9pm, with a smaller peak between 6-8 am. When the chance of a dishwasher being on from the left figure is multiplied with the average power draw from the power curve in figure 2.4 this results in the power curve in the right figure. The power demand as shown in the right graph reaches its maximum at around 8 pm with roughly 87 W. [24]

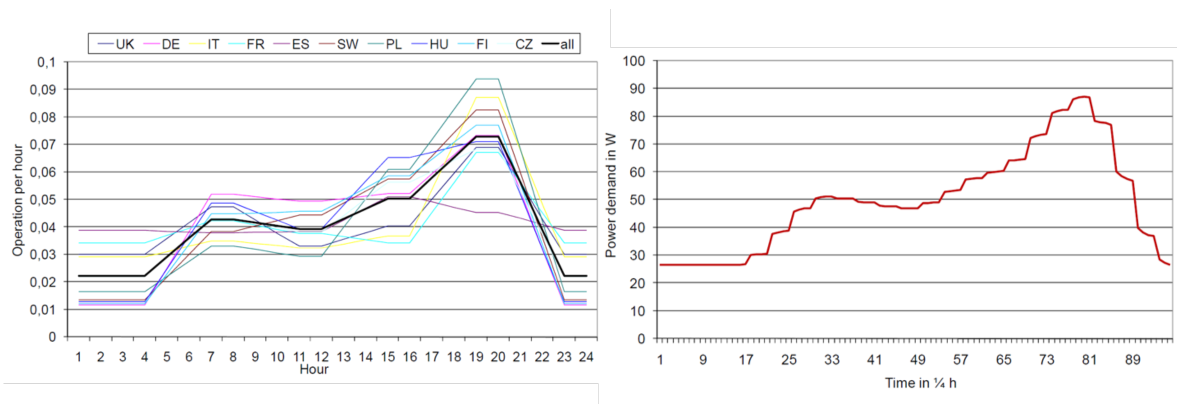

Figure 2.6: *Start time of dishwashing in 10 European countries (probability estimate) (left). Daily load curve of a dishwasher (averaged start time function)(right) [130]*

Other appliances

For the expedient use of detection algorithms for load profiles of electrical appliances, the distinction of different appliance's profiles is needed. For this distinction a sufficiently accurate understanding of load curves is necessary [131]. Most electrical appliances can be grouped into few categories of load types. These are based on the consumption pattern of alternating current (AC). Barker et al. (2013) divides them into resistive (loads with heating element), inductive (loads with AC motors) or nonlinear (devices using switched-modepower supplies (SMPS); mainly electronic devices, such as desktop computers and TVs).

Many large household appliances, consist of multiple different components, which each make use of one of the mentioned load types by which a combination of resistive, inductive or nonlinear power consumption is created (composite loads). [131]

Zeifman et al. (2011) suggests a characterisation appliances based on their power draw for permanent consumer devices (e.g. wired smoke detector), on-off appliances (most common household appliances such as toasters and light bulbs), finite state machines (FSM) (loads with several definite switching states, e.g. washing machines, freezers/fridges, dishwashers dryers), continuously variable consumer devices (e.g. dimmer lights, television). [38]

To compare dishwasher to other similar appliances, figure 2.7 presents load profiles of selected major household appliances with composite loads in one-second resolution. The figure shows typical profiles for a (a) washing machine [43], (b) dryer, (c) fridge/freezer, (e) dishwasher, (f) oven, (g) AC unit [132]. Washing machines, dryers and dishwashers consist of a heating unit (resitive load) and AC motor for e.g. spinning clothes and circulating water (inductive load). They often make use of repetitive cycles of these elements. The fridge/freezer and AC units shown in Figure 2.7 (c) shows small, repeating fluctuations that reoccur for each compressor cycle.

Chart (d) and (h) show the combination of multiple of these appliances together resulting in a household's total electric activity profile.

With the understanding for other load profile characteristics it becomes possible to extract specific signals even if an overlay of signals, as can be seen in (d) between 2-3 hr occurs. Techniques to do this can vary as elaborated in more detail in chapter 2.2.

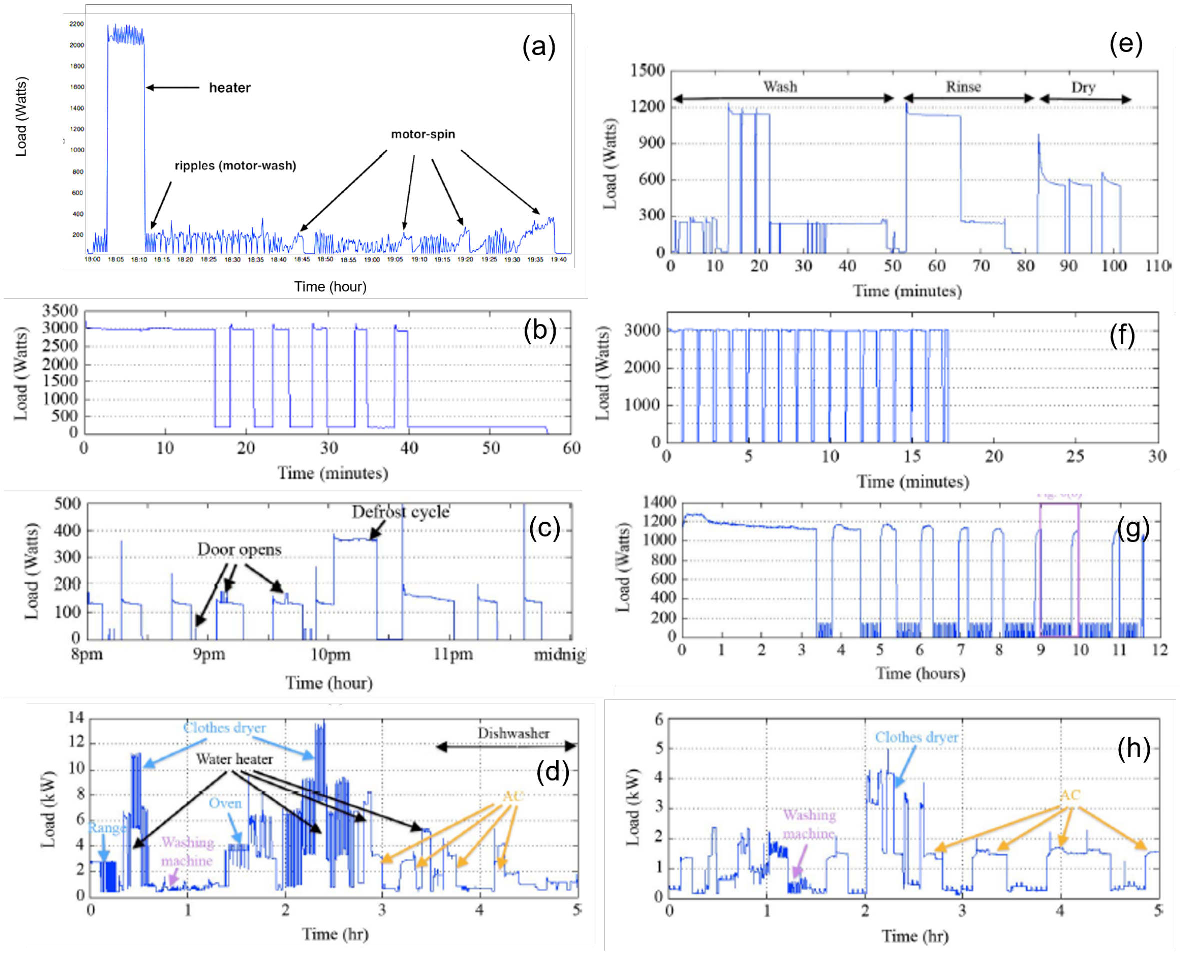

Figure 2.7: *(a) washing machine, adapted from [43];(b) dryer;(c) fridge/freezer; (e) dishwasher; (f) oven; (g) AC unit; (d) and (h) household load profiles, adapted from [132]*

2.4 Appliance Energy Efficiency

Now it is established how NILM could be used to detect appliances and how energy consumption of dishwashers is characterised and can be analysed, this section progresses with how this could translate to energy efficiency. First energy efficiency is defined, next standard efficiency settings for dishwashers are given, as according to European regulation. Finally, an overview is made of research investigating what factors could influence dishwasher energy efficiency in real-life application.

2.4.1 Defining energy efficiency

The EU Energy Efficiency Directive uses a very broad definition and describes energy efficiency as the pursuit of obtaining the maximum benefit from energy use. It is defined by the "ratio of output of performance, service, goods or energy, to the input of energy" [133]. Due to the broad nature of the definition, assessing energy efficiency can range from energy intensity considerations of entire systems to efficiency of individual activities such as energy efficient machine operation. Whereas energy saving is referred to as a reduction in the input energy, without any reference to the desired output.[134]

As described in Introduction to Energy Analysis, by Blok and Nieuwlaar (2017), energy efficiency is the amount of energy (E) needed to produce a certain measurable output (P). This is often listed as a percentage between 0% and 100%, and described with the following equation for efficiency (η):

$$\eta = \frac{P}{E} \tag{2.2}$$

For energy conversion processes this can be a relevant way to describe the useful energy output (E_{out}) resulting from energy input (E_{in}). For end-use applications, such as appliances, the inverse of energy efficiency, specific energy consumption (SEC), is more often used:

$$SEC = \frac{E}{P} \tag{2.3}$$

SEC describes the energy needed to produce one unit of useful output. This depends on (i) the technical efficiency under uniform conditions and (ii) its operation, taking factors such as ambient conditions and usage behaviour into account. Several pieces of equipment with a similar unit of output can then be compared with each other to see which one is more efficient. In order to do this, the energy efficiency index (EEI) can be used, generally defined as the energy use per aggregate output.

The aggregated specific energy consumption can be an ideal candidate for the reference energy use, to calculate the EEI:

$$EEI = \frac{Actual\ energy\ use}{Reference\ energy\ use} * 100 \tag{2.4}$$

The energy index compares the actual energy consumption of the process with a defined reference point.[135] In terms of dishwashers we could think of SEC as the energy consumption to do a single wash (energy per wash) and define the reference energy as the energy needed to do a an aggregate of one whole year of washing for one household, see EU dishwasher efficiency calculation in next subsection.

2.4.2 Standard efficiency settings

Efficiency regulations

Industrial energy efficiency measures are addressed by different legislative acts. Household appliance energy efficiency is currently established in two binding directives with Commission Delegated Regulations for Dishwashers in place, where some of the above mentioned parameters were considered:

- Ecodesign Directive 2009/125/EC addresses supply of energy efficient products. The performance standards set in the Ecodesign Directive concern the improvement of energy efficiency of products and market clearance from inefficient products. [136]

- Commission Regulation (EU) No 1016/2010 implementing Directive 2009/125/EC of the European Parliament and of the Council with regard to ecodesign requirements for household dishwashers [137]. Repealed by Commission Regulation (EU) 2019/2022 [138] (not considered in this research, see Appendix A.0.1).

- Energy Labelling Directive (2010/30/EU), repealed by Regulation (EU) 2017/1369 sets a standardised energy label to support consumers in the assessment of energy efficient products. The regulation required a growing amount of appliances according to a standardised energy usage rating scheme indicated on a scale raging from G (least efficient) to A+++ (most efficient) [139]

- Commission Delegated Regulation (EU) No 1059/2010 supplementing Directive 2010/30/EU of the European Parliament and of the Council with regard to energy labelling of household dishwashers. [140] Repealed by Commission Delegated Regulation (EU) 2019/2017 (not considered in research) [141] (see Appendix A.0.1)

- Directive 2012/27/EU on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU [142]

Furthermore Directive (2009/72/EC) requires smart metering systems to be implemented by the member states, since they may encourage consumers to more efficiently use their energy and may enable smart home management.[143]

Efficiency calculation

For the cleaning performance Regulation (EU) No 1016/2010 implementingthe Ecodesign Directive for dishwashers states an efficiency index for the items "sufficiently placed" in the dishwasher body. The energy efficiency labelling regulation of household dishwashers is determined on the basis of this Energy Efficiency Index (EEI). The EEI is calculated for an estimated annual energy consumption as follows [137]:

$$EEI = (\frac{AE_c}{SAE_c}) * 100 \tag{2.5}$$

AE_c = annual energy consumption of the household dishwasher
SAE_c = standard annual energy consumption of the household dishwasher

Based on the measured energy per wash under standard testing conditions (E_t), the annual energy consumption AE_c is calculated as follows and expressed in kWh/year:

$$AE_c = E_t * 280 + \frac{[P_0 * \frac{525{,}600-(T_t*280)}{2} + P_l * \frac{545{,}600-(T_t*280)}{2}]}{60 * 1000} \tag{2.6}$$

E_t = Energy consumption for the standard cycle in kWh
P_l = Power in 'left-on mode' for the standard cleaning cycle
P_o = Power in 'off-mode' for the standard cleaning cycle
T_t = Programme time for the standard cleaning cycle, in minutes
280 = total number of standard cleaning cycles per year

Based on the dishwasher size in ps (number of place settings), the standard annual energy consumption SAE_C is calculated in kWh/year:

1. for household dishwashers with rated capacity $ps \geq 10$ and width $> 50cm$:

$$SAE_c = 7,0 * ps + 378 \tag{2.7}$$

2. for household dishwashers with rated capacity $ps < 9$ and household dishwashers with rated capacity $9 \leq ps \leq 11$ and width $\leq 50cm$

$$SAE_c = 25,2 * ps + 126 \tag{2.8}$$

Dishwasher energy consumption is calculated by assuming 280 rinsing cycles per year, which accounts for 5 to 6 rinsing cycles per week. The values are based on the manufacturer's standard programme for normally soiled dishes. The annual water consumption is determined on this basis. A benchmark for the best available technology on the market for household dishwashers with 13 place settings was stated with an energy consumption of 0,83 kWh/cycle, corresponding to an overall annual energy consumption of 244,9 kWh/year, of which 232,4 kWh/year for 280 washing cycles and 12,5 kWh/year due to the low power modes.

For the classification of the efficiency, a scoring system was introduced based on the efficiency index with D (least efficient) to efficiency class A+++ (most efficient) [137]. As mentioned under Efficiency regulations (above) the labelling has slightly changed in 2019. The new system has simplified the equations to calculate the EEI on a per wash basis, instead of the AE_c above that makes assumptions on number of annual washes and takes the standby mode into account. (Appendix A.0.1). All dishwashers investigated within this research fall in a time frame making use of this previous regulation, for which the efficiency table is presented here below:

Energy efficiency class	Energy Efficiency Index
A+++ (most efficient)	EEI $<$ 50
A++	50 $\leq$ EEI $<$ 56
A+	56 $\leq$ EEI $<$ 63
A	63 $\leq$ EEI $<$ 71
B	71 $\leq$ EEI $<$ 80
C	80 $\leq$ EEI $<$ 90
D (least efficient)	EEI $\leq$ 90

Table 2.2: *Energy efficiency classes [137]*

As can be seen the EEI changes in stepsizes of 10%. Hence a given label is 10% more efficient than a label for a stepsize lower. As could be seen in equation 2.7 and 2.8, the standard energy consumption and hence calculated EEI, is also dependant on size or capacity of the dishwasher. The capacity of a dishwasher is expressed by number of standard place settings (ps). [137], as specified in 'European norm for electric dishwashers for household use' (EN 50242) [144]. Place setting is a fixed number of crockery and cutlery items in a given composition and size that is used for one menu sequence (see table 2.3). In addition to the specified number place setting items, seven serving and cutlery items are utilised, see table below:

Items per couvert		Additional items per wash	
Dinner plate [cm]	ø26	Oval plate [cm]	32/35
Soup plate [cm]	ø23	Serving bowl [cm]	ø16
Dessert plate [cm]	ø19	Serving bowl [cm]	ø13
Saucer [cm]	ø14	Serving spoons (2x) [mm]	260
Cup [ml]	200	Serving fork [mm]	192
Drinking glass [ml]	250	Gravy ladle	
Knife [mm]	203	Serving bowl [cm]	19
Fork [mm]	184		
Soup spoon [mm]	195		
Teaspoon [mm]	126		
Dessert spoon [mm]	156		

Table 2.3: *Standard items for dishwasher capacity, based on [144]*

2.4.3 Real-life inefficiency

The implementation of Energy Labelling and Ecodesign regulation on major household appliances has been foreseen to create an energy savings potential of 13% [35]. Stamminger et al sees the directives only affect the market slowly, since ownership of a dishwasher is about 10 years on average. It is therefore crucial to assess the real-life usage of a dishwasher and its efficiency. In the 'Preparatory Studies for Eco-design' real-life energy usage is estimated to be 12.2% higher than under standard test conditions, even 29% when including the energy consumption used for manually pre-rinsing the dishes. [24]. This section gives an overview of factors affecting the efficiency of a dishwasher in real-life usage. Example studies of other household appliances are used if relevant literature available and comparison seems applicable. The following items have been stated to determine the consumption of water and energy according to [128]:

- Frequency of usage
- Load size used
- Selected programme and nominal temperature
- Additional rinsing
- Low power mode (start delay + standby)
- Machine efficiency
- Ambient conditions

Many of these determining factors for dishwasher energy use in real-life are driven by behavioural factors (i.e. choice of program, loading, etc.) impacting the energy use of a dishwasher [145]. Hence, the behavioural component is discussed as a final overarching point.

Frequency of usage
With regards to the use of machine capacity the annual energy consumption (AE_c) equation (see equation 2.6) assumes 280 total number of standard cleaning cycles per year. This is set as a "standard" number of cleaning cycles, but no further mentioning is made about an efficient number of annual cleaning cycles, since this is not a machine but household behavioural characteristic.

According to Van Holsteijn en Kemna this could potentially be an overestimation of about 23%. [146] citing 4.1 washes per week or 214 washes per year. The EuP study in 10 different countries in Europe showed that an average of 4,06 washing cycles/week was declared, adding up to 203 washing cycles per year (50 weeks) [24]. The Ecodesign directive assumed a usage of 220 times annually [137] or 4.5 times per week taking 3 holiday weeks into account or 4.2 times per week not taking holiday weeks into account.

The efficient frequency is related to a combination between number of users, dishwasher capacity and expected cleaning performance. [147][130].Comparable results were found for washing machines, where it was shown that the frequency of operation largely depends on the household size, as this defines the amount of wash to be treated per cycle. [148]

A table retrieved from Van Holsteijn en Kemna summarises data from VEWIN and TNS Nipo research on water usage in Dutch households. The studies include the dishwasher usage by number of family members in previous editions (see figure 2.8). The weekly usage is dependent on household size and can be seen to be 2.2, 4.2, 5.1, 6.4 and 6+ for households ranging from 1 to 5+ people respectively (in 2007) and was shown to have increased slightly over time. The last edition of the VEWIN study (from 2016) showed this had actually slightly decreased and stabilised at an average of 0.17 times per day per person, compared to 0.25 in 2007, while the number of households owning a dishwasher had further increased to 75% on average, also depending on household size. [149].

Vaatwasser

bron: VEWIN Watergebruik thuis 2001, 2004, 2007

jaar	2001			2004			2007		
huishouden	penetratie	frequentie pppw	frequentie hhpw	penetratie	frequentie pppw	frequentie hhpw (1)	penetratie	frequentie pppw cycli per pers. per week	frequentie hhpw (1)
1 pers	25%			30%	2.9	2.9	37%	2.2	2.2
2	41%			53%	2.0	4.0	61%	2.1	4.2
3	54%			59%	1.6	4.8	70%	1.7	5.1
4	62%			68%	1.5	6.0	74%	1.6	6.4
5+	75%			74%	1.2	>6.0	70%	1.2	6
gem.pppw	42%	1.68		50%	1.75		54%	1.75	
gemiddeld hh (1)			3.86			4.0			4.0

(1): Calculatie VHK op basis van freq. pppw * aantal pers. in hh. Voor gemiddeld hh is 2.3 pers. gehanteerd.

Figure 2.8: *Dishwasher penetration and weekly usage by household size [23]*
pers = persons, gem = average, pp = per person, hh = per household, pw = per week

Load size

Analysis by Van Holsteijn en Kemna of the CECED database shows that the majority of new dishwashers on the market in Europe in 2012 were 12 ps (46.6%), followed by 13ps (28.9%), 14ps (9.5%) and 9ps (8.2%), averaging at 12.1ps. In comparison in 2005 12ps dishwashers represented 81.8% of the market followed by 9ps (12.2%) and averaging at 11.6. However, consumers often may not fully fill up there machine because they assume it is full due to inefficient placing or the consumer might need cleaned dishes. It is estimated this might reduce a 12 setting machine to average 9ps in real life. [146]

A non-efficient capacity exploitation may impact the energy consumption negatively since machines are designed to reach their efficiency maximum at maximum capacity. In the study by Richter et al. for 20% of the washes, the maximum load capacity of the baskets were not exploited.[150]. In the VEWIN study 82% of households stated to usually fully fill their dishwasher, another 13% fill it for three quarters, 2% even less. [149].

Programme and temperature

For the evaluation of efficiency related characteristics for washing appliances, such as dishwashers and washing machines, the combination of the most used washing program and its nominal temperature are two key parameters. [147] Stamminger investigated in a study of about 2500 households in ten European countries, that the main washing temperature lies between 50/55°C or 60/65°C, averaging on 59,3°C (see figure 2.9). This was estimated to add an average of about 10% of energy, compared to standard test settings. [24].

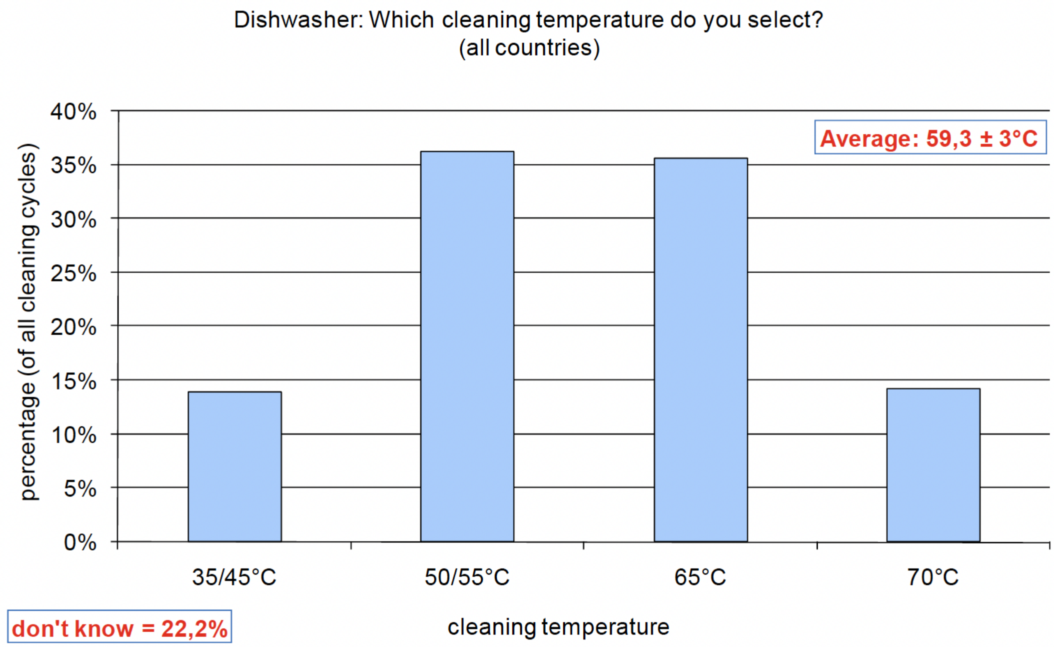

Figure 2.9: *Relative occurrence of dishwashing temperatures in Europe (average of 10 countries) [24]*

Another consumer survey on 200 households revealed that 52% of the interviewees with a dishwasher favoured washing temperatures of 65°C resulting in a higher energy consumption per washing cycle than stated by the respective energy labels [150]. Van Holsteijn en Kemna state there is evidence that the eco program is not the most used. As this is the only program considered in testing, they conclude there is not enough stimulation to improve energy efficiency of other programs via legislation. [146]

Other studies on electrical appliances report similar findings. For refrigerators for instance it was found that 25% of households surveyed, operate the device at lower temperatures than recommended according to their efficiency classes.[151] Besides assumptions on performance, this might be caused because many consumers would be unaware of the temperature connected to the program. Dishwashers and refrigerators often do not show the temperature at the usage panel, but only contain it in the user manual. [128]

Additional rinsing

According to dishwasher manufacturers pre-washing of dishes is not necessary and scraping and wiping of leftovers before placing dishes in the washer is recommended instead. Not many consumers soak dishes, but manual pre-rinsing is more common in some researched countries: 42% in Italy and 25% in Sweden, compared to only 4% in Germany and 8% in the United Kingdom and with varying degrees of water use, averaging at 15 liters. [150] Consumers might clean dishes on beforehand out of habit, even though a pre-rinse cycle is usually available on the dishwasher. However, only 3.2% of consumers are stated to always use the extra rinse cycle, while 5.2% often make use of this functionality. Using a rinse option only uses minimal extra energy and about 5 liters of water, compared to an estimated 0.1kWh and 3 liters of water per place setting, where about half of the dishes are estimated to be manually pre-rinsed in an average of 31% of European households. [24].

Choosing to use an extra (cold) rinse option on the dishwasher can therefore increase the water usage, but only minimally affect the energy consumption of the wash. Manually (warm) rinsing of the dishes can add additional extra water and energy usage. Changing this consumer behaviour could therefore have considerable saving potential for the overall dish washing process. However as this research only focuses on detection of the energy consumption of the dishwasher itself and not the full process and since the usage of automatic rinse options is only very minimal, this additional energy consumption is not further considered within the scope of this research.

Low power mode (start delay + standby)
The use of a start and delay/standby function may result in a higher energy use, but the additional energy usage in this "inactive" use phase is rather small and most consumers (45%) say to never use them. Those who use it often (27%) and those who sometimes use it (15%) used this function at least once a week with 66% choosing a delay time of 0-3h. [130]

Ambient conditions (inlet water)
With regards to external environmental factors, research on refrigerators revealed that ambient temperature is an important influential factor on the energy consumption [152]. With reference to dishwasher functionalities, this factor could potentially be considered as the impact of variation in the inlet water temperature on the energy consumption per wash. The temperature difference that needs to be overcome by heating of the inlet water to the desired temperature of the chosen washing programs is dependant on the water inlet temperature, which again is dependant on the external environment/ambient temperature. The water inlet temperature in Germany for instance might be considerable lower on average. While no further information was available, this estimated to be about 10°C, compared to the 15°C as used under standard testing. It is noted that in countries with higher water inlet temperature this could also lead to lower energy usage vice versa. [24]

Machine efficiency and age
The comparison of household dishwashers under standard settings with a capacity of 12 to 14 plate settings shows that the annual energy consumption may vary by over 100 kWh between A+++ and older appliances [153]. The machine age has been shown

to impact energy usage significantly [128]. As Figure 2.10 indicates the energy usage of newer models coming to the market has reduced considerably over the years.

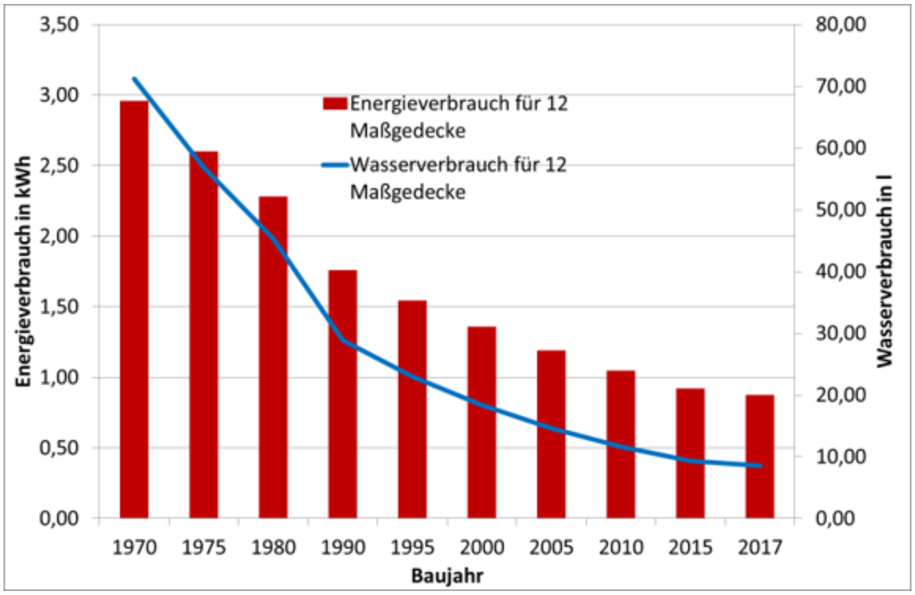

Figure 2.10: *Energy use in kWh (left) and water usage in l (right) over year of production for average of tested dishwashers coming to market [130]*

A more detailed study was conducted on energy usage of actual aged washing machines. Besides less well developed programs, inefficient usage of detergent and material fatigue were mentioned as facotrs reducing quality of the wash over time. It was shown that for similar washing performance 20°C to 50°C higher washing temperature was needed for 15 and 30 year old machines respectively to reach the same washing quality as a new, modern machine. [154] In addition to machine optimisation, Stamminger suggests that over the lifetime of a dishwasher incrustations and other ageing factors like malfunctions also reduce its efficiency over time. The higher energy use of an older machine therefore is a combination of both its initial energy use when coming to market and reduction of efficiency over time. [130]. This suggests that also dishwasher age would be relevant to consider for energy efficiency, while bearing in mind that different programmes use a different amount of energy [148].

2.4.4 Impact of behaviour

Several pieces of research have shown that consumers could not only save time, but potentially water, energy and money as well by doing the dishes with a dishwasher instead of by hand. However, a comparison between washing by hand or with a dishwasher depends on assumptions made about behaviour. According to [150] the main discrepancies between real-life consumption per wash and laboratory testing (EN50242, 2003) are caused by the following three behavioural factors:

- Dishwasher loading
- Pre-treatment of the dishes
- Program choice

The study therefore included the manual pre-rinsing as well and looked at real-life used program choice and loading. Data was gathered based on a photo (of each dishwasher load) and survey diary (number of items, duration of the wash etc.) that 20 households in each of 4 different countries in Europe (Germany, Sweden, Italy, UK) kept to log all their daily dish washing activity for two weeks. An average energy consumption of 25Wh/item (50% interval of 20-29 Wh/item) for houses with a dishwasher was found compared to an average of 34Wh/item (20–50 Wh/item) for houses without. Despite this potential saving of 28% energy by using a dishwasher instead of manual washing, the researchers conclude that further energy could be saved by decreasing the washing temperature, spending less water on manual pre-rinsing and making more efficient use of the capacity. In terms of loading the research showed 20% of dishwasher baskets had more than 40% free space left. Unnecessary energy was spent on pre-rinsing (especially in Sweden and Italy), while only scraping off leftovers is recommended by dishwasher manufacturers. Finally, it was found that more than half of the households (52%) use a washing program above 65°C most of the time, with an average of 59°C. Hence, using higher amounts of energy than according to the label calculations.[150]

The assessed studies in this literature research share the common conclusion that a) a lack of consumer awareness may be a root-cause for inefficient behaviour [151] [150] and that b) the results should be transferred into transparent and easy information to raise consumer awareness [145]. A study by [9] on the saving potential of behavioural change in washing machine usage suggests that a reduction of up to 50% in energy and water consumption are possible, when optimised consumer behaviour is reached. The scenario analysis based on over 5000 consumer survey responses from the A.I.S.E. study across 23 European countries, mentions a combination of more efficient machines on the market (10-30%), better use of capacity (about 25-50%) and lower temperature washing programs (10-20%) as the main areas for improvement with ranges depending on regions in Europe for which these savings most apply. The previous section on real-life inefficiency suggests potential behavioural savings for the dishwasher might be possible as well, but how much these savings could be is yet to be tested.

2.5 Literature conclusion

After the literature analysis, the first sub-question can be answered. The second sub-question can be partially answered as a better understanding of NILM usage for appliance detection from the smart meter has been gained, which shows this could be further developed to move from NILM detection to consumption insights. In the methodology chapter the insights on dishwashers energy consumption and efficiency will then be combined with the NILM literature to extend knowledge into the domain of energy system analysis, by that fully answering the second sub-question.

1. **How is electric energy usage and efficiency for dishwashers characterised based on traditional analysis?**

   The load profile of a dishwasher follows a distinctive pattern, based on different steps part of the washing cycle, such as water heating and rinsing. The average energy consumption depends on factors such as chosen washing program, efficiency label and size resulting in an energy consumption of about 1.1 kWh per wash. Usage frequency depends on family size, amount of dishes used and capacity utilisation of the dishwasher. The average annual number of washes in the Eco Design regulation (No 1016/2010) is assumed to be 280. Energy consumption and usage patterns are both analysed by traditional approaches, such as surveys and smart plugs. Energy consumption characteristics as used in the Eco Design and Efficiency Labelling regulation are based on findings from traditional approaches and lab settings, but do not take into account factors affecting the efficiency in real-life usage impacted by user behaviour.

2. **How can a smart meter based NILM system be developed to detect energy consumption, usage pattern and efficiency characteristics for dishwashers?**

   The application of NILM for disaggregtion of household appliances follows a reoccuring framework that can be applied to automatically recognise appliance activity within unseen load profiles. The details of the chosen approach are dependant on the frequency of measurement and types of power variables measured. Lower frequency detection ($< 1/s$) allows for a more scalable non-intrusive approach, but cannot be developed as precise as high-frequency applications. Smart-meter data (commonly 10s) approaches have been focused on the detection of major appliances in one or only several (controlled) households. The NILM model developed by Basu et al. (2017) has successfully been implemented in an end-to-end solution to disaggregate several major appliances (including dishwashers) for 4 households in real-life usage. Features of this scalable detection system could be used to further develop a model for energy consumption and efficiency analysis.

3

Methodology

3.1 Methodology introduction

The methodology uses findings from the first and second research sub-questions to develop an approach to prepare for answering sub-questions 3, 4 and 5. First a general overview of the used data sets and main model steps is given in 3.1. Section 3.2 goes into more detail on how the model was developed to detect energy usage, usage frequency and efficiency for dishwashers from the smart-meter signal, developed and validated with a data set of smart plug data. A more detailed description of used algorithms and assessment methods can also be found in Appendix B. In 3.3 the statistical methods, aggregation and segmentation methods used to analyse the output from a large scale application of the model on one full year of data is described.

Answers the following sub-question:

2. How can a smart meter based NILM system be developed to detect energy usage and efficiency for dishwashers?

3.2 Method Overview

This section gives a general overview of the methodology developed to couple a meter based NILM appliance detection system with energy and efficiency analysis. This section starts with a summary of the approach, gives an overview of used data sources and a schematic overview of the model.

3.2.1 Summary of Approach

In this subsection, a schematic overview of the modelling phases, the components and steps is given.

As described in literature section 2.2, NILM could potentially be used as scalable and continuous approach for energy assessments since it can be used to disaggregate electricity data from the central smart meter without much intervention of the user. However, so far NILM has mostly been applied in controlled, small-scale settings and not to study energy consumption patterns and efficiency of real-life appliance usage.

In this research a model was developed in Python to calculate dishwasher energy usage and washing frequency and then classify efficient usage based on data from the central meter. The code is not available publicly, since the majority of code has been created in the Eneco environment and is proprietary. The developed system was deployed on a very large set of meter data to non-intrusively analyse energy consumption, usage pattern and efficiency characteristics of real-life dishwasher usage in households.

Stepwise approach

The development of this framework is based on the literature described in subsection 2.2.3, in particular the schematic overview of NILM system application 2.2. The stepwise approach of the research (presented in figure 3.1) builds on existing load identification research and shifts the focus towards the consumption insights, including consumption analysis and efficiency analysis. As shown in the first two rows it can be divided into the two phases and three components:

- **Phase 1: Model Development**

  Component (A) Smart plug analysis

  Literature section 2.3 describes the characteristics of dishwasher energy usage and section 2.4 defines what is efficient usage. To gain further insight into these characteristics of dishwasher load profiles and usage patterns in real-life, an analysis was carried out on plug data.

Component (B) Smart meter model development
The plug data was then used as benchmark to develop and validate a meter data based machine learning model, extending the NÌLM detection methodology developed by K. Basu (2017) (see literature subsection 2.2.3) with estimation of energy consumption per wash, weekly usage frequency and classification of relating efficiencies.

- **Phase 2: Model Deployment**

  Component (C) Smart meter model deployment
  The developed algorithms were applied on meter data to investigate energy consumption, usage pattern and efficiency characteristics non-intrusively at large scale.

For each of these research components the following three main analytic steps can be distinguished:

- **Step 1: Data preparation**
  Gathering and processing of different data sets.
- **Step 2: Consumption insights**
  Model development and analysis of energy consumption and usage frequency.
- **Step 3: Efficiency insights**
  Model development and analysis of energy efficiency.

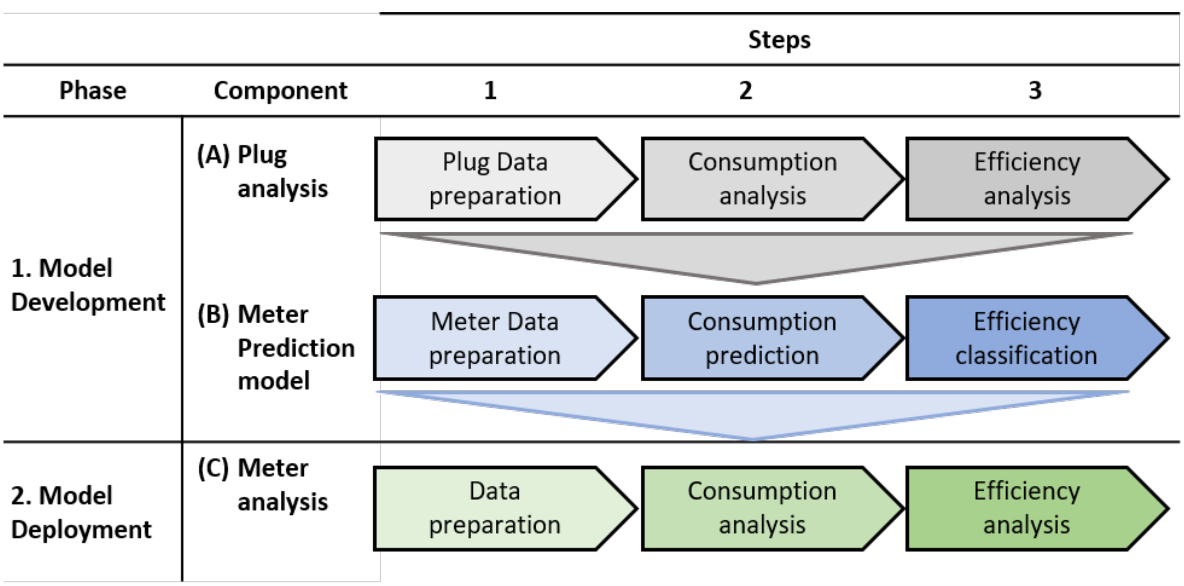

Figure 3.1: *Stepwise Approach*

3.2.2 Data

This subsection gives an overview of the used data sets and sources as well as their characteristics. Detailed explanation of the data sources, retrievement and processing can be found in the subsequent sections.

Table 3.1: *Overview of data sets*

	Data description	**Households**	**Timespan**	**Sample interval**
Development				
Meter data (learning set)	Household aggregated load profiles	100	11/16-03/17	10 sec
Plug data	Dishwasher load profiles	100	11/16-03/17	10 sec
Plug survey	Household size Machine characteristics Model type number Usage behaviour	41	01/16-02/17	
User manuals	Washing program characteristics	26	2000-2018	
Deployment				
Meter data (deployment set)	Household aggregated load profiles	129.137	01/18-12/18	10 sec
Meter survey	Household size Machine characteristics Washing program	10.873		
Various other	Weather measurements Public holiday dates		01/18-12/18	per day

For the two research phases, the following main data sets can be distinguished:

- **Development Phase (1)**
  Three months of dishwasher load profile (smart plug measured) data for 100 households. Household aggregated load profile (smart meter measured) data for the 100 households with plugs installed. A consumer survey (answered by part of the households) and user manuals retrievable online (for part of the households answering the survey).

- **Deployment Phase (2)**
  A full year of household load profile data for nearly 130.000 households to deploy

the model on and survey data for a share of these households. Various other data sets to compare/explain patterns. A survey was conducted under a sub-sample of nearly 11.000 households to segment the results for household and machine characteristics. Various other data sets such as outside temperature and vacation schedule were used to contextualise the usage patterns.

The data for this thesis has been gathered in line with the European GDPR regulation [155]. Data was only gathered in consent with the users and has been anonymised for the analysis.

3.2.3 Model overview

This subsection aims to describe the schematic model overview, relating the aforementioned stepwise approach and the data sources at more granular level. A more detailed description of the model development is found in section 3.3 meter estimation model. The plug analysis was done as a preparatory step to develop the estimation model and hence is completely integrated within the model development section. The methodology of deployment and analysis on large scale meter data (Phase 2 and Component C) is described in section 3.4.

The next page contains a schematic overview (figure (3.2)) of the analytical steps involved in development (left side) and deployment (right side). The colour coding of the steps relates to the colour coding previously also shown in the stepwise approach (figure 3.1). The squared elements in the figure represent analytical steps, the oval elements indicate model outputs. Dashed rectangles are drawn around a set of elements to indicate that the outputs/steps as a whole are prerequisite for another step or that they are feeding together into a set of elements.

Phase 1: Model development

The first set of symbols under the model development relates to the data sources as described in 3.2.2. As pointed out above for the model development **Component (A)** plug analysis and **Component (B)** meter estimation model are relevant. As described in 3.1 the first step in the modelling process for both components is the data preparation (colored in light grey and light blue respectively).

- **Step 1: Data Preparation**
  The data preparation is distinguished into two subsections: Data gathering and data processing. These steps are further described for the smart meter and the smart plug data in subsection 3.3.1. Data for the model development (table 3.1) is gathered into a cloud storage. A NILM dissaggregation algorithm developed by Basu et al (2017) [43] (described in section 2.2.3) is deployed on both meter and plug data. This is needed to detect dishwashers on the meter signal. While plugs

are appliance specific and hence already the profile is separately available, the algorithm is also applied on the plug data to retrieve similar detection features and compare the results. The outputs of this process step are the meter detection features and the plug detection features. Separately energy per wash is calculated for the plug data using area under the curve between start/stop times of individual washes. The output is the smart plug energy usage. The results are created as count of weekly washes for both plug and meter, features and energy per wash are clustered per household to be used for the regression.

- **Step 2: Consumption insights**
  Feature Analysis: The first step of the consumption estimation is the comparison of the features (feature analysis) retrieved from the plug and meter (see section 3.3.2 for more details). The feature analysis is then used for three subsequent steps:

  - Findings are used to better understand power profiles and inform improvements for the dissaggregation algorithm
  - Findings are used to determine the most relevant dishwasher characteristics to focus on for the further research.
  - Summary statistics are calculated for energy per wash and weekly washes
  - Additional consumption analysis on the plug, in which the plug energy usage and the appliance survey are analysed. This can be found in the Appendix C.

  Regression models: The second step of the consumption estimation is the regression model. For both estimations (energy per wash and usage frequency) multiple different regression approaches were tested. These were prepared and compared based on meter profile data input, using plug profiles as validation. The number of weekly washes were calculated based on detected washes on the meter, regressed on actual number of washes according to its related plug. The estimation of energy per wash was based on dishwasher's load profile characteristics (detection features) regressed on energy consumption according to its related plug.

- **Step 3: Efficiency classification/analysis**

  Efficiency threshold models: For this analysis efficiency thresholds were developed to serve as proxy for efficient behaviour for both energy per wash and number of weekly washes. To establish how real-life energy usage compares to the user manual and EU efficiency regulation efficiency, threshold for machine characteristics were identified that were based on equation(s) in EU efficiency label regulation. User manuals were compared with the plug measurements and with the EU regulation based energy efficiency thresholds. More details see section

3.4.3. For the usage frequency efficiency a threshold model was developed based on a bottom up frequency efficiency model, see section 3.3.3.

Efficiency classification: To develop and test whether estimated energy per wash and number of washes per week can be classified as efficient an efficiency classification method was developed. The classification of number of weekly washes and energy per wash as efficient was based on the comparison of plug measurements versus meter estimations compared to the efficiency threshold developed in the efficiency threshold models, see section 3.3.3.

Phase 2: Model deployment

The model deployment is depicted in the right side of figure 3.2. The data sources (see section 3.2.2) feed into two separate steps: the model development and the deployment. The framework follows the same main steps as in literature figure 2.2, but with the emphasis on detailing the consumption analysis and efficiency analysis and taking the NILM disaggregation as a preparation step.

- **Step 1: Data Preparation**
  To deploy the meter estimation model the developed algorithms were run over a full year of meter data of more than a hundred thousand households, see 3.2.2. The NILM diassaggregation algorithm developed by Basu et al (2017) [43] (described in section 2.2.3) is deployed on the meter data. The output of this process step is the meter detection features. The features are then aggregated and weekly washes counted for each household.

- **Step 2: Consumption analysis**
  To analyse dishwasher consumption patterns at large scale and relate to time, machine characteristics and consumer behaviour the summary statistic of total energy consumption per wash, per week and per year as well as frequency of usage per week and per year were retrieved. To analyse the time dependency a comparison for the frequency of usage per week of the year was made in relation to events such as holidays, the weekly usage pattern was analysed and the energy usage per wash was compared to the outside temperature. Energy usage and number of washes were segmented by survey results, including stated machine size, stated and detected number of weekly washes by family size and detected energy usage by stated washing temperature and machine efficiency label.

- **Step 3: Efficiency analysis**
  To analyse scale of inefficient dishwasher the usage efficiency classification based on weekly usage was compared to what to be expected based on family size. For the efficiency classification the estimated energy consumption was compared to expected consumption based on washing temperature and machine efficiency label. Potential energy savings were estimated.

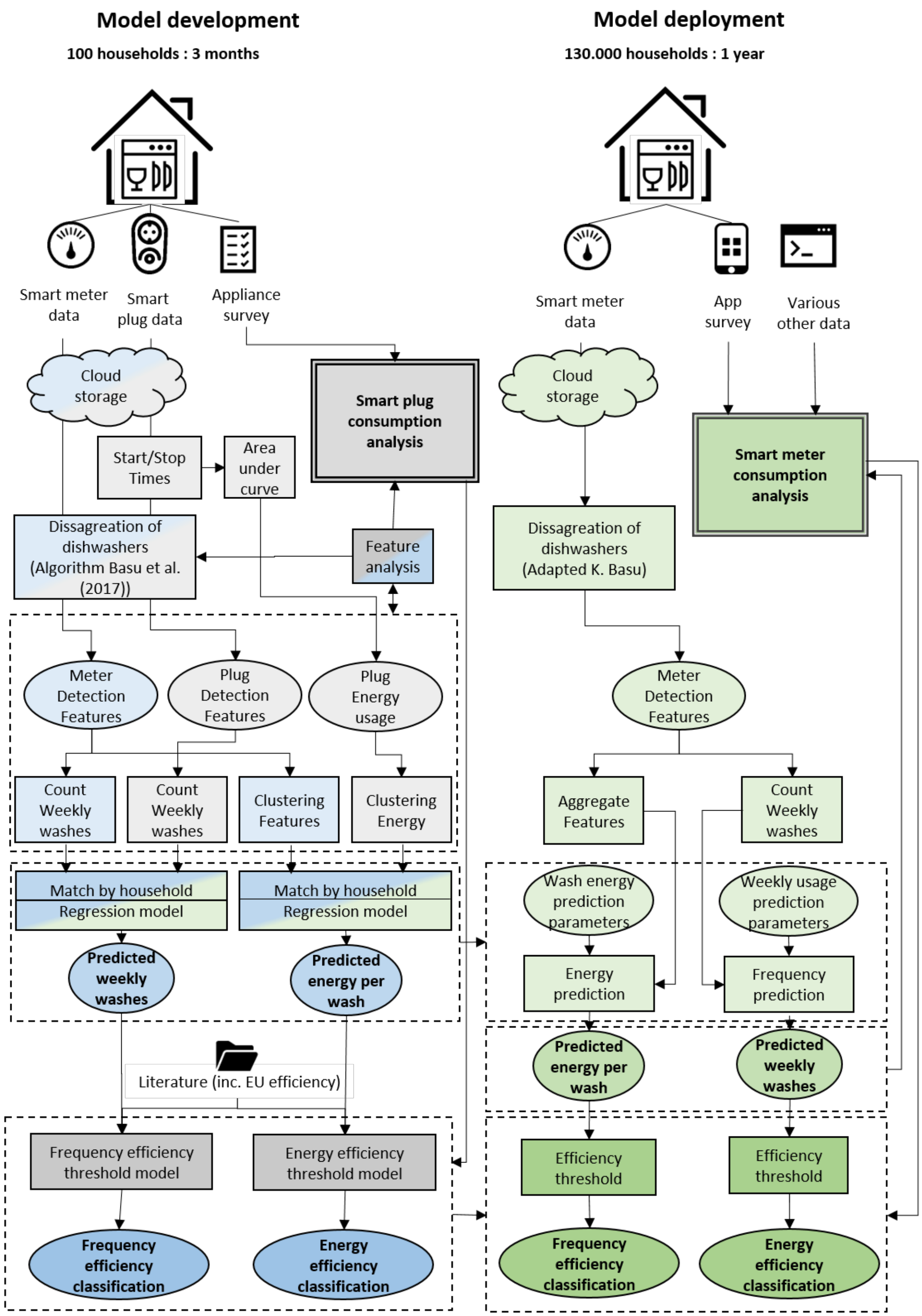

Figure 3.2: *Model overview*

3.2.4 Algorithms and validation

Besides the NILM system for which literature was presented in subsection 2.2.7, several common algorithms/approaches were applied in different stages of the model as presented (on the left, development side) in figure 3.2. This section gives a generic overview of these algorithms and why these were specifically selected. Furthermore, this section gives a brief description of how different steps in the model were validated. The methodology section on model development (section 3.3) continuous with how these different approaches have been applied as part of the analysis. The following information is only meant as a quick summary/refresher for readers less familiar with some of the terminology in subsequent chapters, Appendix B elaborates on these topics in more detail.

Algorithms

The three main algorithm groups of relevance for this research are clustering, regression and classification.

Clustering
Clustering is the aggregation (e.g. to a mean or mode) of multiple observations belonging to the same group/cluster [156]. This can be thought of as multiple washes of the same household. A comparison of different methods for clustering was made for the application on three problems encountered with the data handling: Identification of the most likely estimate (MLE) for nested design of the observations, facilitation of pair-matching and noise filtering. Further details on the three problems can be found in appendix section B.0.1.

To solve these problems, different clustering or aggregation methods can be used. A simple approach is to aggregate the data to a summary measure (e.g. mean or mode) [156] (equations see appendix B.0.1). In addition to aggregating to mean and mode, two clustering algorithms were tested. Using a clustering algorithm such as K-Means or Gaussian Mixture Model (GMM) is an unsupervised learning method, that provides the opportunity to recognise an on beforehand unknown number of clusters in the data, by that accounting for a potential multi-modal character of the data. Furthermore, it provides the opportunity to filter out some noise. [63] [157] However, using these algorithms also adds additional complexity and computational cost. Therefore their performance was compared to basic aggregation to a mean and mode value. As the GMM approach in early stage of analysis showed to be superior over the K-Means approach, only GMM was considered compared to mean and mode for final analysis as presented in this report. For more on the application of this method see subsubsection 3.3.2 and 3.3.2, for more detailed information, see appendix subsection B.0.1.

Regression
Regression is an often-used statistic tool in empirical research to describe the relationship between a dependent variable and one or multiple other variables. In its most basic form, it is a linear function described as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i \tag{3.1}$$

Where Y_i is the dependent variable representing observation i part of N observations, X_i is a corresponding covariate, and β_0 and β_1 are unknown regression coefficients representing the linear line's intercept parameter and the slope coefficients of the covariate, respectively. In this research the regression was not used to infer the relationship between measured variable Y and corresponding measured covariates X^n, but was used to estimate unknown values $\hat{Y}_i$ based on a set of n measured variable(s) X_i^n. [158] [159] For more on the application of this method see subsubsection 3.3.2 and 3.3.2, for more detailed information, see Appendix subsection B.0.1.

Classification
Classification was used to group the consumption analysis outcomes in classes of efficient and non-efficient. Several common machine learning classifiers, including Logistic Regression, Naive Bayes, K-Nearest Neighbour and Support Vector Machines (SVM) [160] where considered. However, it was decided not to deploy these. Instead thresholds were set to separate the results from the regression analysis into different classes. For more on the application of this method see subsubsection 3.3.3, for further detail on considerations see Appendix subsection B.0.1.

Validation

In order to develop the model, the full training set of 100 households was used and then the trained model was deployed on the large scale unseen meter data set. However, in order to understand how accurate different steps of the model are, the performance has to be tested. To validate the results the trained algorithm has to be deployed on a set of unseen (meter) data, where reference (plug) information is also available. To do this the available data can be split up in a training set and a validation set. However, since the available number of training households is sparse, the variability in the data can cause a random sampling of the training data and validation data to produce different results. Depending on the sampled training and validation set this can result in either over- or underestimation of the performance. A common practise is to cross-validate. The data is randomly split in a test and validation set many times and the algorithm's performance stored for each iteration. The comparison of the algorithms was done with 1000 iterations of a 50/50 split cross validation. [123]

In this research three types of results have to be assessed: clustering, regression and classification results. Since the clustering is used as a preparatory step for the regression, the outcomes of the clustering and regression analysis can be assessed together based on the error of estimation. The classification quality can be based on the ratio of true and false classifications.

Error of estimation
While several metrics could be used in order to do this, the Root Mean Squared Error (RMSE) provides a common often used metric to compare different models:

$$RMSE = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(\bar{Y} - Y_i)^2} \tag{3.2}$$

However, as the RMSE gives results in absolute values for the unit of analysis (e.g. kWh) in order to compare models across different units it helps to normalize the results (e.g. to %). This can for instance be done by dividing them by the mean value, resulting in a Normalized Root Mean Squared Error (NRMSE):

$$NRMSE = \frac{RMSE}{\bar{Y}} * 100\% \tag{3.3}$$

The NRMSE (Normalized Root Mean Squared Error) served as the comparison metric. The NRMSE output of the 1000 iterations cross-validation, were aggregated to an average with standard deviation. The best performing algorithms are selected based on a low average NRMSE and small 95% interval. A selection of algorithms with the lowest NRMSE were further compared based on a visual inspection of their performance (further details see appendix B.0.2.

Accuracy of classification
Classification accuracy is relevant both for the outcomes of the NILM detection system and for the efficiency classification. While the detection system classifies if it found a dishwasher at some point in time, this means that each other point in time it basically classifies no dishwasher is active. When a washing activity is detected correctly this is a true positive and when misdetected a false positive. When a dishwasher actually was active, but not detected this is a missed detection or false negative, all moments no dishwasher was active and not detected are true positives. [123]

Likewise for the efficiency classification, when dishwasher usage is classified as inefficient both on the plugs and the meter, this is a true positive. When dishwasher usage would be classified inefficient based on the plug data, but gets classified as efficient based on the meter data, this is false positive. Efficient on the plugs and the meter is a true negative and efficient based on the plugs but not the meter is a false negative. [123]

The accuracy of the classification depends on the balance between these outcome, according to the following equation:

$$Accuracy = \frac{True\,predicted}{Total\,predicted} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.4}$$

However, a more specific way of representing the performance of the classification is according to the following equations:

$$Precision = \frac{TP}{TP + FP} \tag{3.5}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.6}$$

$$F1 = 2 * \frac{precision * recall}{precision + recall} \tag{3.7}$$

The precision resembles how well the algorithm performs at not misdetecting (ratio of false positives to true positives), while the recall resembles how well the algorithm performs at not missing detections (ratio of false negatives to true positives). The F1 score is an accuracy measure weighted for both precision and recall. Ideally the F1 would be as close to 100% as possible, meaning both precision and recall are close to 100%, which would be perfect accuracy. However, realistically measurements diverge from this ideal state. The classification can be optimised either for high precision or high recall but not both at the same time. It's a trade-off between either being precise about what is detected at the cost of missing some detections (higher precision) or ensuring most detections are made at the cost of some being false (higher recall). [62] [123]

To test how different values of the threshold affect the rate of false positives vs true positives an ROC analysis was conducted. The threshold is varied in gradual steps and the classifications are recalculated for each step. Based on this, a precision, recall and F1 can be derived for each step, respectively. The ROC outcomes are then plotted on a curve vs. a line that would represent a random guess. The larger the area between the ROC curve and the random guess line the better the classification algorithm performs overall. The optimal precision, recall and F1 points are on the different inflection points along this curve. [123]

Based on these outcomes a general efficiency threshold can be set that aligns with the goal this classification mechanism might have in application. Which could either be to reduce false positives or false negatives.

3.3 Smart Meter Model Development

This section describes in more detail how a model was developed to detect energy usage, usage frequency and efficiency for dishwashers from the smart-meter signal, building on an existing appliance detection system. The model uses smart meter data as input. Smart plug data was used to calibrate and validate the model in the development stage. As both smart plug and smart meter data are needed, first the data gathering and processing is described for both. The appliances detection developed prior to this research by Basu et al. (literature section 2.2.3) is referenced how this was applied in the data processing. Next the developed model to estimate usage frequency and energy consumption per wash are described. Finally, the approach to classify energy efficiency is described.

3.3.1 Data Preparation

The two main data categories to develop the Smart Meter Model are smart meter data and smart plug data. This subsection describes the preparation of these two main data sets.

Smart meter data

The smart meter data consists of the complete aggregated load profile of the household. The first aim is to gather the smart meter data and disaggregate dishwasher activity from the complete smart meter profile.

Data gathering
Dutch households with a smart thermostat from Eneco (called Toon) can receive real-time insight in their power consumption profile. Data can be analysed for all users that give permission for their data to be used for research and who have the necessary software installed. Data is sampled from the smart meter's P1 port with 10 second intervals, but can also be gathered from analogue meters, where a laser counter gets installed when placing the Toon.

The data set consists of continuous time series for each single household. The continuous power measurements are uploaded to a cloud database and stored in watts with a 10 seconds interval by date and time stamp and labelled with an anonymised household code. Figure 3.3 shows an example of such a time series of power measurement for one household at the end of 2016.

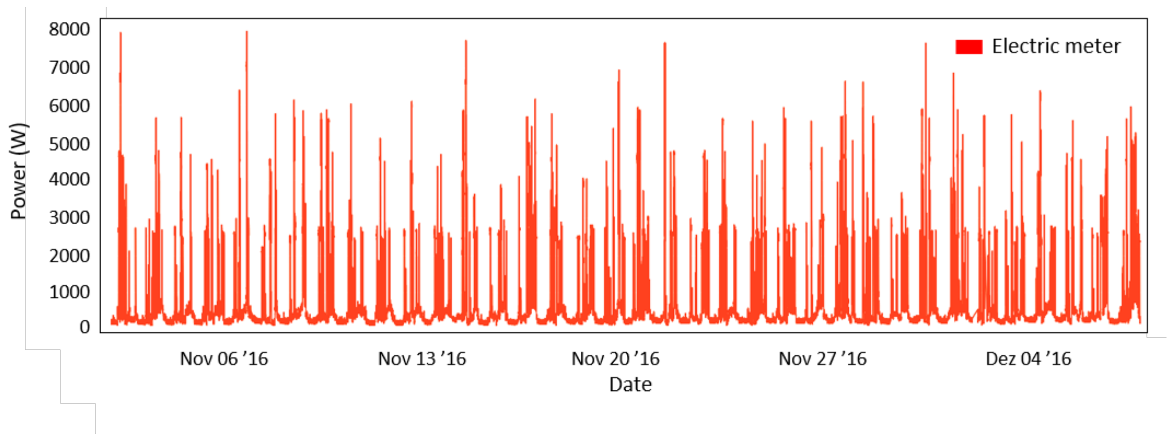

Figure 3.3: *Smart meter power profile of one household in November 2016*

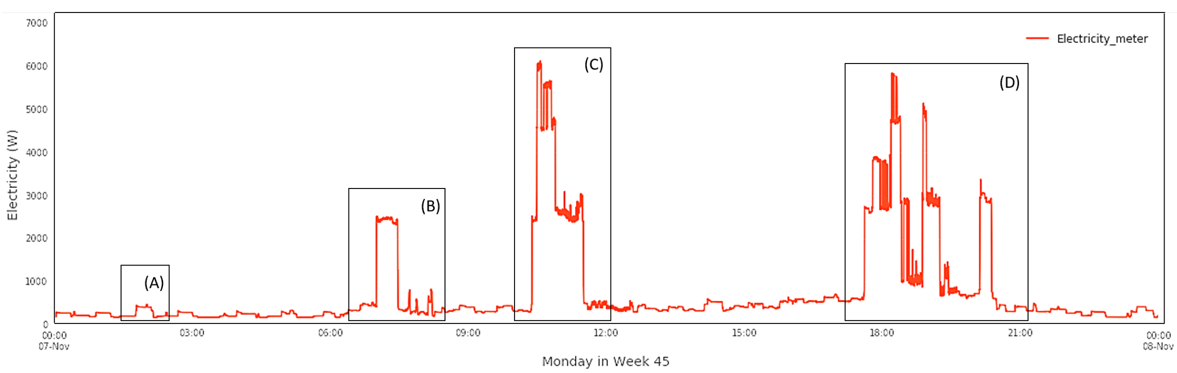

Figure 3.4: *One day smart meter power profile of Monday week 45, 2016, with patterns of:*
(A) Fridge/freezer (repeated throughout the day)
(B) Washing machine
(C) Tumble dryer and washing machine
(D) Tumble dryer, washing machine and dishwasher

Data processing
The methodology of Basu et al. (2017), described in literature subsection 2.2.3, was used to detect washes from the smart meter signal. The detections provided by the algorithm of Basu et al. including the anonymised household code and a time stamp were stored with the feature values related to the washing cycle of a dishwasher. These detections provided by K.Basu were queried with SQL from the central databse and used for further development of the Smart Meter Model in Python, as described in the following sections.

Smart plug data

The smart plug data consists of the load profile measurements of the specific appliance it is connected to. In order to train the model to disaggregate dishwasher activity from the smart meter signal and validate if detections are made correctly the smart plug data can be used. An additional survey was conducted to gain more insight in characteristics of the sample and type numbers of the dishwashers (see Appendix C.0.1)

Data gathering
Eneco customers can place smart plugs in the socket of a device of interest and connect with their Toon to gain insight in the power profile of the individual appliance. As part of a demand side management trial by Eneco, a group of 130 households were sent smart plugs for washing machine, dryer and dishwasher. Participants were requested to install the plugs before November 1st, 2016 and not allowed to disconnect before May 1st, 2017.

The data consists of 10 second power measurements, similar to the smart meter data. The main difference is that only one single appliance is measured. The data was labelled with the same anonymised household code matching as the smart meter data, with the added label for appliance type as provided by the participants.

The number of households from which dishwasher data could be retrieved reduced down to 100 households. Reasons ranged from non-functioning plugs, connection issues, people taking plugs out (in between) prior to the requested date to participants not being able to install plugs, for instance because their dishwasher is built-in, making it hard to reach.

Because several users did not directly install their plugs in the first weeks and the longer the project progressed the higher the chance that participants had taken the plugs out to test on other devices, a selection of three months was made. Three months was chosen as an optimal period to strike a balance between enough data points to train the model, but short enough to be able to provide insight within a time frame of less than a 100 days. A selection of the data for the period December 2016 until March 2017 (winter 2016/17) was chosen and prepared for analysis.

Data processing

- *Detection*
  To prepare the smart plug data for the consumption analysis, individual washes have to be detected in the continuous plug power profile. To detect a wash the start and stop time of the individual wash have to be recognised.

  To detect the start time the moment of first instance of power consumption has to be recognised. The only power consumption detected by a smart plug should be the consumption of the appliance that it is connected to. Therefore, the energy consumption should be zero if the appliance is not in use. However, in practice it was observed that small malfunctions can occur. Furthermore, standby functions, such as start delay consume several watts, when in use. Consequently the starting moment of 30 consecutive minutes of more then 10 watts of power consumption was taken as a start of a new consumption cycle of the appliance.

To detect the stop time, the last moment of power consumption has to be recognised. Since it is not uncommon that an appliance makes short intermediate stops within a program, a stop of more then five consecutive minutes with power consumption below 10 watts was taken as the end point of a program. As it could happen that two consecutive dish washing cycles would occur, the risk arises that two consecutive cycles would be counted as one. However, it was assumed that it would take at least five minutes between two consecutive cycles, as it takes several minutes to get the cleaned dishes out and refill the dishwasher, particularly as dishes are still hot after a washing cycle. While some situations are thinkable where a new wash is directly started, for instance because the user does not like the cleaning performance and/or the soap dispenser did not open up the first time and the user might restart a new (different) washing program, these cases would stand out as a wash with very long duration and high energy usage and could be filtered out based on that. Figure 3.5 shows a cutout of a single dishwasher wash.

Figure 3.6 shows there can be some variation to this profile with regards to factors such as occurrence and duration of different phases. Profile 1 shows a standard pattern for a program with two heating/drying phases and an energy consumption of 1.21 kWh. Dishwasher profile 2 shows a comparable pattern, but a brief pause of the heating, resulting in a frequency count of 3 heating moments, with calculated energy consumption of 1.24 kWh only slightly higher. Dishwasher profile 3 shows a program with three heating/drying moments and an energy consumption of 1.58 kWh. Dishwasher profile 4 shows a program with four heating/drying cycles, but only an energy consumption of 1.32 kWh as the total wash and heating cycles are considerably shorter.

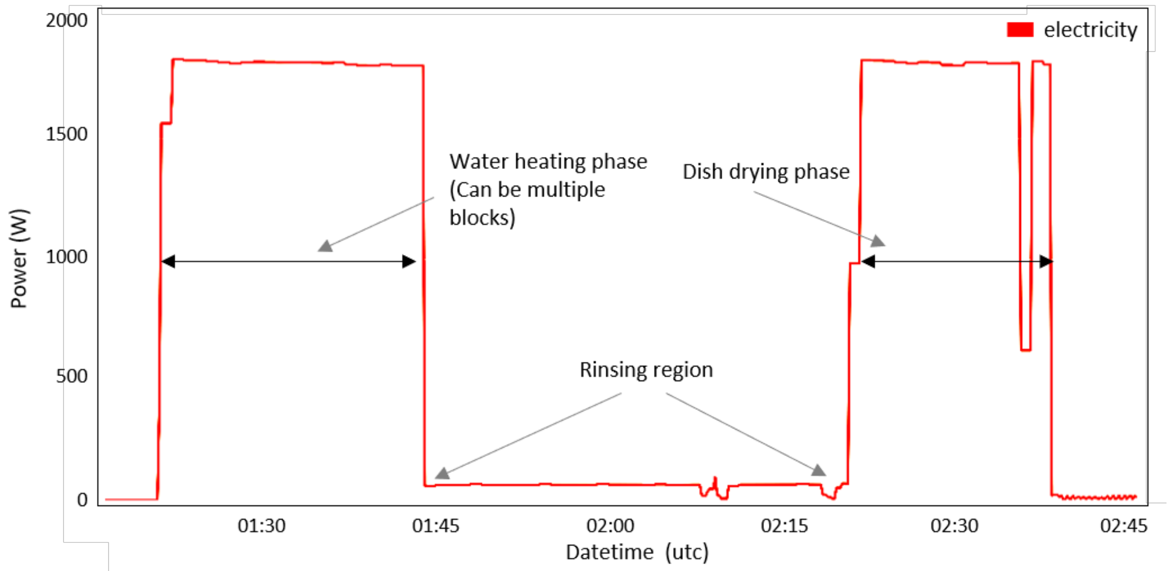

Figure 3.5: *Plug extracted load profile of dishwasher showing two distinct heating phases and rinsing activity in between*

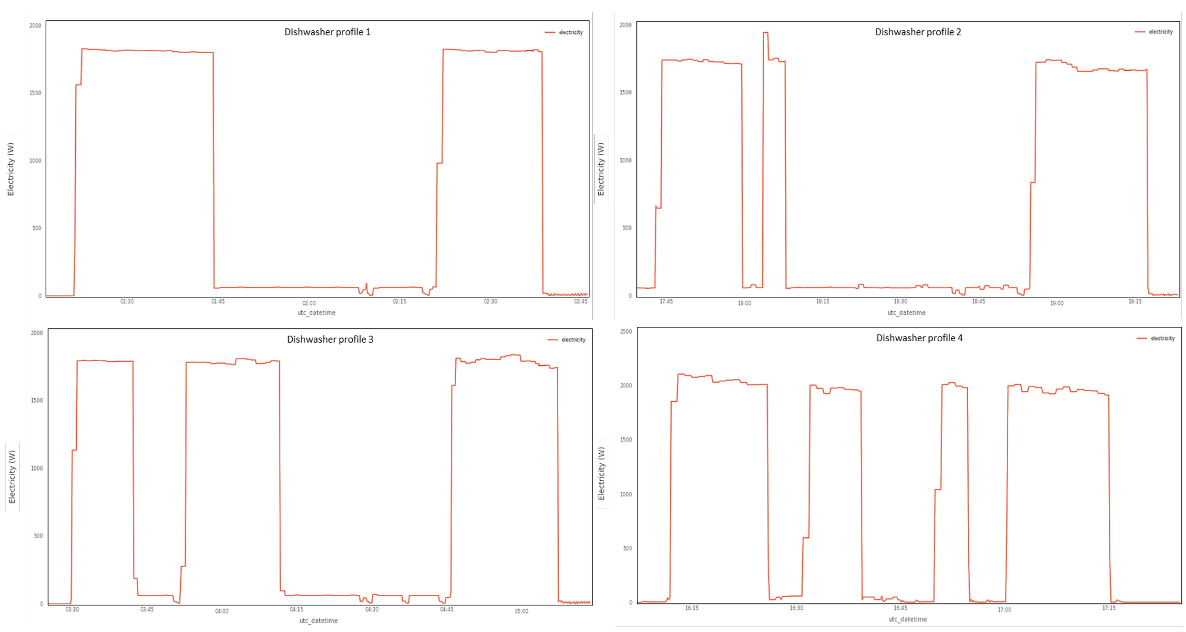

Figure 3.6: *Variation in dishwasher load profiles (exemplary, other variations possible)*

- *Feature extraction*
  The same feature extraction method from the algorithm of Base et al. to detect features of the dishwasher within the meter signal (see 3.3.1), was used to

detect different components of the washing program from the plug data as well. This provides the opportunity to analyse the washing profiles in more detail and compare plug and meter results on a component level instead of on the full wash only.

- *Usage frequency*
  For usage frequency analysis the data per household was split in calendar weeks starting Monday morning and ending Sunday night. The detected washes per week were aggregated to weekly washes, accounting both the total number of washes per household in a week and the average energy consumption per wash.

3.3.2 Consumption estimation

Following the detection of the dishwasher, the aim of this step is to analyse how often it is used and how much energy it uses per wash. Different algorithms to estimate number of washes per week and the energy per wash were tested and compared. The chosen algorithms and tested variations are discussed below. For more detail see Appendix B.0.1.

Energy per wash estimation

As multiple appliances are usually active at once within a real-life household, the energy of a wash can not be measured by the area under the curve for a smart meter detection, such as was done for the smart plug. Therefore, to approximate the energy consumption for a wash detected on the meter, another estimation algorithm was developed. The energy estimation method uses the features generated in the detection process by the method of Basu et al. to estimate the energy consumption of the detected wash. In order to do this, the potential features were analysed first, to determine the most suitable.

Feature analysis
The following output variables from the detection algorithm were considered potentially relevant for estimation of the energy per wash:

- **Wash duration:** the length (duration) of the total wash in minutes between detected start and stop time
- **Heating duration:** the length (duration) in minutes of a single detected heating moment
- **Heating power:** the height (power) in Watts of a single detected heating moment
- **Heating moments:** The total number of detected heating moments within the washing cycle

- **Heating energy:** The different heating moments within one detected washing cycle were aggregated and multiplied with the heating power to calculate the heating energy of the wash in kWh.

How close the energy estimation method comes to approximating the actual energy consumption relies on two aspects. The precision with which the individual features are detected and how much they individually influence the total consumption. To assess how different features affect the energy consumption, the plug calculated energy consumption was matched with both the plug detection and meter detection. This can also give a first indication on how well a feature is detected on the meter. To further assess how well the meter detects the different features, the distribution for both meter and plug and the correlation between meter and plug are compared for each feature of interest. See the Feature analysis in Results subsection 4.2.1

Estimation
The energy estimation aims to calculate the amount of energy consumed by a wash based on the features detected by the meter. This is done using a similar type of feature regression approach as described for washing machines in literature section 2.3.2. In order to train the regression model the energy of a wash according to the plug measurements has to be regressed with the meter detection features. While plugs do detect all washes, not all washes that actually happened are detected on the meter (false negatives) and not all washes detected by the meter, actually happened (false positives). Therefore the energy consumption measured with the plugs and meter detection variables cannot be pair-matched for every individual wash. When pair-matching single washes, only the true positives would match. All false positives and false negatives would get discarded, which would result in bias in the training data. This would show low detection error, but would actually perform poorly on unseen data. Therefore, instead of training on individual washes, only aggregated plug and meter values are investigated, resulting in one single value per household.

Per household i the energy consumption per wash, calculated with area under the curve for plug detections ($Energy^i_{plug}$) is first regressed on the meter detection feature set ($X^i_{n,meter}$) for n included features.

$$Energy^i_{plug} = \beta_0 + \beta_1 X^i_{1,meter} + ... + \beta_n X^i_{n,meter} + \mu^i \tag{3.8}$$

The measurement error of the detection algorithm can be divided into two aspects: a bias or structural error and a random error. In comparison to the random error, which can be thought of as noise and works in both directions, a structural error can be caused by the design of the detection algorithm or bias in the underlying data that was used to develop it.

As the first goal is to recognise as many correct appliances as possible, sometimes it can be more beneficial to over- or underestimate certain features to decrease the risk of misrecognition. It can also be that this over- or underestimation is unintentional and could be improved upon. In either way, intended or unintended structural under- or overestimation of features get accounted for by the learned β values. The random detection error for every feature are accounted for in the total error term μ^i. The regression parameters are then used to estimate $\widehat{Energy}^i_{meter}$ based on new, unseen meter detection feature sets ($X^i_{n,meter}$), as shown in the following estimation equation:

$$\widehat{Energy}^i_{meter} = \beta_0 + \beta_1 X^i_{1,meter} + ... + \beta_n X^i_{n,meter} + \mu^i \tag{3.9}$$

To develop the estimation model, a variety of combinations of approaches were tested and compared for the different features. For more information on the different regression and aggregation approaches, including clustering methods such as GMM, see Appendix B.0.1. The following approaches were considered:

- **Regression approach**
  Two approaches to the regression were tested. One is to fit the regression based on the training data (linear regression). The other is to use a heuristic for the parameter values instead such as the average share of heating energy compared to the total energy (heuristic approach).

- **Aggregation**
  The energy consumption per wash over a period of multiple weeks was expressed by both the aggregation to the average and aggregation to the mode. In addition, clustering based on a Gaussian Mixture Model (GMM) was applied (see appendix subsection B.0.1). This was done both to test the ability to distinguish different washing programs and to potentially filter out misdetections.

The performance of the best performing algorithm is presented in Results subsection 4.2.1. For more on performance comparison see Appendix B.0.2.

Usage frequency estimation

The smart meter detection algorithm, as developed by Basu et al., was shown to have a general detection accuracy of over 80% (ratio of false positives and false negatives are not provided). Regardless if this number could go up with more and better data and changes to the algorithm, a share of misdetections will stay inevitable. Depending on whether the sensitivity of the algorithm is set to penalise false positives or false negatives more, this will result in over detection or under detection of the weekly washes by the meter compared to the actual number of washes (as measured by the smart plugs). To estimate the number of weekly washes more closely, an estimation approach was added.

Estimation

The usage frequency estimation was developed to calculate the actual number of weekly washes for each week based on the occurrence of washes detected by the smart meter detection algorithm. To teach the algorithm to estimate this, the number of weekly washes detected on the smart meter profile were first regressed on the actual number of washes according to its related smart plug (see B.0.1 for more on regression). The number of weekly washes according to the plugs ($Washes^{i,j}_{plug}$) for household i in week j is given by the number of washes detected on the meter for that household in that week ($Washes^{i,j}_{meter}$). This is regressed for regression parameters β_0 and β_1 with an error $\mu^{i,j}$. The following equation shows the regression of detected washes on the actual number of washes according to the plugs.

$$Washes^{i,j}_{plug} = \beta_0 + \beta_1 Washes^{i,j}_{meter} + \mu^{i,j} \tag{3.10}$$

The regression parameters are then used to estimate $\widehat{Washes}^{i,j}_{meter}$ based on new, unseen $Washes^{i,j}_{meter}$, as shown in the following estimation equation:

$$\widehat{Washes}^{i,j}_{meter} = \beta_0 + \beta_1 Washes^{i,j}_{meter} \tag{3.11}$$

There are several variations of these regression-estimation steps possible. For more information on the different regression and aggregation approaches see Appendix B.0.1. The performance of the following variations was compared:

- **Regression approach**
  The generic linear regression as shown above was compared with a force zero linear regression, where the β_0 is set to zero. A generic linear regression results in the lowest total estimation error. However, with this method $\beta_0 > 0$, hence zero weekly washes falls outside the estimation domain. The force zero fit therefore can perform better on the lower numbers of weekly washes and reduces the risk of estimating the occurrence of washes, while the owners were actually not at home that week.

- **Aggregation**
  Depending on the intended message both the mode and the average weekly washes can be relevant. Therefore the number of weekly washes over a period of multiple weeks was expressed by both the aggregation to the average and aggregation to the mode.

- **Aggregation Order**
  The aggregation order can matter as both have their trade-offs. First aggregating of the weekly washes and then regression, reduces the noise, but also the available data points as each household reduces to a single, similar weighted point. Vice

versa when first regressing and then aggregating, more information, but including noise is maintained.

The performance of the best performing algorithm is presented in Results subsection 4.2.2. For more on performance comparison see Appendix B.0.2.

3.3.3 Efficiency classification

After the estimation of the energy per wash and the number of weekly washes, the final step is to be able to classify if a dishwasher is used efficiently. In order to do so, focus points for dishwasher efficiency first had to be identified and are described in the following subsection. Based on that both energy per wash and weekly washes were identified as the main overarching focus points and a model to serve as proxy for efficiency was developed for both, described in the subsections hereafter.

Dishwasher efficiency

When defining clean dishes over a specified period of time as the desired output for dishwasher usage, (in accordance with the definition of energy efficiency in section 2.4.1) energy efficiency for dishwasher usage would be achieved by reducing the total energy consumed by the dishwasher over that same period, without sacrificing the desired output of clean dishes. The components affecting real-life dishwasher efficiency were grouped (see figure 3.7).

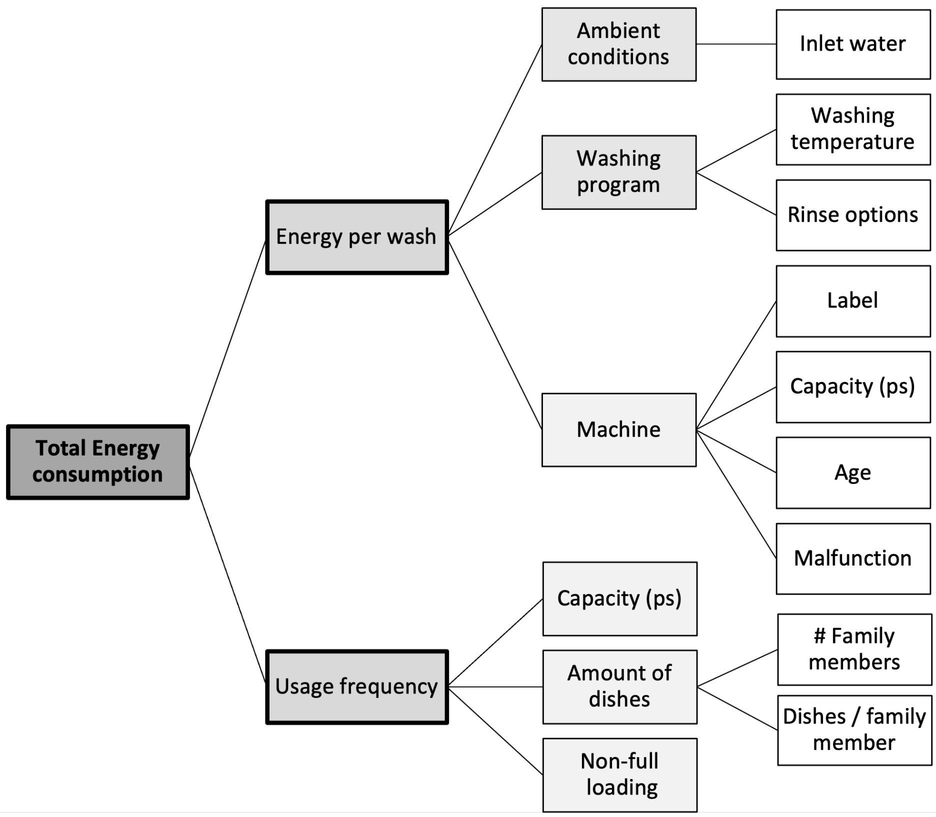

Figure 3.7: *Factors influencing energy efficiency for dishwashers. Structured through consultation with a field expert [130]*

As described in 2.2.1 the total energy consumption depends on the average power usage and chance of operation at that specific moment in time. In literature section 2.3.1 factors related to dishwasher energy consumption were discussed. Besides owning a dishwasher (share of penetration) the two most important identified factors contributing to the total energy consumption for a defined period such as a week or year are the energy per wash and the total number of washes over a specified time period, referred to as usage frequency.

Then individual sub-factors (section 2.4.3) affecting real-life efficiency, where grouped depending on how they contribute to energy per wash and/or usage frequency. The three important sub-factors affecting energy per wash are the machine characteristics, chosen washing program and ambient conditions. The usage frequency is mainly based on usage of the machine capacity, the household size and dishes produced by family members. The machine capacity is the only parameter affecting both. According to the subsection on the impact of behaviour (2.4.4), usage behaviour is the most important cause for inefficiency. Behaviour relates to the chosen washing program, amount of dishes produced and loading capacity utilised.

Classification threshold
For both energy consumption usage frequency efficiency thresholds were defined. The outcomes of the meter estimation were then measured against the set threshold to classify if the estimated outcome can be categorised as efficient usage. This meter classification is then validated with the classification according to the smart plugs (see subsection 3.2.4 for more on classification performance analysis). The following general thresholds were used as proxy to classify inefficient washing behaviour:

- Energy usage per wash: A general efficiency target of 1.05kWh per wash was set.
- Usage frequency: A general efficiency target of 4 washes per week set.

Note: These generic thresholds are also varied based on household and machine characteristic see section 3.4.3.

The following two subsections describe the developed methods to define these efficiency threshold for energy per wash and usage frequency.

Energy per wash efficiency

The EU energy efficiency label calculations, described in theory section 2.4.2, set a framework for efficient energy usage for a dishwasher. However, efficiency is defined as an index number for annual energy consumption. To calculate an efficient energy consumption threshold (in kWh) on a per wash basis, the equations from the EU efficiency label have to be slightly rewritten. For a standard wash, defined as the Normal (50°C) program, the following equation for annual energy consumption of the household dishwasher (AE_c) was used as starting point:

$$AE_c = Et * AAW + X \tag{3.12}$$

With Et = energy consumption of a standard wash [kWh], AAW, the average annual number of washes (280) and X some small additional standby consumption. When assuming X = 0, the energy per wash can be rewritten to:

$$Et = AE_c/AAW \tag{3.13}$$

The Energy Efficiency Index (EEI) in the efficiency label calculations is calculated as the ratio between annual energy consumption of the dishwasher (AE_c), compared to a standard dishwasher (SAE_c) :

$$EEI = AE_c/SAE_c * 100 \tag{3.14}$$

Rewriting this to AE_c gives:

$$AE_c = EEI * SAE_c * 100 \tag{3.15}$$

Substituting that into equation 3.12, we get:

$$Et = (EEI * SAE_c/100)/AAW \tag{3.16}$$

With the standard annual energy consumption of the household dishwasher (SAE_c) depending on machine size in number of place settings (ps), either:

1. for household dishwashers with rated capacity $ps \geq 10$ and width $> 50cm$:

$$SAE_c = 7,0 * ps + 378 \tag{3.17}$$

2. for household dishwashers with rated capacity $ps < 9$ and household dishwashers with rated capacity $9 \leq ps \leq 11$ and width $\leq 50cm$

$$SAE_c = 25,2 * ps + 126 \tag{3.18}$$

This shows that (Et), the energy consumption of a standard wash, is dependent on the dishwasher's energy efficiency label (EEI), the machine size (ps) and the assumption of 280 washes per year (AAW). Hence, when rated capacity and energy efficiency for a machine are known, maximum energy usage of the standard (50°C) wash can be calculated and used as an individual threshold for energy efficiency of that dishwasher.

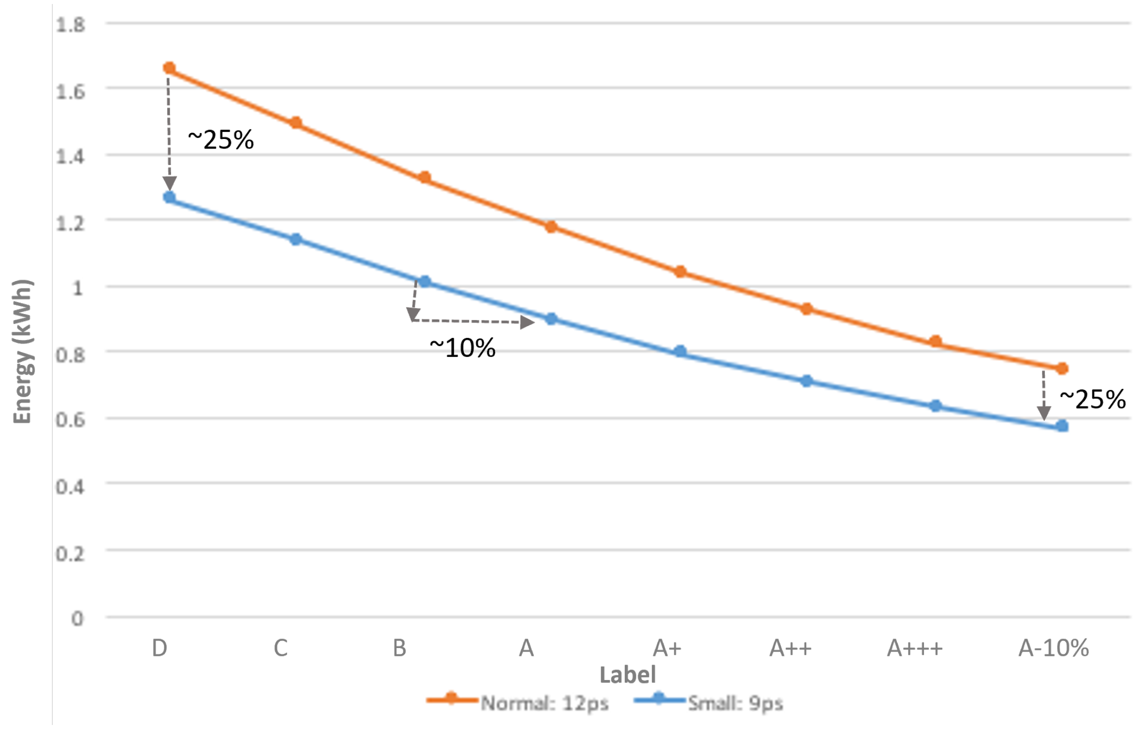

Figure 3.8: *Energy consumption calculated for different efficiency labels, based on EU energy label equations [140]*

The energy consumption for different efficiency labels and machines sizes was calculated. For an overview see appendix A.0.2. In figure 3.8 the results are plotted for a 12ps normal and 9ps compact machine for different efficiency labels. As can be seen the energy consumption decreases with about 10% per step in label efficiency (as earlier discussed when presenting table **??**e literature section). Furthermore, the difference between the compact (9ps) and normal (12ps) is about 25%.

In the next figure, the energy consumption for different sized dishwashers is calculated for an A+ label, shown in 3.9. For the smaller (sized width $\leq$ 50cm) machines the amount of couverts (ps) shows a stronger linear increase, increasing 0.3kWh over 5 steps. The normal (width $>$ 50cm) machines are not as much influenced by number of couverts (ps), increasing less than 0.1kWh over the same number of steps.

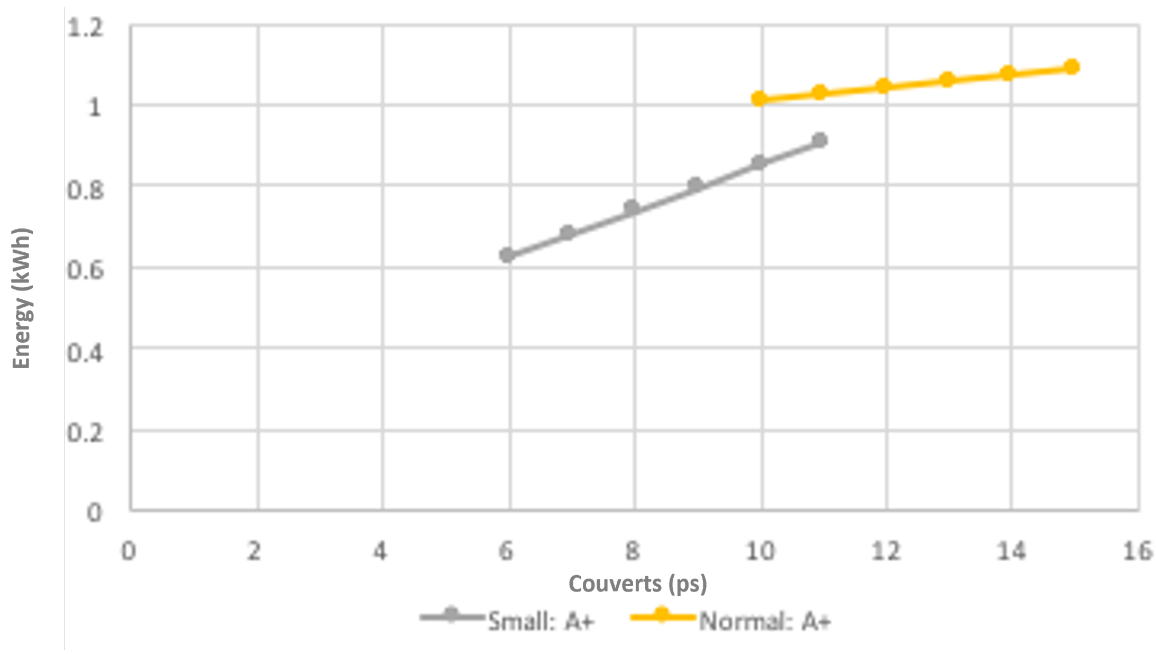

Figure 3.9: *Extrapolated energy consumption for different rated capacity (ps), based EU energy label calculation [140].*

To establish a standard threshold to serve as proxy for efficiency, the following characteristics were chosen based on consultation with a field expert (R.Stamminger, 2017). Most people have normal sized dishwasher (60cm), with 12 or 13 place settings. A+ with eco program is stated as efficient in the Netherlands by Milieucentraal. According to the label equations this corresponds with 1.05 kWh. The latest machines on the market were then also confirmed with an analysis of 500 available models on Dutch consumer website Consumentenbond, see appendix A.0.3, showing the latest models are often 12 to 14 ps and A+ or A++.

Therefore, as an initial generic threshold the value of 1.05kWh was used as proxy boundary for efficient energy consumption. The energy consumption per wash estimated by the energy estimation model (described in subsection 3.3.2) can then be compared with this established threshold. This can be used to classify if a household consumes more or less energy per wash, than the efficiency threshold. For the performance of this efficiency classification see Results section 4.2.3.

Usage frequency efficiency

In the EU efficiency regulation no statements are made concerning what constitutes efficient usage frequency. Only an assumed number of standard annual washing cycles of 280 washes per year is used within the calculations (see 2.4.2). The usage frequency efficiency analysis aims at developing an understanding for the average and potential optimal usage frequency of a dishwasher, when taking consumer and machine characteristics (e.g. family size, machine size) into account.

In order to do this, two different approaches were considered. The first would be to simply calculate the average weekly washes and take this as a threshold, or slightly more sophisticated, make user groups, based on family and/or machine size, and calculate the average for each group. Values above the average are interpreted as inefficient, values below are interpreted as efficient. While this could provide a social comparison between users, it provides no further information on whether this average is actually efficient. Furthermore, as most households in the sample group have similar sized machines, this would be unsuitable for further empirical distinctions.

Therefore, a second approach was considered. This approach is a bottom up estimation of what number of weekly washes could be seen as efficient usage, based on consumer and machine characteristics as well as behavioural assumptions. The bottom up approach was based on the household size and machine size:

H: Household size (people)

ps: Rated capacity and number of place settings

In addition certain assumptions have to be made about the amount of dishes produced per person per meal:

TS: Number of table setting items per serving cycle per person were set to 11, according to EU standard EN 50242 (see Table 2.3)

AS: Number of additional large serving items per six serving cycles were assumed to be 7 according to EU standard EN 50242 (see Table 2.3)

sf: Size factor for large servings was assumed to be 3 to account for the relative difference in size for a standard couvert item and an additional item

Instead of a step up, adding a large amount of additional items for each 6ps, the additional number was simply divided by 6. The number of table settings per dishwasher size would then be calculated as:

$$Capacity = ps * TS + AS * sf/6 * ps \tag{3.19}$$

This means the rated machine capacity (ps) is based on 11 items per place setting (ps) and a large additional setting per 6 place settings. Table 3.2 shows the calculated number of items assumed to fit in a dishwasher of different sizes:

Table 3.2: *Table settings per dishwasher type*

	Place settings		Items	
Type of dishwasher	**Min**	**Max**	**Min**	**Max**
Table dishwasher	4	6	58	87
Dishwasher (45 cm)	8	10	116	145
Dishwasher (60 cm)	12	16	174	232

Assuming 11 items per place setting and a large additional setting per 6 place settings

To calculate how many dishes are produced within a household of size H, assumptions have to be made about the amount of servings per person per day and week:

D: Servings per day per person were assumed to be 3.

W: Number of days per week eating at home were assumed to be 6 (sensitivity)

On first sight 3 servings per day might seem a bit high, assuming 3 main meals per day and people potentially also not being at home the whole day, particularly not during lunch. However, this number, instead of for instance 2, was chosen to account for other dishes resulting from snacks, drinking and food preparation as well.

Furthermore, instead of assuming additional items (AS) for every six serving cycles, instead it was assumed these are produced only once a day for dinner. This assumption was made based on the actual items this category includes, such as dinner bowls (see Table 2.3). This assumption also tries to cover the potentially higher amount of dishes produced by a single person and relatively decreasing effect for every additional household member (as noted in [24]).

Finally, to account for the relative high number of daily servings six days per week eating at home were assumed. The total number of items per week was then calculated:

$$Items = [(H * TS * D) + (AS * sf)] * W \tag{3.20}$$

Table 3.3 shows the calculated number of items assumed to be produced for different household sizes. The number of weekly washes is then total number of items that need to be cleaned on a weekly basis divided by the machine capacity, for each household size. This resulted in table 3.4 of supposed minimum and maximum weekly washes, depending on family size and machine capacity:

Table 3.3: *Total amount of items with increasing number of people*

Number of people	Items per person per serving	Total items per day[1]	Total items per week[2]
1	18	54	324
2	14.5	87	522
3	13.3	120	720
4	12.75	153	918
5	12.4	186	1116

[1] For 3 serving cycles per person and one large additional setting [2] For 6 days servings

Hence, when the family size and machine size are known, the specific maximum efficiency threshold can be calculated (see Table 3.4).

Table 3.4: *Assumed frequency efficiency per week*

	Number of people									
	1		**2**		**3**		**4**		**5**	
Type of dishwasher	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Table dishwasher	3.7	5.6	6.0	9.0	8.3	12.4	10.6	15.8	12.8	19.2
Dishwasher (45 cm)	2.2	2.8	3.6	4.5	5.0	6.2	6.3	7.9	7.7	9.6
Dishwasher (60 cm)	1.4	1.9	2.3	3.0	3.1	4.1	4.0	5.3	4.8	6.4

To establish a standard threshold to serve as proxy for effieincy, the following characteristics were chosen. Most people in central Europe use a normal sized dishwasher (60cm), with 12 or 13 place settings. The average family size sampled in the plug survey was 3.3. According to the model the minimum and maximum usage frequency for 3.3 household members lies between 3.4 and 4.6 washes per week, hence averaging at 4 washes per week. Considerably below the 5.4 resulting from the 280 washes assumed in the energy label calculation, slightly below EuP and much in line with VEWIN research by household size in the Netherlands (subsection 2.4.3).

Therefore, as an initial generic efficiency threshold the value of 4 washes per week was used for further analysis. The number of weekly washes estimated by the frequency estimation (described in subsection 3.3.2) can then be compared with this established threshold. This can be used to classify if a household washes more or less times per week than the efficiency threshold. For the performance of this efficiency classification see Results section 4.2.3.

3.4 Smart Meter Model Deployment

This section describes the methodology used to analyse the output from one full year of data on which the NILM detection system was applied for nearly 130.000 households. The section starts with a description of the data preparation and processing of the smart meter data and an accompanying survey. The subsection on consumption analysis outlines the aggregation and segmentation methods used to describe energy usage and usage frequency statistics, dependency on time of the day, week and year and by categories such as household size and machine characteristics. The last subsection describes how these results were used to analyse efficient usage of dishwashers compared to the established efficiency thresholds.

3.4.1 Data preparation

The next component of this research is to deploy the developed model on a large-scale set of households, without smart plugs. In addition to the smart meter profiles a survey was conducted with a sample of the users. This subsection describes the preparation of these two main data sets.

Smart meter analysis data

The smart meter data consists of the complete aggregate load profile of the household. The different steps of the model were used to prepare the data for the consumption and efficiency analysis of the dishwashers.

Data gathering
Gathered smart meter data for 158,037 households in the Netherlands from the beginning of January until the end of December 2018 was prepared and used for analysis. The profiles were measured either by the smart meter or analogue meter and sent to the Toon smart thermostat, which logged the data and sent it to a cloud database. This data was accessed using big data library PySpark within a DataBricks user interface. Data was already pre-processed to detect dishwashers out of the smart meter signal, by K. Basu. The queried detection data was then used to apply the developed smart meter model on (see section 3.3).

Data processing
Not all the data for every household was gathered from the same starting date and some smart meter devices showed data logging gaps. To represent a year round of data, two selection steps were applied:

- Only households were selected, for which detection data was available for a threshold of at least 40 weeks in 2018. In this year public and school holidays combined to 92 days (3.03 months, see appendix table E.1). It was therefore assumed that most people would not be absent from the household for more than twelve weeks per year.

- The households were filtered to at least show some activity before the end of February and some activity after October.

After the selection the dishwasher power profiles were detected within the aggregated household power profile, by deploying the appliance detection algorithm. The smart meter estimation algorithm could then be applied on the detected feature values. By applying the estimation algorithm energy consumption per wash and the usage frequency could be calculated for each individual household.

The energy detection of a dishwasher cycle was restricted to a maximum energy consumption per wash of 3 kWh. The average energy consumption in 1970 was just below 3 kWh and has been going down to an average of below 1kWh for the more recent models on the market (figure 2.10). It was assumed that no machine from before 1970 was in use. Outliers above this threshold were filtered out of the dataset because they were assumed to be misdetections. Applying these filters on the smart meter data led to a number of 129,137 households, which is a data loss of 18%.

Smart Meter survey

To gain more insight into how different household and machine characteristics affect energy consumption, results of a survey were added.

Data gathering
Those costumers of the Toon smart meter, who gave consent for the use of their electricity data in research were offered to use a mobile app to track their energy consumption. A survey was developed for the app-users to retrieve additional features on machine characteristics and usage behaviour. These included: household size, number of weekly washes, machine size and energy efficiency label.

The survey option appeared, when starting up the Toon app called "Verspillingscheck". Users would receive information on their energy usage and were prompted to fill out the questionnaire to help them gain a better understanding on factors affecting the energy consumption of their appliance. In total 10,873 households took part in the survey.

Data processing
The smart meter survey data was matched with the smart meter data based on the anonymous serial numbers of the households.

3.4.2 Consumption analysis

The large scale smart meter data was analysed to retrieve information on overall energy consumption and usage frequency statistics. Dishwasher consumption patterns were compared for dependency on time from events such as holidays, day and weekly patterns and seasonality. Finally, the survey results were matched to segment the outcomes based on machine and household categories.

Consumption overview

For the consumption overview, summary statistics for common statistical parameters (mean, std, min, max etc.) were generated per household. A summary of the total energy per wash, per week and per year as well as the frequency of usage per week and per year was retrieved. The calculation of energy per wash as well as weekly washes for each individual household was based on the detection-estimation model as described in section 3.3.

Summary statistics
The summary statistics for energy per wash were calculated based on the average energy per wash for each household. They were calculated non-weighted meaning each household was considered equally, regardless of number of washes.

The number of weekly washes was based on averaging the number of washes per week for all 52 weeks for every individual household. To represent weeks in which no wash was detected, these weeks were added as zero values. The weekly energy consumption is calculated by summing the energy consumption of each wash in that week and, likewise to the weekly washes, averaging these 52 weeks over the course of the year. The annual number of washes and total energy consumption was then summed over the whole year for each individual household and summary statistics were calculated.

Data distribution
In order to better understand how summary statistics for the different data sets compare, the energy consumption and usage frequency for all nearly 130.000 households for whom activity was detected from the smart meter, was compared with the respective data available from the 100 households with smart plugs installed and the sub-group of nearly 11.000 households who filled out the survey. Plug data was gathered from Nov 2016 until Feb 2017. To ensure no effect of seasonality, the same months are used for this comparison. Since smart meter data is available for the full year 2018, the months January, February and December were taken from 2018.

Results can be seen in subsection 4.3.1.

Consumption time dependency

While the statistics in the previous subsection look at energy consumption per wash and number of washes per week, averaged based on a full year of data, consumption patterns can potentially vary throughout the day, week and seasons. Since a major change was made to the sensitivity of the detection algorithm in week 15 of gathering the data, the time line has been adjusted to match the time period before and after this change in week 15. For more on this change, see appendix E.1.

Seasonality on energy per wash
For the assessment of seasonal influences, the weekly aggregated average energy consumption for each week was plotted. The pattern was then compared to weekly mean outside temperature of the Netherlands measured by weather stations per day in 2018, downloaded from KNMI (Royal Netherlands Meteorological Institute) [161]

Seasonality on washes per week
To asses the effects of seasonality and effects of events such as holidays throughout the year on usage frequency, the average weekly washes were plotted as well. The pattern was compared with data on public holidays in the Netherlands in 2018 (see appendix table E.1) to assess the impact of these events.

Week-hour activity pattern
To analyse the chance that dishwashers are being used at different times of the day and week, the amount of washes per week-hours was assessed. For this aggregation, all counts for dishwasher start times at a specific hour of a week were summed. All week-hours of a respective week were then summed for the year of available data. To assess the relative share of the washes per week-hour, all washes found were divided by the total number of households. The resulting data adds to 100% of all detected washes in the year, with each hour of the week representing how many washes were started on that particular hour throughout the year, revealing a pattern for dishwasher usage.

Results can be seen in subsection 4.3.3.

Consumption category dependency

To compare dependency on household and machine characteristics, the survey data was matched with the smart meter data and statistics were compared for the different subgroups. Available factors such as machine and household characteristics were matched with energy per wash and usage frequency, in accordance with dependencies shown in figure 3.7 and as modelled as factors for the efficiency thresholds in section 3.3.3.

Effect of dishwasher capacity on energy and frequency
Both the energy per wash and the washing frequency were shown to relate to the machine capacity (ps). Therefore, the dishwasher size stated in the survey was combined with the average detected energy and average weekly detected washes of the meter data.

Effect of family size on usage frequency
The other major factor affecting usage frequency, was number of household members. To see the effect of household size, the stated number of washes and the detected washes where combined and then grouped by the stated household size. The stated average weekly washing frequency was compared to the detected smart meter average.

Effect of efficiency label and washing temperature on energy
Finally, the efficiency label and washing temperature are the factors mainly affecting washing temperature. Hence, the stated efficiency label and washing temperature where grouped and connected with average detected energy.

Results can be seen in subsection 4.3.2.

3.4.3 Efficiency analysis

After gaining understanding about factors that affect the energy consumption and usage patterns of dishwasher usage, the final step is to determine how efficient households use their dishwashers. First the estimated energy consumption and number of weekly washes are compared to the generic efficiency thresholds. Next these findings are compared by category to look at the relative efficiency. Finally the findings are summarised to calculate how much households could potentially reduce their annual dishwasher energy consumption.

Generic efficiency classification

For both energy per wash and weekly usage the estimated usage was compared to the generic binary thresholds established to serve as proxy for efficient usage (as described in subsection 3.3.3). Households over this set threshold are classified as (potentially) inefficient. Using a binary threshold as initial indication for efficient dishwasher usage provides the opportunity to assess the potential usage efficiency without additional information on such factors as family size, machine size and efficiency label, potentially making this applicable to all nearly 130.000 monitored households. However, the additional survey information can be used to gain a deeper understanding about the household and machine characteristics, in order to better understand reasons for some identified potential inefficient usage as well. In order to compare the results on several household and machine characteristics in the next steps, the efficiency classification was only applied for the nearly 11.000 households who filled out the survey.

The set generic threshold could be varied to optimise the amount of false positives, false negatives or overall accuracy. Sensitivity to these adjustments was therefore compared. The survey results were taken as reference measure in order to calculate how often the model makes the same classification as how the household would be classified based on their survey results.

Relative efficiency classification

Classifications made for the generic threshold were then compared to a classification in relation to the household and machine characteristics of the individual household. In order to make these specific thresholds, the same model approach was used as described in subsection 3.3.3) to find the generic thresholds. However, instead of using certain average values based on the literature, for each individual user, who filled out the survey, the information from the survey, such as efficiency label, machine size and number of household members was then used to calculate a threshold to serve as a more specific proxy for efficient usage for their situation. The results for each category were then segmented to show how the generic and specific thresholds compare in classifying efficient dishwasher usage.

Potential energy savings

To summarise these findings and indicate how much energy households could potentially save on average on their dishwasher usage, the two generic thresholds were combined and compared to the annual energy consumption. Based on the 1.05 kWh per wash and 4 weekly washes an annual energy threshold of 218kWh was established. Households were then segmented based on their assigned efficiency classification from the previous steps (inefficient usage frequency, high energy consumption per wash) to show how much households could potentially save on their annual energy consumption based on the different potential improvements.

3.5 Methodology conclusion

The methodology uses findings from the first and second research sub-questions to develop an approach to prepare for answering sub-questions 3, 4 and 5. Sub-question 2 was already answered in part in the literature chapter by looking at existing literature on NILM research. This was found to mainly focus on the detection of appliance activity, but less on real-world application steps after detection. The methodology developed in this research extends this to be able to analyse energy and efficiency of real life dishwasher usage, by that answering sub-question 2:

> **2. How can a smart meter based NILM system be developed to detect energy consumption, usage pattern and efficiency characteristics for dishwashers?**
>
> The research consists of two phases: development and deployment. The development consists of two components; the plug analysis was done as a preparatory step to develop the estimation model and hence is completely integrated within the model development section. Each of the three components then consisted of 3 main steps: Data acquisition and processing, which includes the detection, then consumption estimation and analysis and finally the efficiency classification and analysis.
>
> In order to do this, first a model was developed in Python to calculate dishwasher energy usage and washing frequency and then classify efficient usage based on data from the central meter. The developed system was deployed on a very large set of smart meter data to non-intrusively analyse energy consumption, usage pattern and efficiency characteristics of real-life dishwasher usage in households.
>
> The model developed in this research extends the NILM methodology developed by K. Basu (2017), which provided the possibility of detecting dishwasher usage on the smart meter power signal. This research used 100 households with plugs connected to the dishwasher and smart meter data sampled at 10 second intervals to develop a more granular level of detection. Energy consumption per wash was estimated, using linear regression on several of the detected features such as total heating period, heating power and number of heating moments. Energy consumption as calculated from the plug data was used to calibrate the regression parameters of the NILM algorithm.
>
> In order to assess the efficiency of the dishwasher usage a binary threshold model

was developed, serving as a proxy for efficient energy consumption per wash in kWh and efficient usage frequency in number of washes per week. In order to establish values for the energy efficiency threshold the Eco design regulation was used, which contains calculations for dishwashers based on efficiency label and size and assuming a standard washing program. For usage frequency only an assumption of 280 washes per year is available, but no information is given on what usage frequency could be considered as efficient. Therefore weekly usage was estimated with a model based on number of family members and daily usage of dishes. Both efficiency thresholds were calibrated to improve accuracy based on the 100 validation households. The developed model was then deployed to analyse energy consumption, usage pattern and efficiency characteristics for dishwashers for nearly 130.000 households with smart meters for the period of one full year.

4

Results

4.1 Results introduction

The results chapter consists of two main parts, analysis of the developed model and analysis of the results from its deployment. Research question 3 is answered in the first section. The first part of the results (4.1) shows how the newly developed system performs in detecting energy usage, usage frequency and efficiency for dishwashers from the smart-meter signal. To develop the model, characteristics of dishwasher usage first had to be analysed based on data from the plugs, hence part of research question 4 and 5 are answered in this comparison already as well. To assess the accuracy of the model, the output based on the smart meter data is compared with the smart plug data. More detailed analysis of the smart plug data can be found in Appendix C and a comparison of several other tested algorithms in Appendix D. Section 4.2 displays the results of the real-life application of the newly developed NILM-based detection system. The results for the consumption analysis consist of energy usage and usage frequency averages depending on time, household and machine characteristics and ends with an efficiency analysis, by that answering question 4 and 5. Some additional results are given in Appendix E.

Answers the following sub-questions:

3. How does the newly developed system perform in detecting energy usage and efficiency?
4. How much energy do dishwashers in households consume and how often are they used, depending on time, household and machine characteristics?
5. How efficiently are dishwashers used in households?

4.2 Smart Meter Model Development

This section describes how the newly developed system performs in detecting energy usage, usage frequency and efficiency for dishwashers from the smart meter signal. To asses the performance of the model the output based on the smart meter data is compared with the smart plug data. The summary statistics of an analysis of the smart plug data (see appendix C for more detail) is compared with the summary statistics for the developed smart meter model. The accuracy for different model approaches was compared (see appendix D for more detail). The main results for the best functioning approaches are presented in the following subsections on energy per wash, usage frequency and efficiency classification.

4.2.1 Energy per wash estimation

This paragraph presents the accuracy of estimation of the energy consumption per wash based on smart meter data. The estimation of the energy consumption per wash utilises the detection features in a regression model to estimate the energy consumption as discussed in methodology subsection 3.3.2. First the predictive quality of selected features is shown, followed by an overview of performance for several estimation methods.

Feature analysis

To reveal which features are most relevant for the energy estimation, the predictive quality is assessed. The predictive quality of the different selected features depends on how important they are for the energy consumption and how well they are detected. The summary statistics of the considered features have been compared for both smart-plug and smart meter detections. Table 4.1 presents the selected features summarised for all individual investigated (5691) washes, as detected by the smart plugs. The table shows the data from the smart plugs as this best reveals the actual value of the features, a similar table for smart meter data can be found in appendix D.0.1.

As can be seen in table 4.1 the mean energy consumption of one total wash is 1.21 kWh. This total energy consumption can be split up into different processes. At 82%, the largest share of the energy is used for the water heating, translating into an average energy usage of 1 kWh. The remaining 0.21 kWh is used by other processes such as rinsing, drying and pumping water. The average total duration of water heating is 30 min, which is spread over an average of 2.7 heating cycles, making up about one third of the total duration of the wash.

However, as washes can get missed or misdetected by the smart meter, not every detection can be matched with its individual smart plug detection and therefore cannot be matched with the energy consumption as measured by the smart plug. The energy as measured by the smart plug can therefore only be connected with the smart meter data on a per household aggregated level. Comparing the smart plug values with the results averaged and aggregated for 100 households (appendix D.0.1) only slight differences were noted. Most notably the energy per wash changes slightly to 1.22kWh, which we will use as average number detected by the smart plugs in further analysis.

Table 4.1: *Summary statistics of potential energy consumption parameters for all washes detected by the smart plugs*

	Plug detected consumption parameters					
	count[a]	**mean**	**std**	**min**	**median**	**max**
Energy usage of total wash [kWh]	5691	1.21	0.27	0.44	1.18	2.29
Energy usage of water heating [kWh]	5691	1.00	0.25	0.36	0.97	1.99
Share of energy for water heating	5691	82%	8%	40%	84%	99%
Duration of total wash [min]	5691	85	30	21	85	159
Duration of water heating [min]	5691	30	8	11	29	69
Heating power [kW]	5691	2.0	0.1	1.3	2.0	2.3
Number of heating moments	5691	2.7	0.8	2	2	5

[a] Number of washes

To assess the relevance of these parameters with regards to energy consumption per wash, the correlation is presented in table 4.2. Some parameters such as duration of water heating and energy usage of water heating show high correlations, while parameters such as duration of total and heating power show lower correlation. The area, expressed by the energy usage of water heating consequently shows a very high correlation with energy usage. As it is responsible for an average of 82% of the energy usage for smart plug detections and 78% as detected by the smart meter (appendix D.0.1), it results in a 0.94 correlatation for averaged smart plug detections and 0.82 for averaged smart meter detections.

Table 4.2: *Pearson correlation of selected parameters with energy usage in kWh per wash for all smart plug detections and smart plug and smart meter detections averaged per household*

	All plug		**Avg plug**		**Avg meter**	
	count[a]	**corr**	**count**[a]	**corr**	**count**[a]	**corr**
Energy heating water	5691	0.92	100	0.94	100	0.82
Duration total wash	5691	0.32	100	0.21	100	0.22
Duration heating water	5691	0.90	100	0.90	100	0.82
Heating power	5691	-0.03	100	-0.03	100	-0.02
Heating moments	5691	0.66	100	0.61	100	0.51

[a] Number of washes

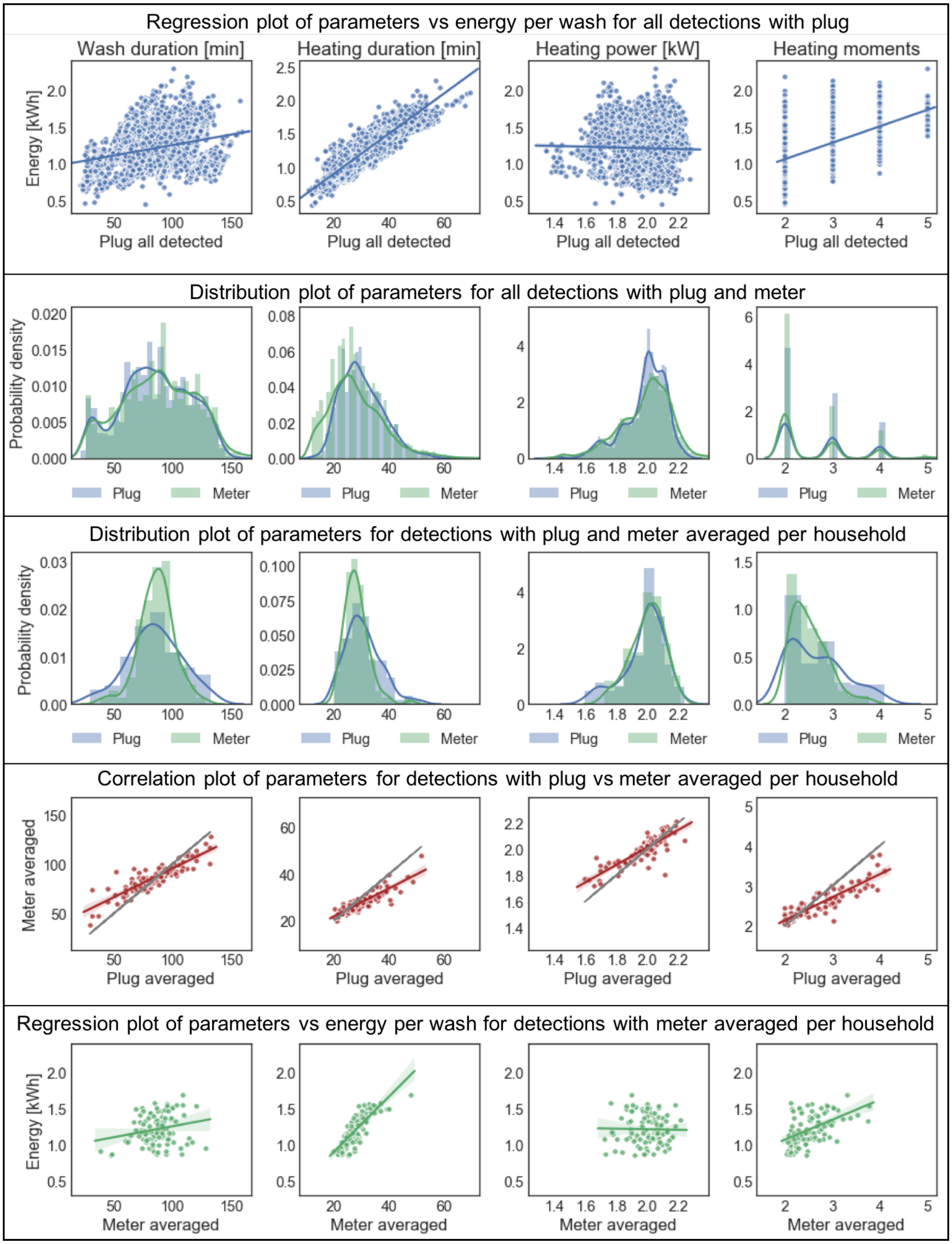

Figure 4.1: *Overview of selected features (columns) detected by smart plug and smart meter (referred to as plug and meter)*

Figure 4.1 gives a graphical representation of the assessment for the different features (same feature below each other in each row). In the top row a regression of selected features compared to energy for all smart plug detections are shown in a scatter plot. In the second row the distribution of these features for all detections with smart plugs and smart meters are shown. The third row depicts the distributions of the same features averaged per household. In the fourth row the scatter plots for the household averaged detections for smart plug and smart meter are plotted against each other. On the horizontal axis the smart plug values and on the vertical axis the smart meter values are given. The dark grey line shows perfect correlation, where the smart meter would detect exactly the same as the smart plugs. The red line shows the actual relationship, were deviation from the grey line represents the level of over- or underdetection. The fifth row shows the scatter plot of the per household averaged smart meter detected values versus the matching per household averaged smart plug detected energy consumption per wash. The findings for each of the four features are described below:

Total wash duration [min] in the first column the rather broad spread of datapoints is in line with the low correlation between wash duration and energy of 0.32. In row two and three, the distribution is overlapping considerably accurate. The difference here is that the averaged smart meter data is centering more around the mean than the averaged smart plug data and detects far outliers less well. When considering the respective scatter in row 4, it can be seen that the scatter for this parameter aligns well around the smart plug-meter correlation line, though the lower smart plug detections are slightly overestimated for instance at 40 minutes the smart meter might detect 70 minutes and at some of the longer washes the duration is underestimated on the smart meter.

Duration of the water heating [min] in the second column can be seen to be highly correlated with energy usage (0.92 as shown in table 4.2). The distributions in both the second and third row show that the smart meter distribution (green) is shifted to the left. In the respective scatter plot in row four, this is stipulated by most detections being below the correlation line. Hence, the heating duration is often underestimated by the smart meter. Nevertheless, despite underestimation, the detections do show to tightly scatter around the regression line in row five.

Heating power [kW] is presented in column three. Heating power seems to be scattered quite broadly. A visual inspection of the smart plug data showed that the smart plug detection was not performing so well on detecting the power of the water heating. Often small signal breakages distorted how high a heating block was estimated, hence smart plug data is not such a precise representation of reality in this case either. Nevertheless, averaged smart meter detection data does fall in the right range and overlaps actually quite close with smart plug detection.

Number of heating moments in the far right column shows a step up pattern, which resulted in a correlation factor of 0.66. The second, third and fourth row clearly indicate an underdetection in the number of heating moments, which seams a reasonable explanation for the heating duration being underdetected, as apparently every now and then a heating moment at either the beginning or end of the wash is missed.

Heating energy [kWh] is not shown here as this was not a seperate feature of the detection algoritm. However, it can be simply calculated by multiplying the heating duration with power. According to table 4.1 this represents 82% of the total energy consumption as measured by the smart plugs and according to table 4.2 has the highest correlation with energy.

Estimation model

To further investigate the relation between energy and the most relevant detection feature, i.e. heating energy, the two variables are plotted against each other. Figure 4.2 shows the average kWh of a wash detected by the smart plugs on the vertical axis and the average heating energy per wash detected by the smart meter on the horizontal axis. The diagonal black line shows the case of a perfect linear relationship between the 2 variables. Each dot represents one household, the size represents the total number of observed washes.

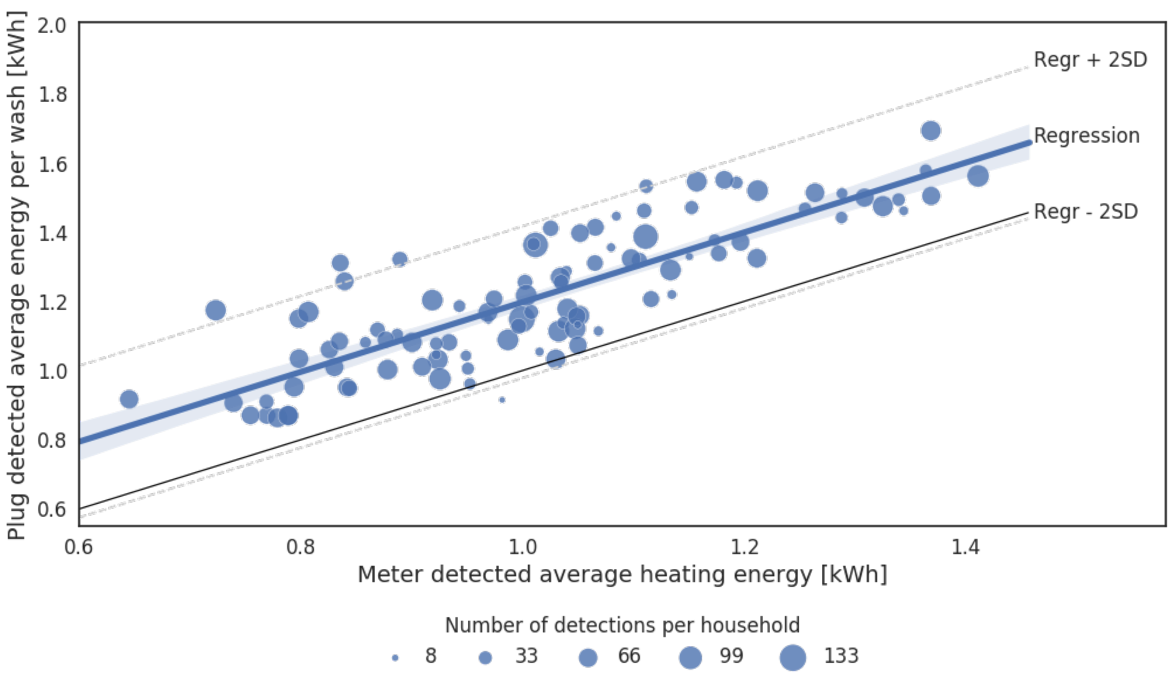

Figure 4.2: *Average energy usage per wash detected by smart plugs and by smart meter*

As can be seen, the regression line (dark blue) through the scatter plot is horizontally shifted upwards by roughly 0.2 kWh, with some slight variation towards the edges (regression error in shaded blue). This indicates that the heating energy is an adequate approach to determine the average energy consumption, in particular when adding a standard offset of 200Wh (about 16%) for the other energy processes. From the 95% interval indicated by the dashed grey line a standard deviation of roughly 0.1 kWh can be read off. However, it has to be noted the offset depends on the ratio between heating energy and the other activities, both the offset value and ratio would therefore likely change and therefore need to be updated over time for newer machines.

Estimation approaches
To further investigate what approach performs best, several different alterations were investigated. The estimation approaches can be split in two categories, one is the heuristic approach, the other is using a regression model. The heuristic method is based on set assumptions, while the regression model aims to capture the detection bias with the regression parameters (see 3.3.2). Based on the features that showed good correlation in the previous subsection, several combinations were assessed. The estimation methods where chosen on basis of simplicity and accuracy (for more detailed accuracy assessment of different investigated combinations see appendix table D.3). The following three estimation methods were selected for further comparison:

Predict 1: A heuristic approach that uses the heating energy as the first variable and adds an average offset (mean energy usage of the category 'other') at 200Wh.

Predict 2: Linear regression estimating the parameters based on heating energy + heating power + heating duration + wash duration + number of heating moments.

Predict 3: Linear regression only using number of heating moments + heating energy

To match energy (from smart detections) with features from the smart meter detections the data has to be aggregated for each household. As aggregation method the average and mode were calculated and tested for different estimation methods (detailed results see appendix D.0.1). Additionally for the best performing algorithm the GMM clustering was considered and tested (detailed results see appendix D.4). The estimation of average energy consumption per wash could be achieved with a lower estimation error than the estimation error of the most common wash (mode) and GMM .

Summary statistics
Summary statistics for energy usage per wash for the 100 test housheholds (number of households specified as count in table below) are shown in table 4.3. The data is based on averages for each household. The three selected estimation models are compared with the energy as retrieved from the smart smart and the direct smart meter detections. The smart meter category represents direct calculation of energy from the heating block only (multiplying maximum power with total heating duration), which

can be thought of as the most straightforward energy estimation method (or Predict 0). Predict 1 follows the most straight forward heuristic approach. Predict 2 includes all regressors while for Predict 3 several of the less relevant variables were pruned to only contain the ones with the highest predictive quality.

Table 4.3: *Summary statistics for averaged dishwasher energy usage per wash*

	Smart plug	**Smart meter**	**Predict 1**	**Predict 2**	**Predict 3**
Averaged					
count[a]	100	100	100	100	100
mean[b]	1.22	1.02	1.23	1.22	1.22
std	0.20	0.17	0.17	0.18	0.18
min	0.86	0.65	0.86	0.86	0.84
median	1.18	1.01	1.23	1.21	1.21
max	1.69	1.41	1.62	1.66	1.62

[a]Count is number of households [b]Average energy per wash for 100 households, averaged over all washes in a period of 12 weeks.

The distribution of the three estimation methods is depicted by the respective probability density functions of average energy usage per wash per household in figure 4.3.

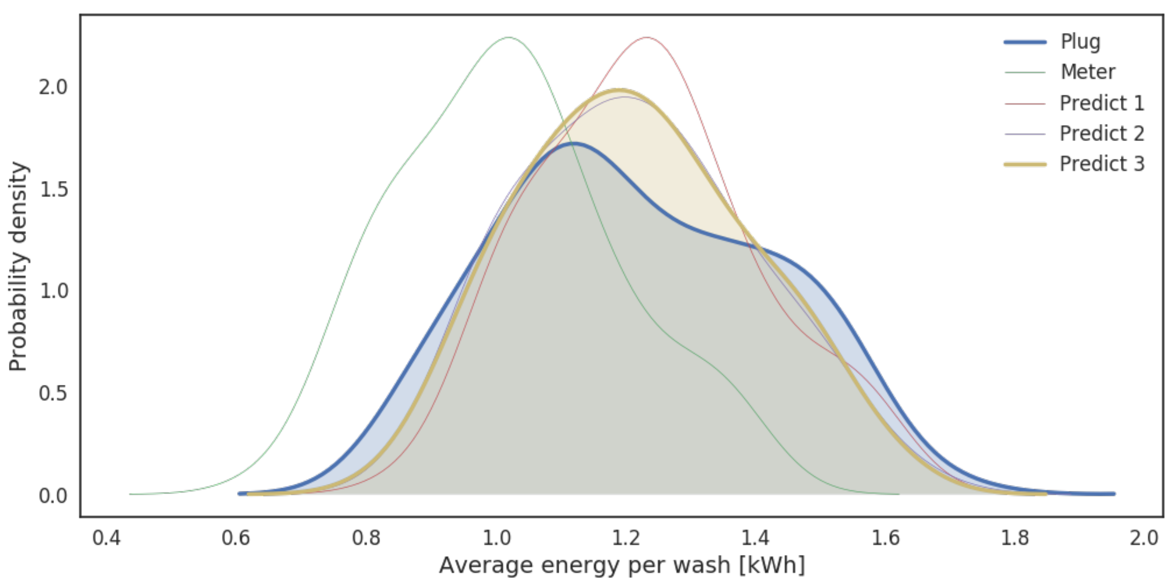

Figure 4.3: *Probability density functions of average energy usage per wash per household for smart plug, smart meter and different estimation methods*

The average energy consumption (mean) of the smart-meter detection diverges 0.20 kWh from the smart-plug detection (table 4.3). In the distribution figure it can also be seen that the smart meter alone (green) is considerably off from the smart plugs (blue). The estimation methods (Predict 1, 2, 3) all show better overlap in distribution. Predict 3 (yellow) shows the best overlap with the smart plugs. However, a performance difference between the estimation methods can not be made on the basis of this analysis alone.

Uncertainty analysis

To compare the performance of the different estimation methods, the accuracy of estimation was calculated (for more on estimation assessment, see appendix B.0.2). Heating energy shows to be the most relevant regressor, with an NRSME (Normalized Root Mean Square Error) of 9.1%. Number of heating moments shows to be relevant (correlation of 0.66) but produces relatively poor result as single regressor with an NRMSE of 14.4%. However, the combination of the regressors heating energy and heating moments appear to have the lowest NRSME of 8.8%. Predict 1 (9.2%) and Predict 2 (9.1%) perform slightly less accurate (more detailed results in appendix D.0.1).

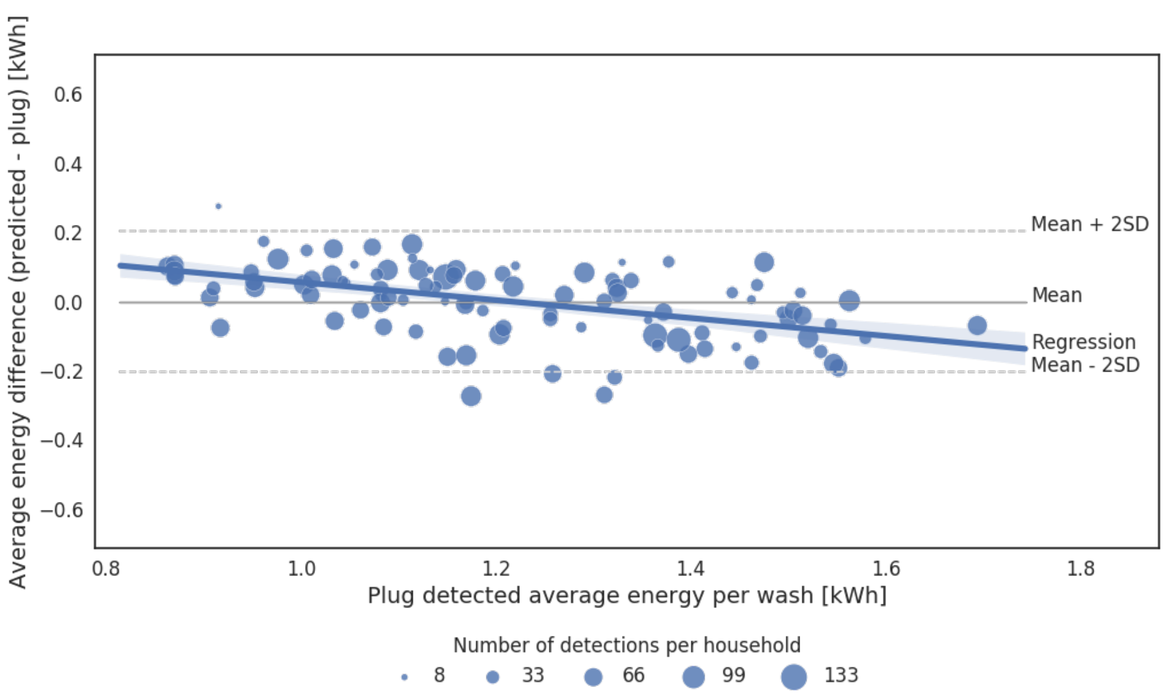

Figure 4.4: *Difference between average energy usage per wash by smart plug and estimated versus smart plug average*

Finally, to better understand the error distribution, figure 4.4 shows the difference between the average energy usage (in kWh) by smart plug measurements and what is estimated using Predict 3. The average energy consumption by both smart and Predict 3 was estimated to be 1.22kWh (table 4.3), resulting in a mean difference of 0 (horizontal grey line). It shows that 95% (dashed grey line) of the estimation errors fall between +0.21 and -0.21 kWh. This is in line with an 8.8% NRSME on an average energy consumption per wash of 1.22 kWh. The regression shows that more often there is a slight over-estimation for energy consumption per wash below the mean energy consumption and more often an under-estimation of energy consumption per wash above the mean. While this is a common result of the offset (β_0) weighing relatively higher for the lower values and relatively lower for the higher values, it is an important factor affecting estimates further away from the mean. While this affect can be shown in comparison to the smart plug detections in figure 4.4, for future estimates where no smart plug data is available, the extent to which a value is over- or underestimated is unknown.

4.2.2 Usage frequency estimation

Not all washes get detected (false negatives) and some detected washes are actually misdetections (false positive). While this varies by household, the NILM detection algorithm showed an average rate of 15 false positives and 26 false negatives per 100 correct detected washes. The detection model can be balanced. Balancing it to reduce false positives, increases the precision and hence can improve the accuracy of energy estimation. However, the higher rated of missed washes (false negatives) results in underestimation of the weekly number of washes. In the previous subsection the relationship between energy per was and detected features was investigated to be able to estimate the energy per wash. A similar relationship can be analysed for weekly washes detected on the smart meter, compared to the smart plugs.

Figure 4.5 shows the weekly smart meter detections on the horizontal axis, plotted against the smart plug detections on the vertical axis. Every dot represents a combination of the number of smart meter detections in a week for a particular household versus what was actually detected by the corresponding smart plug. Bigger dots imply that the combination occurred more often. The black line shows the zero difference line. The bubbles would align closely around this line if the smart meter detects a similar amount of weekly washes as the smart plugs. The blue line shows the regression for the actual relationship between washes detected by the smart meter compared to the smart plugs, including its error in shaded blue. The grey dotted line shows the 95% interval around the regression line.

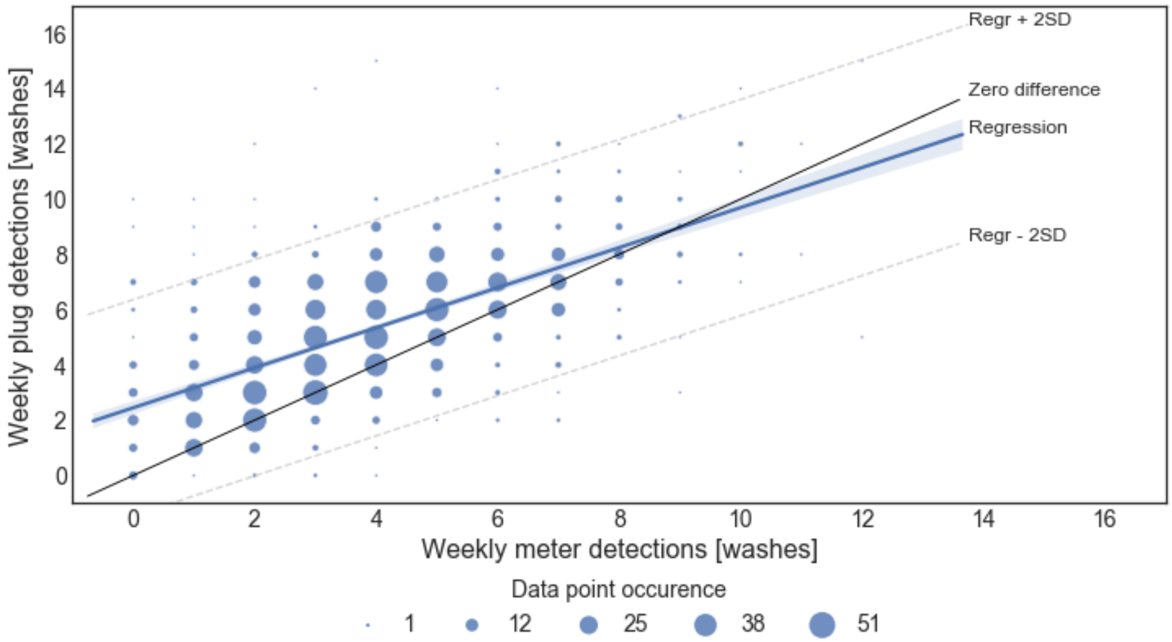

Figure 4.5: *Weekly dishwasher usage count detected by smart meter versus detected by smart plugs*

It can be seen that most bubbles are above the zero difference line, indicating that the number of weekly washes according to the smart plugs are higher than the number of washes detected by the smart meter. The regression line does not line up with the zero difference line as it did for the energy regression. It starts at slightly over three, where underdetection by the smart meter occurs i.e. it more often misses washes (false negatives) and there are not so many false positives. This is because the aim is to rather estimate on high quality detections i.e. detections where one can be very sure that it is a dishwasher instead of creating a lot of noise by mis-detecting dishwashers where the algorithm is less certain, hence these are dropped out.

Shown in table 4.4 the summary statistics for the 100 sample households (number of households specified as count in table below), the number of weekly washes, averaged per household over the period of 12 weeks. A comparison is made between summary statistics of weekly detections for smart plugs compared with smart meter detections and several estimation methods based on the smart meter detections:

Predict 1: Basic linear regression (as depicted in figure 4.5)

Predict 2: Linear regression forcing regression parameter $\beta_0 = 0$, resulting in the regression starting without additional offset

Predict 3: Linear regression as a hybrid of predict 1 and 2. Uses the estimation resulting from Predict 2. However, it uses the difference between the mean value of Predict 1 and Predict 2 and adds this as an offset.

Table 4.4: *Summary statistics for dishwasher: Averaged weekly detections by smart plug and smart meter and estimations on usage frequency according to three different estimation methods*

	Plug	**Meter**	**Predict 1**	**Predict 2**	**Predict 3**
Averaged					
count[a]	100	100	100	100	100
mean[b]	5.2	3.9	5.2	4.7	5.2
std	2.1	1.6	1.2	2.0	2.0
min	1	1.1	3.2	1.2	1.8
median	5.6	3.8	5.2	4.6	5.2
max	11.3	9.2	9.1	11.1	11.7

[a]Count is number of households [b]Weekly dishwasher detections for 100 households, averaged over a period of 12 weeks.

According to the smart plug data, average weekly washes is 5.2. The smart meter underdetects this on average with 1.3 washes, detecting an average of only 3.9 washes per week. Predict 1 matches the same mean value, but the standard deviation and min and max values, show it allows for considerably less variation. Predict 2 provides for more variation, but with 4.7 results in a lower mean value. As the algorithm is forced to keep $\beta_0 = 0$ and aims to reduce the error rate from there, it can result in a discrepancy between the estimated and actual mean value if this reduces the overall error rate. By adding the difference between this mean value and the correct mean value found with the general regression (Predict 1) as an offset, both a correct mean value and wide variation are provided.

Figure 4.6 shows the distribution of smart plug data in blue, combined with the smart meter data (green) and the three estimation methods. As can be seen the smart meter has the least overlap with the smart plug distribution, Predict 1 (red) has the same mean but less variance in distribution, Predict 2 (purple) comes closer but does not have the correct mean, predict 3 (yellow) shows most overlap with the smart plugs.

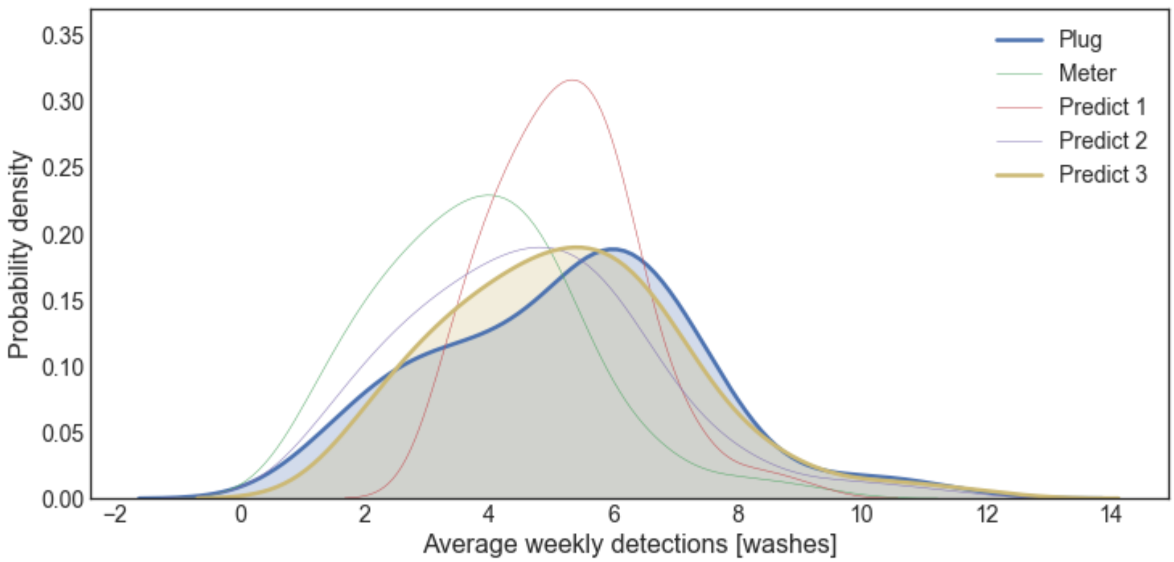

Figure 4.6: *Probability density functions of average weekly detections per household for smart plug, smart meter and different estimation methods*

Uncertainty analysis
To compare the performance of the different estimation methods, the accuracy of estimation was calculated (for more on estimation assessment, see appendix D.0.2). Looking at the results in D.5 it shows that all estimation methods perform better than direct smart meter detection. With a mean NRMSE of 27.2% the average of regression method in combination with Predict 3 appears to create the best result.

Algorithm Predict 3 was used for the difference plot, to compare estimation results with what was detected by the corresponding smart plug for a specific household in a specific week (see figure 4.7). On the horizontal axis the average number of washes per week, according to the smart plug. On the vertical axis the difference with what is estimated. The mean usage frequency according to both smart plugs and Predict 3 was estimated to be 5.2 washes per week (table 4.3), resulting in a mean difference of 0 (horizontal grey line). It shows that 95% (dashed grey line) of the estimation errors fall between +2.8 and -2.8 washes. This is in line with a 27.2% NRSME on a average number of 5.2 weekly washes.

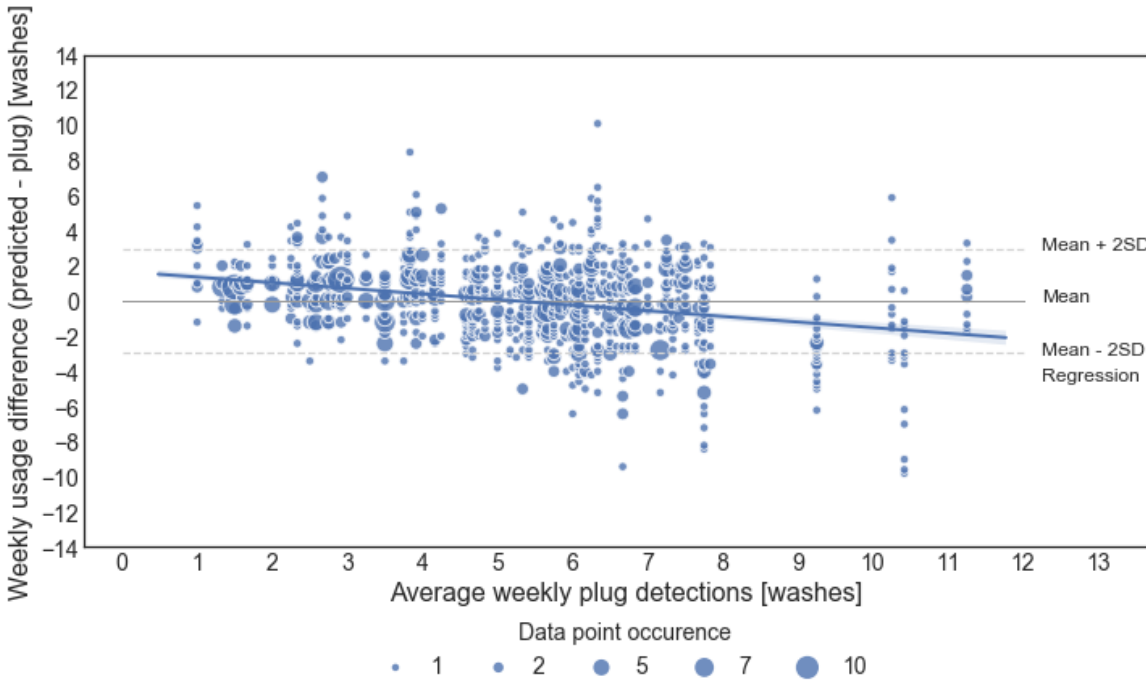

Figure 4.7: *Difference between weekly usage counted by smart plug and estimated versus smart plug average*

4.2.3 Efficiency classification

The estimated energy consumption per wash and estimated weekly usage can be used to classify if a household uses their dishwasher below the set efficiency thresholds. In this subsection the performance of this binary classification is assessed (for more on classification assessment, see methodology subsection 3.2.4). The thresholds can be varied to optimise the amount of false positives, false negatives or overall accuracy. Both the standard and optimal results are shown for set thresholds and the diagnostic ability visually compared.

Energy per wash efficiency

To assess efficiency with smart plugs, the set threshold of 1.05 kWh (see methodology subsection 3.3.3) can be compared with the energy consumption based on the smart plug's power measurement. The same can be done for energy estimated based on smart meter detections. Figure 4.8 shows smart measurements on the vertical axis, with the dark grey line as binary threshold for efficiency. All (22%) households below that (in green) according to the smart plugs are using an efficient amount of energy per wash. All (78%) households above that threshold on average wash inefficiently according to the smart plugs (in red).

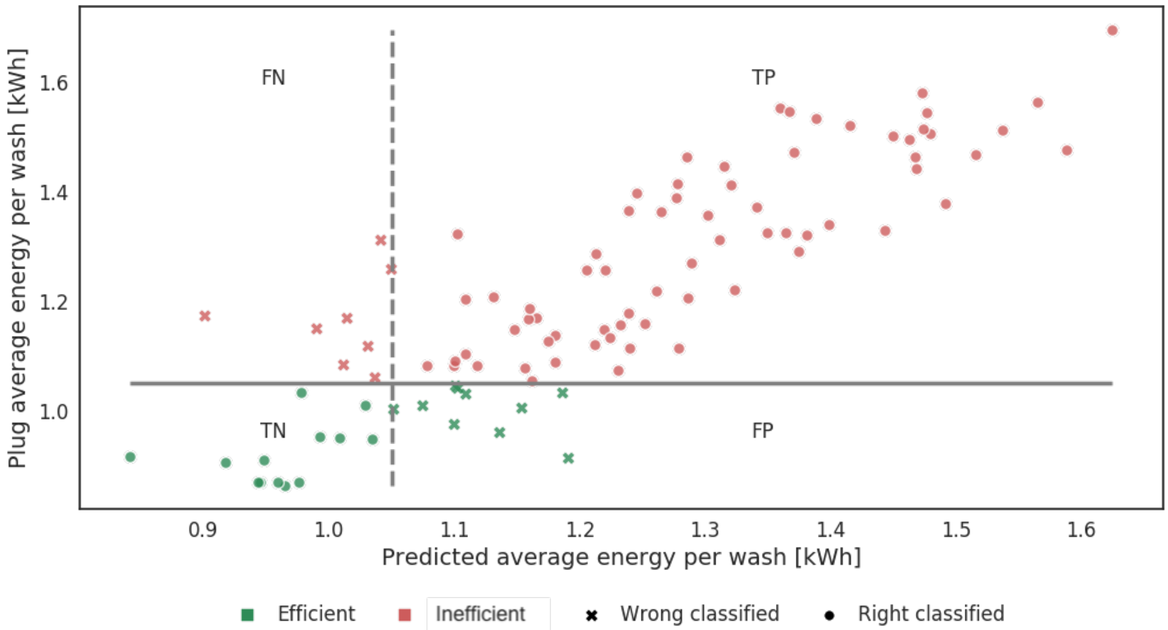

Figure 4.8: *Classification of efficient and inefficient average energy usage per dishwasher wash for estimated versus actual smart plug detections*

When applying the same threshold value to the smart meter estimated average energy per wash, several (18) households are wrongly classified. Either as (8) red crosses depicting classified as efficient while actually being inefficient according to the smart plugs (false negative). Green crosses are the (10) households that are classified as inefficient based on the estimated energy per wash while according to the smart plug data they are actually efficient (false positives). The dots represent households that are either (12) correctly classified as efficient (green) true negatives or (70) inefficient (red) true positives. The straight grey horizontal line is constant at the set threshold. The dotted grey line can be varied based on what shows the best classification accuracy.

Classification accuracy
Based on the above true and false positives and negatives, the precision and recall can be calculated. For a threshold of 1.05kWh a precision of 0.88, a recall of 0.90 and a F1 score of 0.89 were found. Table 4.5 shows the threshold in kWh per wash for the standard threshold, maximum precision, F1 score and recall against the related precision, recall and F1 score. It can be seen that the maximum accuracy (F1 = 0.91) can be achieved at a threshold of 1.01 kWh resulting in a precision of 0.86 and a recall of 0.97. When focusing on maximum precision a threshold of 1.19 kWh would be optimal, resulting in a recall of 0.68 and an F1 score of 0.81. When focusing on maximum recall a threshold of 0.84 kWh would be optimal, resulting in a precision of 0.78 and an F1 score of 0.88.

Table 4.5: *Maximum precision, recall and F1 score under different estimation thresholds for efficiency classification of average energy usage per dishwasher wash*

	Threshold	Precision	Recall	F1 score
0. Standard Threshold	1.05	0.88	0.90	0.89
1. Maximum Precision	1.19	1.00	0.68	0.81
2. Maximum F1 score	1.01	0.86	0.97	0.91
3. Maximum Recall	0.84	0.78	1.00	0.88

Based on calculation of 300 confusion matrices for different thresholds for the estimated average energy usage per wash. The threshold for smart plug values was fixed at 1.05 kWh per wash, while the threshold for estimated average energy usage per wash was varied.

Usage frequency efficiency

A similar plot can be made for average smart plug detections per week (on the vertical axis) against average weekly estimated washes (horizontal). A set threshold of 4 washes per week (see methodology subsection 3.3.3) results in 31% classified efficient. Of those 22 households (green dots) are also correctly classified as efficient with the algorithm and 9 (green crosses) falsely classified as inefficient. Of the 69% red, 7 (crosses) are mislabelled and 62 (dots) correctly identified as inefficient, compared to the set threshold.

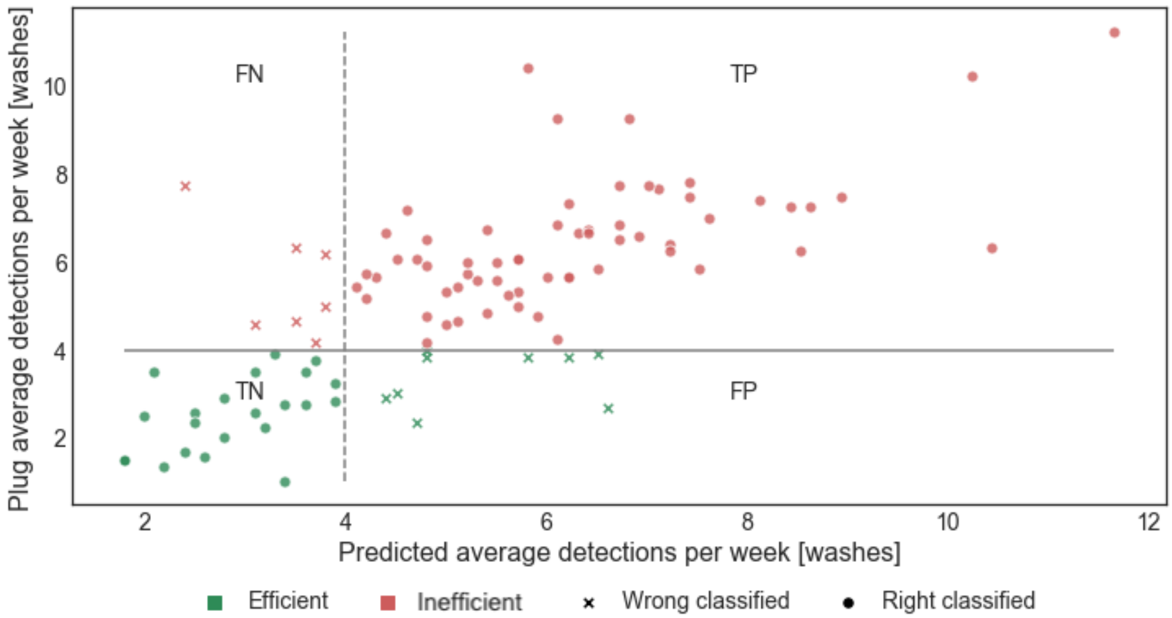

Figure 4.9: *Classification of efficient and inefficient average weekly dishwasher usage frequency for estimated versus actual smart plug detections*

Classification accuracy

For a threshold of 4 washes per week a precision of 0.87, a recall of 0.90 and a F1 score of 0.89 were found. Table 4.6 shows the threshold in days per week for the standard threshold, maximum precision, F1 score and recall against the related precision, recall and F1 score. It can be seen that the maximum accuracy (F1 = 0.89) can be achieved at a threshold of 3.39 days resulting in a precision of 0.83 and a recall of 0.97. When focusing on maximum precision a threshold of 6.61 days would be optimal, resulting in a recall of 0.3 and an F1 score of 0.81. When focusing on maximum recall a threshold of 2.22 days would be optimal, resulting in a precision of 0.73 and an F1 score of 0.88.

Table 4.6: *Maximum precision, recall and F1 score under different estimation thresholds for efficiency classification of average weekly dishwasher usage*

	Threshold	Precision	Recall	F1 score
0. Standard Threshold	4.00	0.87	0.90	0.89
1. Maximum Precision	6.61	1.00	0.30	0.46
2. Maximum F1 score	3.39	0.83	0.97	0.89
3. Maximum Recall	2.22	0.73	1.00	0.84

Based on calculation of 300 confusion matrices for different thresholds for the estimated average detections per week. The threshold for smart plug values was fixed at 4 average weekly usages, while the threshold for estimated average detections was varied.

Diagnostic ability

The diagnostic ability of the efficiency classification for different threshold values can be graphically illustrated with the receiving operating curves (ROC curve). Figure 4.10 shows the ROC for efficiency classification for usage frequency (left) and energy per wash (right).

The ROC curve (yellow) shows the rate of true positives against false positives. The black dashed line depicts an algorithm that would be performing as good as random guessing. The larger the area between the black dashed line and the ROC curve the better the performance of the classification. The corresponding threshold values for the points of maximum precision (1.), maximum F1 score (2.) and maximum recall (3.) are shown in table 7 (usage frequency) and table 8 (energy consumption per wash). The area for the usage frequency is 0.87. The area for the usage energy efficiency is 0.91.

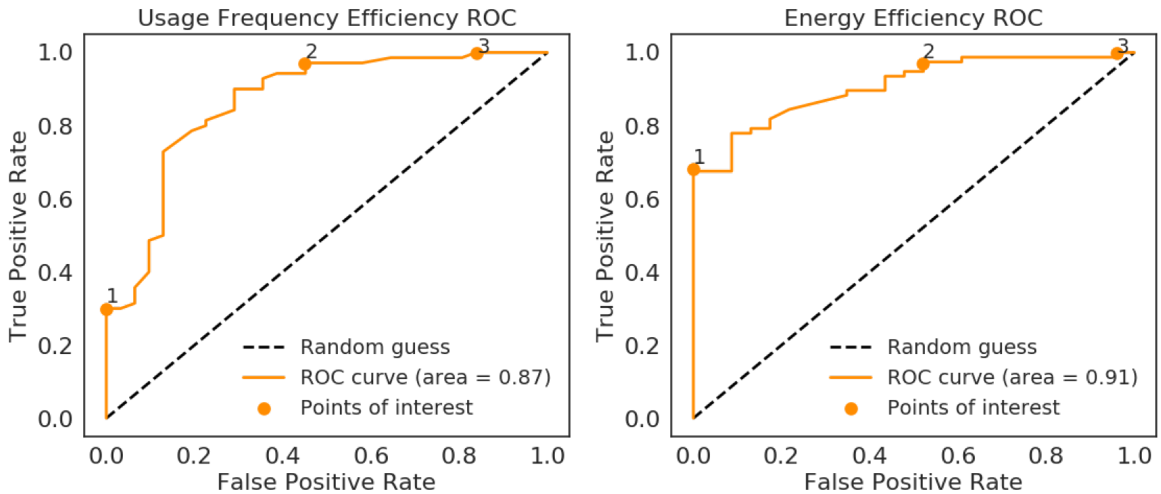

Figure 4.10: *Receiving Operating Curves for dishwasher efficiency classification of weekly usage frequency and energy usage per wash*

4.3 Smart Meter Model Deployment

This section displays the results of the application of the newly developed NILM-based detection system on one full year of data from 130.000 households. The results for the consumption analysis consist of an overview of energy usage and usage frequency averages, followed by usage patterns and dependency categories such as household size and machine characteristics. The last subsection describes how efficiently dishwashers are used in households.

4.3.1 Consumption overview

The summary statistics for energy consumption and usage frequency of the smart-meter detections is presented in table 4.7. The application of the smart meter model for 130.000 households over a full year results in an average energy consumption of 1.18 kWh/per wash and 285 kWh/per year. Regarding the frequency an average of 4.6 washes per week and 240 per year were found.

Table 4.7: *Summary statistics of the energy consumption and the usage frequency detected on the smart-meter data aggregated by different time steps*

	Per wash	Per week		Per year	
	Energy	**Washes**	**Energy**	**Washes**	**Energy**
count [a]	129137	129137	129137	129137	129137
mean	1.18	4.6	5.5	240	285
std	0.18	2.1	2.6	108	137
median	1.17	4.4	5.1	227	267

[a] Number of households

Distribution comparison

As explained in methodology section 3.4.1 a sample of nearly 11.000 of the households answered a survey. To see how this sample compares to the complete group of 130.000 households the distribution of energy per wash and weekly washes is plotted for both. Furthermore, as this relies on model estimated values, a comparison is made with the data from the smart plug measurements for the sample of 100 households. The results of the comparison can be seen in figure 4.11.

The distributions of the different user groups are shown, with "All" representing all nearly 130.000 households, "Survey" those nearly 11.000 who filled out the survey and "Plugs" the data coming from smart plugs for the 100 households used for the model development. It can be seen that for energy per wash are comparable the different data sets show a similar distribution, although not completely overlapping. In contrast to the plugs (red), the detection values show very little observations below 0.9 kWh as a result of the estimation algorithms tendency to over-estimate below the average and under-estimate above the average. Furthermore, it can be noted that the mode for the survey group (green) is slightly higher than for the complete group and the plugs. On weekly washes the survey group shows very high overlap with the complete group. Plugs on the other hand show a considerable higher mode, likely caused by the difference in family size between the Plug group (over 3.3 household members) and the Survey group (just under 3 household members). Finally it has to be noted that the Plug data seems to show a considerably less smooth (multimodal) curve. The probability density curve implies that the observations are spread uniformly across a continuous line. While for the 130.000 and also the 11.000 observations this case can be made, for only 100 observations this is spread less evenly and hence results in a less fluent line.

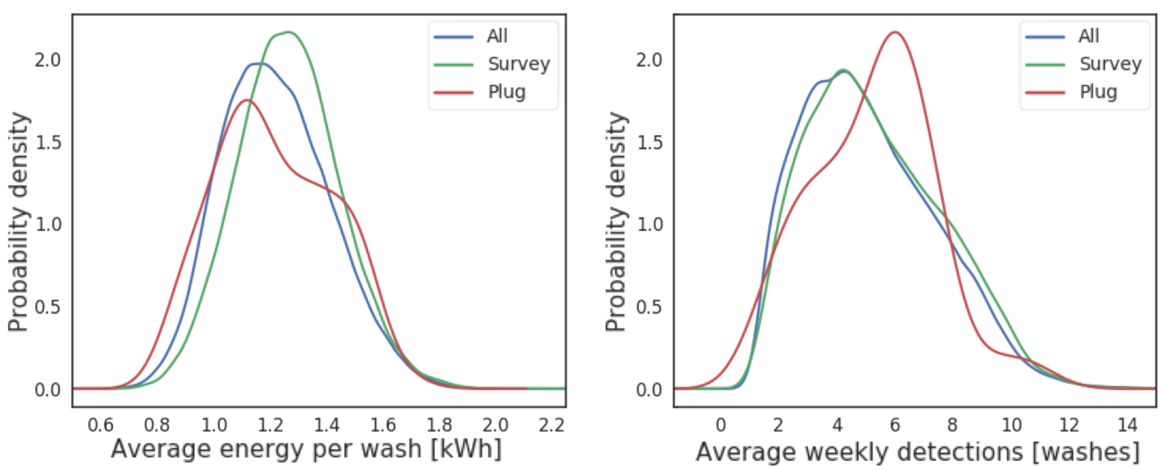

Figure 4.11: *Probability density of the smart-meter, survey and smart-plug detections for the average energy consumption per wash in kWh (left) and the average weekly detections in number of washes (right)*

4.3.2 Consumption category dependency

The survey results can be used to segment the usage frequency and energy consumption per wash for different groups. The results for the categorisation by dishwasher size, household size and average washing temperature are presented.

Effect of machine size

As both energy consumption per wash and number of washes relate to the machine capacity, the dishwasher usage is categorised by the machine size type in table 4.8. Comparing the average detected energy consumption of the machine size "Regular" and "Table" it can be seen that both mean energy consumption are similar to the "Notsure" category at around 1.22kWh. Only the 3% of dishwashers stated to be "Compact" show a mean energy of 1.13 kWh. The mean value for "Compact" is also slightly lower with 4.6 washes per week, compared to 5.0 for "Regular" and 4.9 for "Table" and "Not sure". Both 1.22 kWh and 4.9 weekly washes are slightly above the total for all households (4.7).

Table 4.8: *Aggregation of the smart-meter estimated dishwasher activities by in survey stated machine size type*

	Households		**Estimated energy**		**Estimateded washes**	
Machine size	**count***	**share**	**mean**	**std**	**mean**	**std**
Notsure	58	1%	1.23	0.17	4.9	2.1
Compact	297	3%	1.13	0.18	4.6	2.0
Regular	9578	88%	1.22	0.16	5.0	2.0
Table	940	9%	1.23	0.17	4.9	2.0
Total	10873	100%	1.22	0.16	4.9	1.9

*Number of households

Effect of household size

Next segmentation of washing frequency by household size is shown in table 4.9. As people were also asked to indicate how often they think they wash, both stated and estimated weekly washes are shown. The 1 person households wash an average of 2.8 times a week, much in line with the majority (6%) stating to wash 1-3 times and some (2%) 4-6 times a week, resulting in a weighted average of 2.9 based on the survey results. The next, 2-persons households, on average use their dishwasher 4.4 times (or 4.8 according to the survey), after that for every additional family member roughly 0.6 extra washes are added. This clearly shows not a linear but a reducing increase for every additional family member. The stated washes reveal a similar pattern, although they would indicate a slightly higher estimation of about 0.7 washes per week more than what is estimated. In total 43% of all the households taken together stated 7+ washes per week, higher than the 32% stating to wash 4-6 and 25% stating to wash 1-3 times a week. The detected mean of usage frequency was estimated at 4.9 washes per week, compared to 5.6 for all households according to survey results (table 4.7).

Table 4.9: *Household size versus stated washes and detected washes per week*

	Household size					
	1	2	3	4	5	Total
Households						
count*	816	3713	2247	3083	1014	10873
share	8%	34%	21%	28%	9%	100%
Survey stated washes						
1-3	6%	12%	4%	3%	1%	25%
4-6	2%	13%	7%	8%	2%	32%
7+	0%	9%	9%	18%	7%	43%
mean**	2.9	4.8	5.6	6.6	7.1	5.6
Model estimated washes						
1-3	6%	13%	5%	4%	1%	28%
4-6	1%	17%	11%	14%	4%	48%
7+	0%	5%	5%	10%	4%	24%
mean	2.8	4.4	5.0	5.7	6.3	4.9

*Number of households **weighted average based on shares per frequency range, counting 1-3 as 2, 4-6 as 5 and 7+ as 8

Effect of label and temperature

Lastly, the energy per wash estimations are categorised by efficiency label and washing temperature as presented in table 4.10. The largest share of households own a machine with A++ label. Of the total 10780 households 33% use an A++ machine, followed by 29% households in the A+ category, 25% households in the A- category and lastly 12% households with a dishwasher labelled A+++. Independent of the given efficiency label, most (55%) households stated to use the temperature range between 50-75 Celsius and another 6% said to use an even hotter program most often.

When looking at the estimated average energy consumption in kWh it can be seen that in each efficiency category the mean energy consumption rises with lower to higher temperatures. Overall a difference of 0.14 kWh can be noted. Besides of 1 irregularity (A+++, 30 lower) also the average energy consumption for each respective temperature range decreases with increasing efficiency label.

Table 4.10: *EU efficiency label and washing temperature in degrees Celcius stated by the survey participants versus detected average energy consumption in kWh*

		Households		Estimated energy	
Efficiency label	**Washing temperature**	**count***	**share**	**mean**	**std**
A-	30 lower	17	0%	1.18	0.16
	30-50	804	7%	1.21	0.16
	50-75	1664	15%	1.30	0.15
	75 higher	205	2%	1.35	0.17
A+	30 lower	33	0%	1.10	0.19
	30-50	1159	11%	1.18	0.16
	50-75	1804	17%	1.26	0.15
	75 higher	172	2%	1.31	0.16
A++	30 lower	65	1%	1.10	0.16
	30-50	1505	14%	1.15	0.16
	50-75	1837	17%	1.22	0.16
	75 higher	171	2%	1.27	0.16
A+++	30 lower	42	0%	1.13	0.17
	30-50	597	6%	1.14	0.15
	50-75	652	6%	1.20	0.16
	75 higher	53	0%	1.23	0.14
Total		10780	100%	1.22	0.16

*Number of households

4.3.3 Consumption time dependency

In this section the time dependency of the energy consumption per wash and the usage frequency over the course of a year are presented. Unlike in the previous section on the category comparison, here data for all households is considered.

Effect of outside temperature

Ambient conditions were identified as one of the factors potentially affecting real-life dishwasher usage (see literature 2.4.3). When plotting the energy consumption per wash for each week, averaged for all households, a seasonal pattern was identified. This pattern was compared with the average outside temperature for each week.

In figure 4.12 the average energy consumption per wash in kWh on the left vertical axis is plotted for each week in 2018. On the right vertical axis the average outside temperature in Celsius is given. In addition to the mean energy consumption per wash (in black), the left graph depicts the distribution of energy in each week (light blue). It appears to not be fully equal on both sides of the graph, skewed to the right, in accordance with figure 4.11.

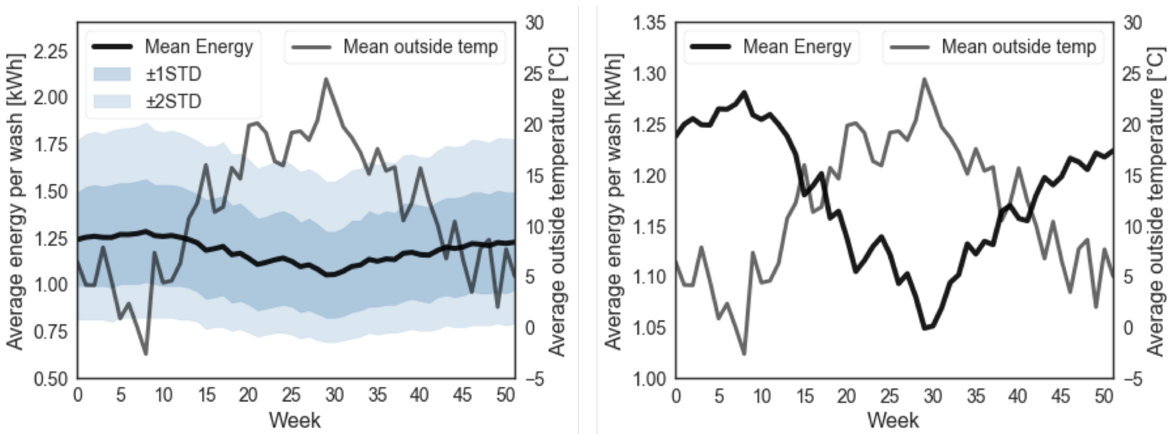

Figure 4.12: *Energy consumption per wash over the year vs the respective average outside temperature in °C with standard deviation (left) and without standard deviation and different average energy consumption scale (right)*

In the right graph in figure 4.12 the same graphical set up is displayed without the standard deviation bands and a changed range on the left vertical axis. By this change, the pattern of the time dependant average energy consumption fluctuation can be seen more clearly. It can be seen that the average energy consumption is decreasing in the summer months with a minimum of 1.05kWh in week 30, which is the end of July in 2018, when the average outside temperature was 24.4°C. The average energy consumption per wash increases in the winter months with a peak of 1.28kWh in week 9, which is the end of February / beginning of March, when the average outside temperature was -2.7°C.

The mean energy consumption reacts in the opposite direction of changes in the mean outside temperature with a Pearson correlation of -0.94, cross-correlation of -0.92 and lag of 0. When regressing the average energy consumption against the outside temperature, it was found that the average energy consumption at 0°C outside temperature is 1.28kWh and that the average energy consumption decreases with -9.1Wh/°C. In line with an average annual energy consumption of 1.18kWh at an average annual outside temperature of 11.4°C found for 2018.

Effect of annual events

Following the analysis of the seasonality on the average energy consumption, the usage frequency can be assessed respectively. In figure 4.13 the average weekly usage frequency (black line, left axis) is plotted for each week in 2018. On the right vertical axis the number of detected users (in 1000 households) (blue bars) for the same period is plotted. Getting a better understanding for these patterns can both be relevant for energy consumption forecasting and to understand how usage frequency and thus measurement compared to an efficiency threshold can differ throughout the year.

As detected for the average energy consumption also the usage frequency shows a seasonal dependency. On average, in the Winter months more washes per week are estimated with a maximum of 5.3 in the first week of January, compared to a minimum of 3.3 in week 31 (end of July/beginning August). This trend can be seen both for the estimated usage frequency as well as for the detected active households.

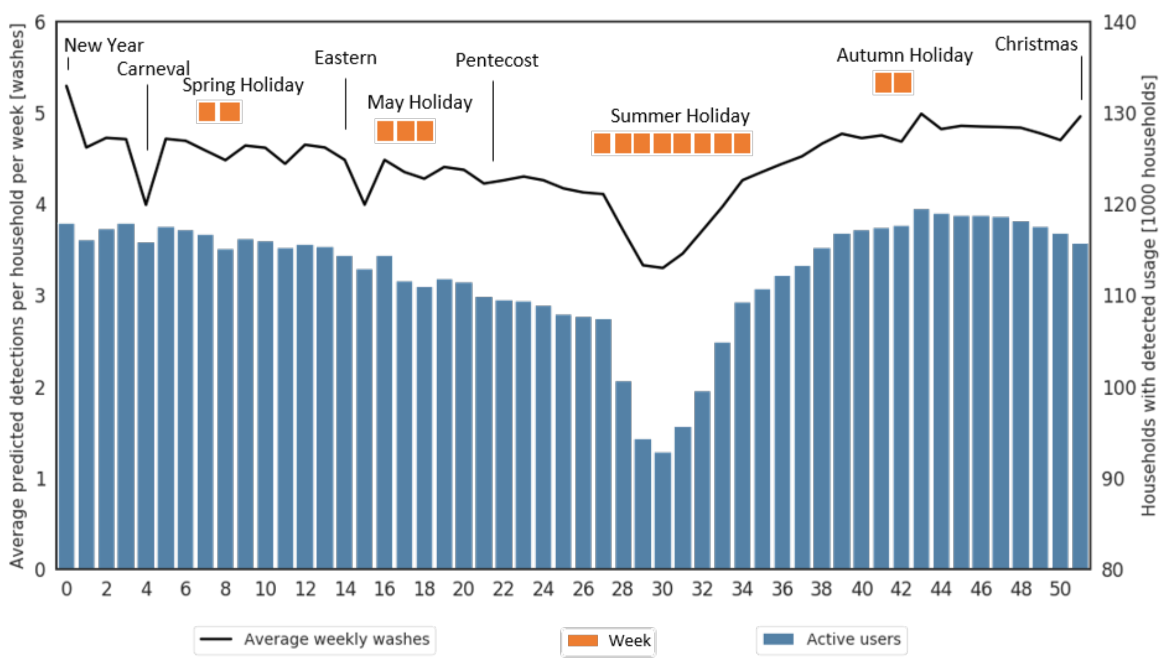

Figure 4.13: *Estimated average weekly usage frequency in washes over the year vs number of detected users in number of 1000 households*

When looking at the fluctuations more carefully the differing average washes per week per household over the year can be related to public holidays and other events. A list of all holidays in 2018 can be found in appendix table E.1. In weeks 51 and week 0 the effect of holidays around Christmas and New Year can be seen. Around week 4 (end of January) the Carnival weekend took place. The Spring break in end of February created a slight decrease in the usage frequency. The dip in week 15 might be related to Eastern the weekend before, which might be extended together with festivities like Kings day (27th April), Liberation day (May 5) and Ascension Day (May 10) falling into the May vacation. Note the actual May vacation is officially only in week 17, however most provinces in the Netherlands extend it by 5 days before or after. The clearest impact can be seen during the summer break period, where many people take summer holidays and go abroad. After the public school Autumn holidays between week 41 and 42 (mid until end of October) an upwards trend can be detected.

Weekhourly usage pattern

In addition to the yearly time dependency also a daily time dependency of the dishwasher activity can be recognised. On the vertical axis of figure 4.14 the share of the dishwashers activated in % of the weekly washes is given. On the horizontal axis the start time of a washing cycle (according to the detection algorithm) is shown. Each bar then represents what share of all washes for all households over the full year were started for each hour of the week.

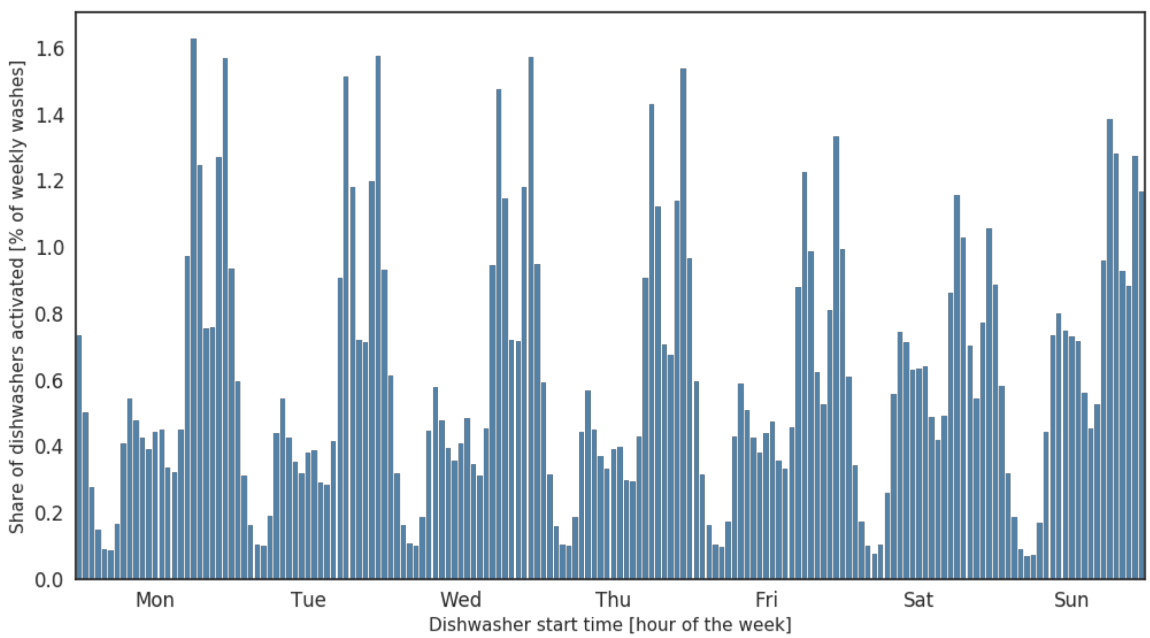

Figure 4.14: *The share of dishwashers activated in % of weekly washes plotted against the start time of dishwashers (based on detection algorithm) in hours per week plotted*

The bar chart displays a repeating pattern of different magnitude over the course of the week. As can be seen it repeats every day with two high peaks in the evening. When considering the distribution between days, a downwards trend for the evening peaks can be seen from Monday onward, with Saturday depicting the lowest peak. On Saturday and Sunday a more even pattern throughout the day can be noticed, with relatively higher peaks spread over the morning hours.

Figure 4.15 shows the single day pattern in more detail, where each bar the hours for all days of the week are averaged into a single typical day. Two main peaks can be noticed around early evening after dinner and late night activity just before midnight. A consistently strong dip can be seen each night, with starting times rapidly decreasing after midnight. The increase in the morning is followed by a smaller but noticeable peak after lunchtime. When looking at the week pattern it can be noticed this pattern is more distinctive during the week than the weekend.

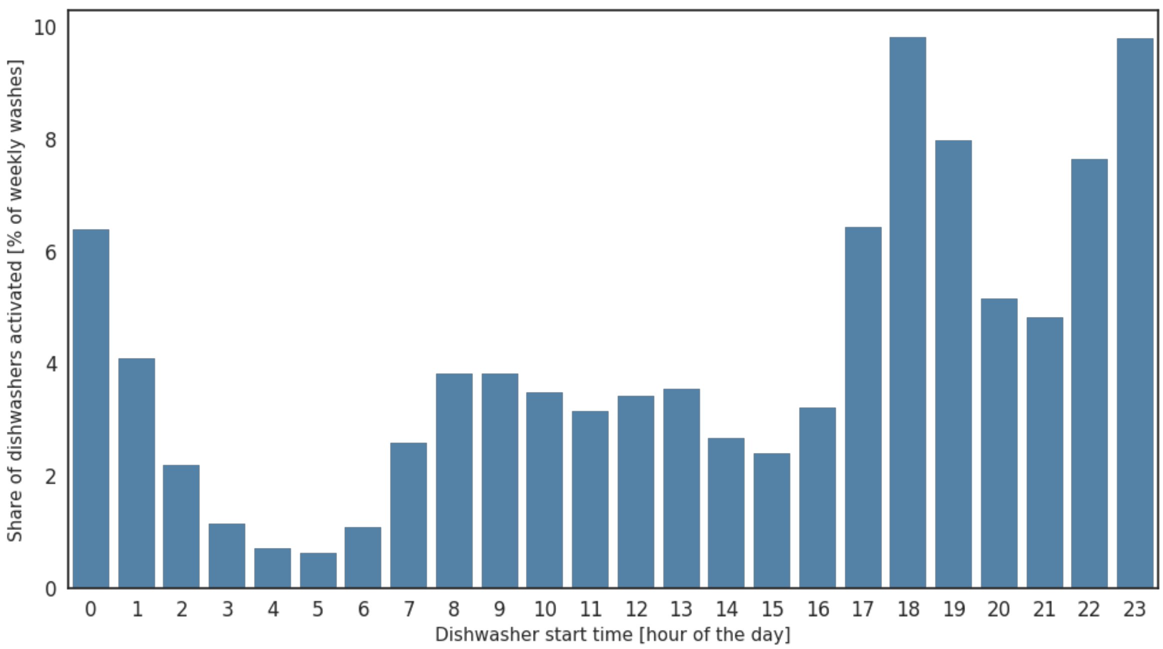

Figure 4.15: *The share of dishwashers activated in % of daily washes plotted against the start time of dishwashers (based on estimation algorithm) in hours per day plotted*

4.3.4 Efficiency analysis

The final step is to determine how efficient households use their dishwashers. The binary efficiency classification (see methodology subsection 3.4.3) was applied for the nearly 11.000 households who filled out the survey. For both energy per wash and weekly usage, the classification based on the meter estimated values related to the set threshold are compared to a classification based on the results from the user survey. An estimation is made of how much households could potentially save on their annual dishwasher energy consumption.

Energy efficiency

In figure 4.16 the distribution of estimated energy consumption is shown for different efficiency labels. For each efficiency label the distribution by most common stated washing temperature is shown. The average energy consumption for each efficiency label is plotted with the dashed black line and can clearly be seen increasing stepwise with decrease in efficiency label. The average energy consumption per temperature range is plotted with a white dot, which shows an even steeper increase within each label than the difference in energy consumption between the labels. This visualises the earlier findings from table 4.10, where the average energy consumption and number of households for each category were shown.

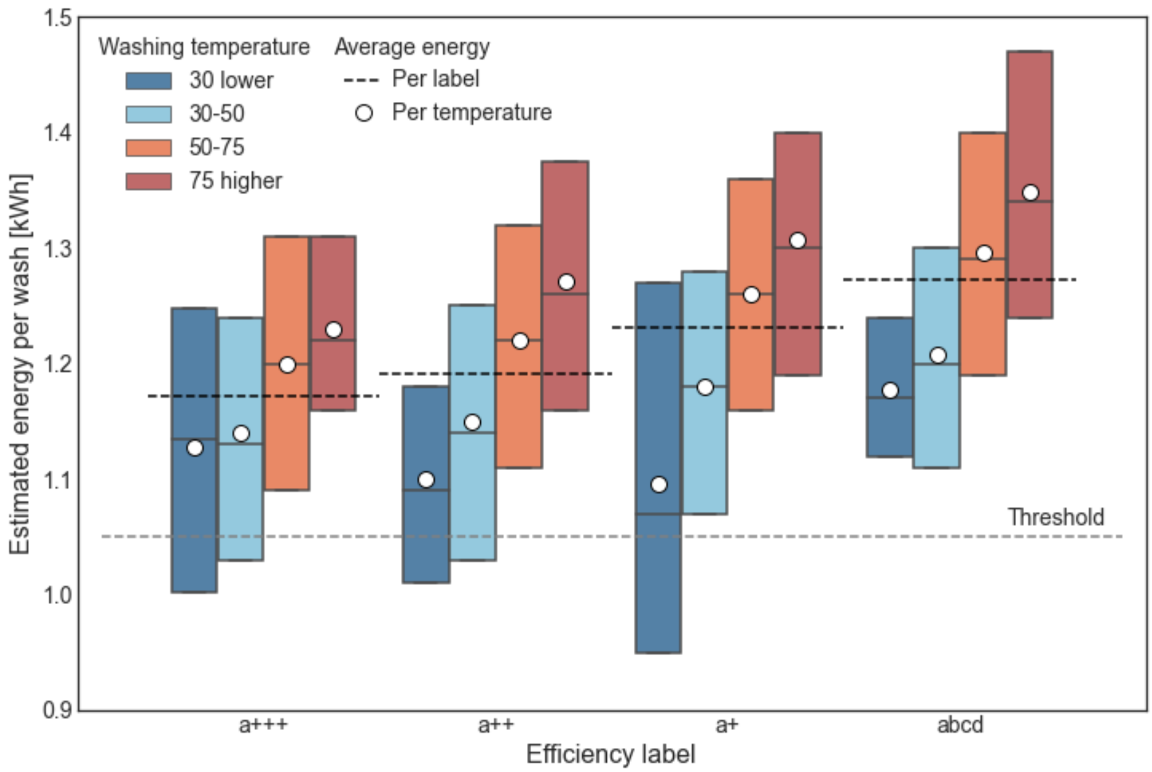

Figure 4.16: *Distribution box (50% interval) of estimated energy consumption per wash by efficiency label and washing temperature*

The grey dotted line running from left to right shows the established generic threshold, that is used as proxy for efficient energy consumption per wash. The same threshold of 1.05kWh was used for which the classification performance was earlier analysed in subsection 4.2.3. It appears that the majority of the shown distributions are above this threshold also visible through the averages per label and temperature, which are above as well. However, placing more emphasis on the change of the distribution in relation to each other, only the 50% interval of the boxplots (the boxes) are shown, excluding the full length of the the 1st and 4th quartiles (the whiskers). The majority of the 1st quartiles among the higher efficiency labels and lower washing temperatures do largely or at least partially fall under the threshold. To see the full boxplots, including whiskers see appendix figure E.2

The estimated energy for these households could then be compared to the binary energy threshold to classify which households are inefficient. The results from the user survey for efficiency label and washing temperature were used to classify which households state to have inefficient dishwashers or use more energy per wash than needed due to their chosen washing program. Households that stated their dishwashers to be below A+ and/or washing warmer than 50 degrees were classified as inefficient. Figure 4.17 shows this comparison.

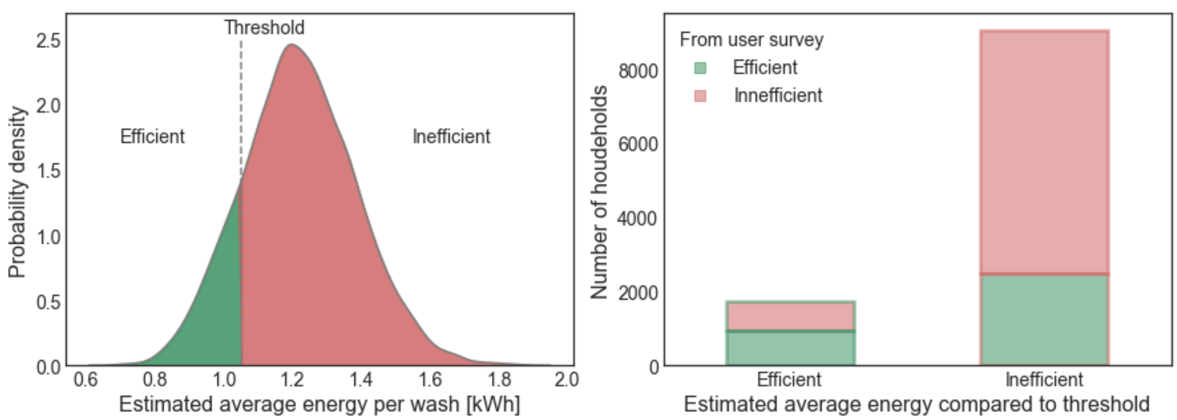

Figure 4.17: *Distribution of households with efficient and inefficient estimated average energy per wash (left), compared to the classification based on results from user survey (right)*

The distribution of energy per wash for all households (left chart) is divided by the set threshold into 16% efficient (green) and 84% inefficient (red). The two bar charts (right figure) are the number of households classified as efficient (green line) and those as inefficient (red line). These two bars match with the green and red area in the left figure and the number of households that are below and above the threshold in the previous figure (3.4.3). The bar charts are then segmented by the classification they received according to the survey results. A share of 8.6% of households (green, left bar) are classified efficient by both the meter estimation and survey approach, while 7.4% (red, left bar) are classified as efficient according to the meter, but not based on the survey results. In contrast 61% of households (red, right bar) are classified as inefficient according to both the meter and survey, while 23% (green, right bar) are classified inefficient by the meter, but not from the user survey. This results in a precision of 0.73, recall of 0.89 and weighted accuracy (F1) of 0.80.

Usage frequency efficiency

A similar figure can be created for the distribution of the average number of weekly washes by household size and dishwasher type. The average number of weekly washes by household (shown with the black dashed lines in figure 4.18) indicate a stepwise increase with each additional household member. As discussed earlier in relation to table 4.9 there is a bigger jump from 1 to 2 people in number of weekly washes and then an increase of about 0.6 extra washes for each additional household member. While the type (hence size) of dishwasher could potentially influence the number of weekly washes, the results (average shown with white dots) do not show much variation. A single outlier seems to be the compact dishwashers used in 5+ sized households. However, as could be seen in table 4.8 only 3% of households stated to have a compact dishwasher. Even a smaller share applies for larger households. Making this rather an outlier due to the low number of data points.

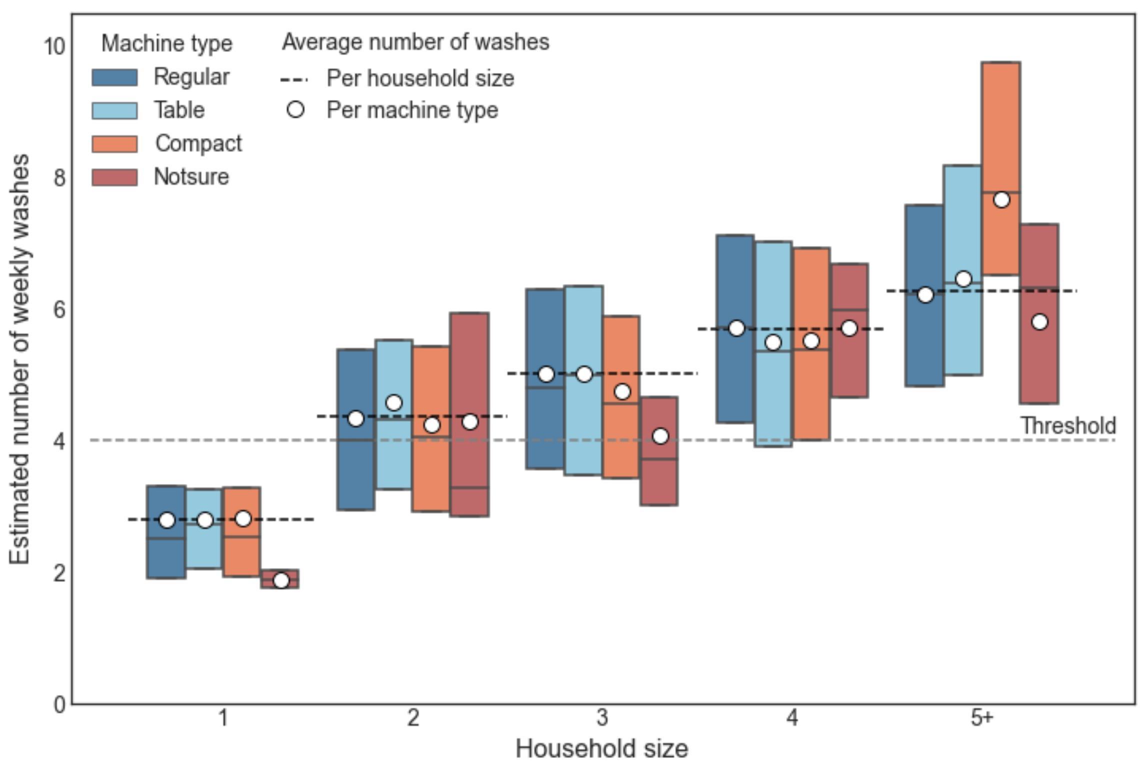

Figure 4.18: *Distribution box (50%) of estimated number of weekly washes for different household sizes and by dishwasher machine type*

Similar to the earlier figure a grey dotted line running from left to right shows the established generic threshold, in this case the proxy for efficient number of weekly washes. The same threshold of 4 washes was used as earlier analysed in subsection 4.2.3. While also in this figure only the boxes are shown (for plot including whiskers see appendix figure E.3) the majority of single member households are below this threshold, while most 4 and 5 member households are above and most 2 and 3 member households are closely around this threshold.

In a similar way, the estimated weekly washes can be compared to this binary frequency threshold of 4 washes per week. The number of weekly washes as stated in the user survey were used to classify which households use their dishwasher more then 4 times per week. Figure 4.19 shows this comparison.

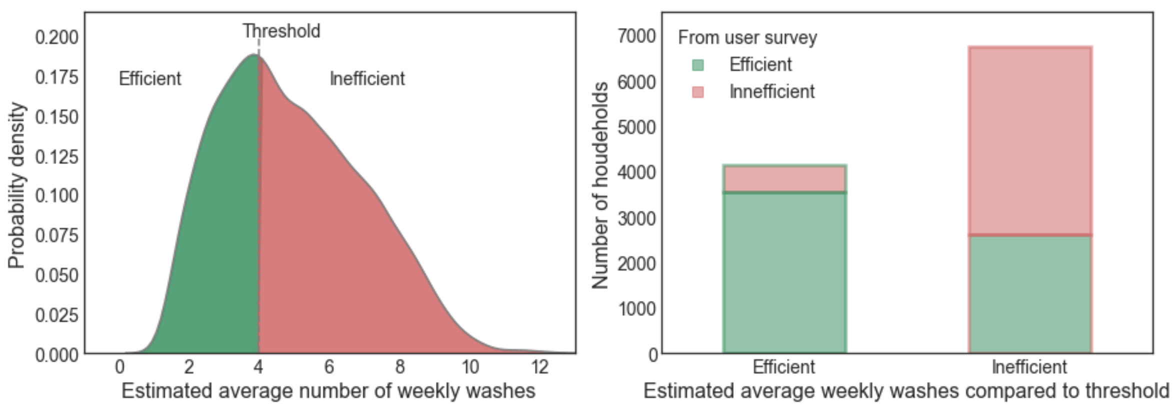

Figure 4.19: *Distribution of households with efficient and inefficient estimated average weekly washes (left), compared to the classification based on results from user survey (right)*

The distribution of weekly washes for all households (left chart) is divided by the set threshold into 38% efficient (green) and 62% inefficient (red). The two matching bars (right figure) are divided by a share of 32.6% of households (green, left bar) classified efficient by both the meter and survey, 5.5% (red, left bar) classified as efficient according to the meter, but not by the survey. Around 38% of households (red, right bar) are classified as inefficient according to both the meter and survey, while 24% (green, right bar) are classified inefficient by the meter, but not the user survey. This results in a precision of 61, recall of 87 and weighted accuracy (F1) of 72.

Relative efficiency

As seen in the previous subsection (figure 4.16 and 4.18), there is a variation in energy usage and number of weekly washes between categories. Efficiency label, washing temperature and household size were shown to affect this. Figure 4.20 (below) therefore segments the findings from figure 4.17 and figure 4.19 by category.

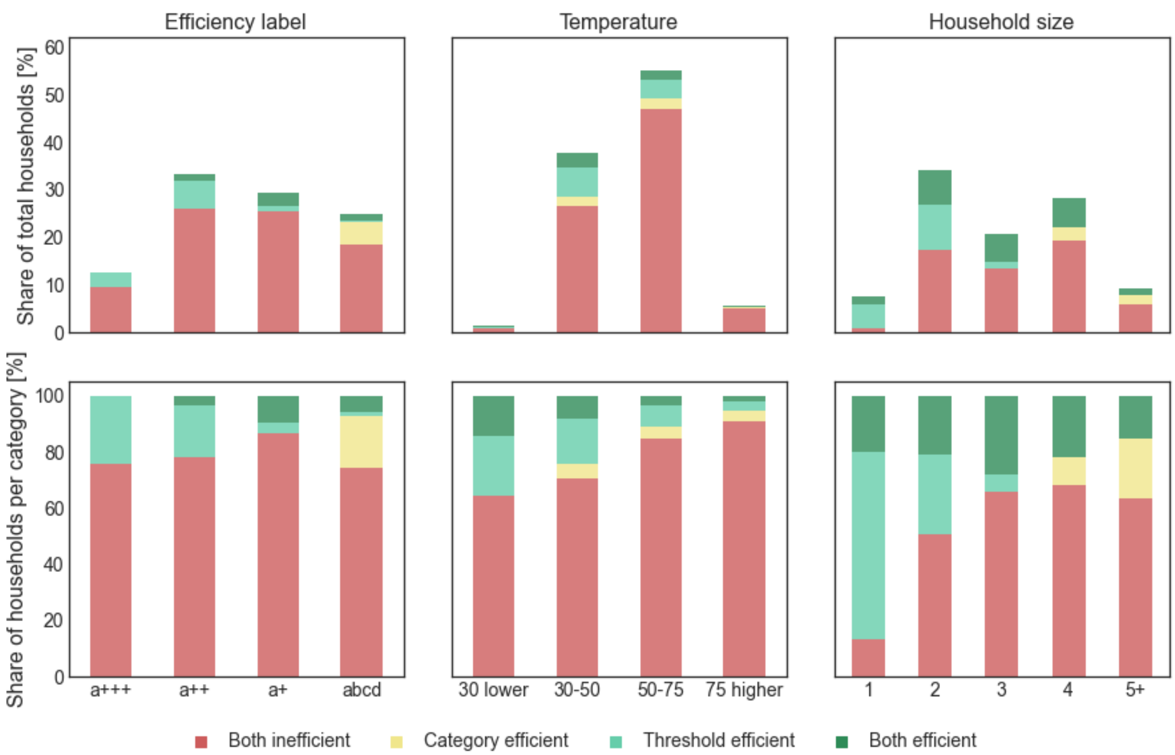

Figure 4.20: *Share of households at different type of efficiency (generic, specific or both) compared for machine efficiency label, washing temperature and household size*

For both efficiency label and temperature estimated energy consumption based on detection, is compared with the energy consumption specific for that category and compared to the generic threshold. For household size the same is done, but for the number of washes based on family size. When the energy consumed per wash or the number of weekly washes respectively is under both the generic threshold and the specific threshold applicable for that category, then the household is 'both efficient'. Hence, it is found the household uses its machine efficiently according to the binary classifier and the actual results. This (dark green) applies for only a relatively small share of households (5-10% depending on specific category).

When the household's usage is both above the generic and specific threshold it is found to be 'both inefficient'. This (red area) applies for most households (70-80% depending on category). The light green and yellow areas are where the binary classifier finds a different outcome than a specific classification. When classified as 'threshold efficient' the consumption (light green) is below the set binary threshold, but the user actually consumes more than expected in their category. This applies especially for more efficient machines and smaller households, because they are under the generic threshold, but still within their category might not be as efficient. Finally the 'category efficient' classified households consume more than the set binary threshold, but within their category would actually be efficient. This applies for some of the abcd labelled dishwashers and for larger family sizes, where a less efficient machine uses more energy than the set energy threshold, but still could be seen as efficient within that category and larger households would more often use their dishwasher.

Potential energy savings

To summarise these findings and indicate how much energy households could potentially save on their dishwasher, the two generic thresholds were combined and compared to the annual energy consumption. Households are segmented based on their indication of potential inefficiencies, to show how much they could potentially save on their annual energy consumption.

Based on the 1.05 kWh per wash and 4 weekly washes an annual energy threshold of 218kWh was established. Figure 4.21 shows the distribution of annual energy consumption for all households, oriented vertical, in the right chart. Similar to the average energy per wash and weekly washes, the households are divided in those (29%) below (green) and those (71%) above the threshold (red). With an average annual energy consumption of 312kWh this provides 94kWh of energy saving potential on average, a potential reduction of just over 30%.

In the left chart the households are divided into potential energy saving categories based on their supposed inefficiencies. The first dark green violin distribution on the left shows the 7% of households that have efficient labelled dishwashers, make use of a low energy consuming washing program and wash less than 4 times a week, by that using 141kWh on average (24.5% below the threshold). The next group (9% of households) is called 'Frequency efficient' as their weekly usage is below 4, but their energy consumption per wash is above 1.05kWh. Nevertheless, with 164kWh they consume 17% below the threshold. The households (1%) that are 'Energy efficient' (lightest green) wash on average below 1.05kWh, but more than 4 times a week, resulting in an energy consumption of 204kWh. This compared to those (10%) 'Energy inefficient' (lightest red) who wash over 1.05kWh, but despite washing less than 4 times a week, with 247kWh still use (9%) more than the set threshold. For those (20%) who wash more than 4 times a week, despite efficient energy usage per wash, results in an average consumption of 321kWh (33% over the threshold). Those (52%) who use more energy per wash and wash more than 4 times a week on average, use 408kWh per year, nearly 61% over the threshold.

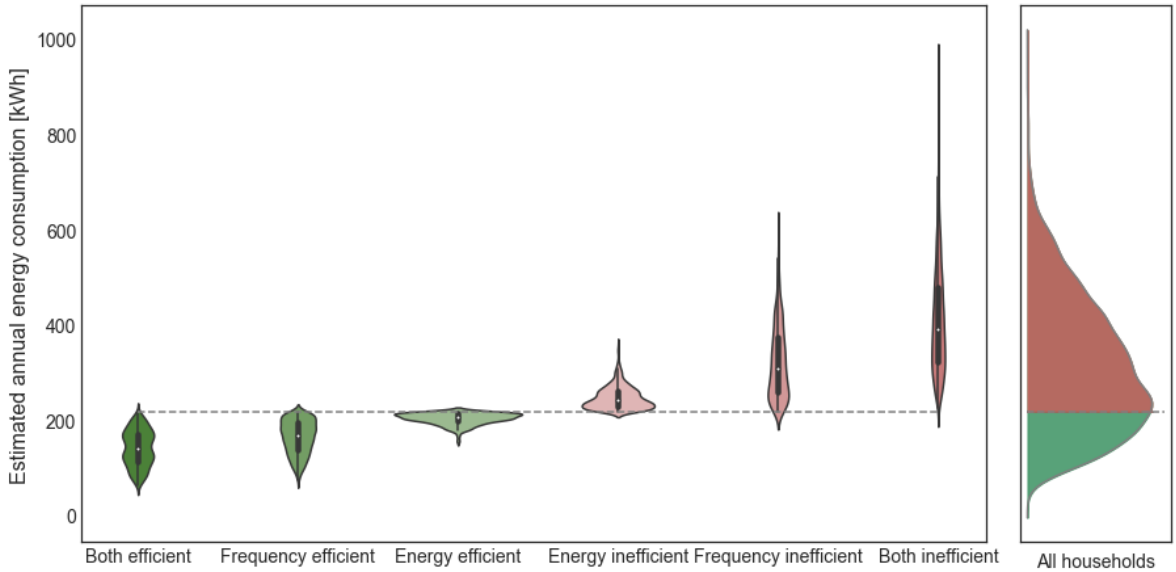

Figure 4.21: *Distribution of annual energy consumption per household by efficiency classification (right) and segmented by inefficiency categories (left)*

4.4 Results conclusion

The results chapter showed the two main parts of this research, analysis of the developed model and analysis of the results from its deployment. To analyse the performance of the model, it also had to be compared to the outcomes from analysis of the plugs, as a benchmark. Below research question 3, 4 and 5 are answered. Research question 3 is completely answered by the performance of the model. Research question 4 and 5 were partially already answered with the plugs, but further extended with insights from the large scale deployment on meter data.

3. How does the newly developed system perform in detecting these (energy consumption, usage pattern and efficiency) characteristics?

Energy estimation
From the comparison of the estimation methods based on smart meter data as described in section 4.2.1 it can be concluded that the best functioning approach to estimate the energy is the regression of the heating energy and number of heating moments (see section 4.2.1) resulting in an error (RMSE) of 0.10kWh and relative error (NRMSE) of 8.8%.

Frequency estimation
The comparison of the estimation methods for weekly washes (section 4.2.2) indicates, that the best functioning estimation approach is a linear regression adjusted to estimate closer to zero, showing an error (RMSE) of 1.4 washes/week and relative error (NRMSE) of 27.2%.

Energy efficiency classification
When applying a generic energy efficiency threshold of 1.05kWh on the energy values that were estimated by the smart meter model, a classification accuracy of 0.89 (F1 score), a precision of 0.88 and recall of 0.90 was found.

Frequency efficiency classification
For the generic usage frequency efficiency threshold, set at 4 washes per week, a precision of 0.87, a recall of 0.90 and a F1 score of 0.89 were found. In contrast to the relative error (NRMSE) of the usage frequency estimation, the usage frequency efficiency classification shows very similar accuracy measures compared to the energy per wash.

4. How much energy do dishwashers in households consume and how often are they used, depending on time, household and machine characteristics?

How much energy do dishwashers in households consume on average per wash how frequently are they used per week and per year and what is the resulting total energy consumption?

In the preparatory smart plug analysis it is was found that the washes of the 100 smart plug users average at 1.22kWh, ranging between 0.86kWh and 1.69kWh with an error of the mean of 0.02kWh. As shown in table 4.1 the energy intensity of about 80% of a wash can be attributed to the heating phases of the washing cycle. On an average wash of 85 min this energy is used in the 30 minutes needed for on average 2.7 separate heating phases. In the preparatory smart plug analysis, an average usage frequency of 5.2 washes per week was found. The NILM system deployed for the large smart meter group revealed an average energy consumption of 1.18 kWh/per wash, 4.6 washes per week, resulting in 240 washes and 285kWh per year.

How does the energy consumption of dishwashers depend on household and machine characteristics?

The majority of households in the conducted smart meter survey (88%) stated to have a "Regular" machine. For these households an average energy consumption of 1.22 kWh/per wash and 5.0 washes per week were revealed. For households which indicated to have a "Table" machine or did not specify the machine type similar values were found. For the (3%) households indicating to have a "Compact" dishwasher an average consumption of 1.13kWh and 4.6 washes was detected.

One person households were found to wash an average of 2.8 times per week and two people households wash 4.4 times on average. For every additional family member the weekly washes increase roughly 0.6 on average.

With 33% the A++ was the most common efficiency label. Most households (55%) stated to use the temperature range between 50-75 Celsius. The temperature of the chosen washing program impacts the energy consumption more than the efficiency label. High efficiency label and low temperature result in an average energy consumption of 1.13 kWh (for A+++, <30°C). Low efficiency label with high temperature (label A and below, <75°C) result in an average energy consumption of 1.35 kWh.

How does the energy consumption of dishwashers in households depend on time of the day, week and year?
It was shown that the average energy usage varies throughout the year. The energy pattern shows a strong negative correlation (-0.94) with outside temperature. It was found that the average energy consumption at 0°C outside temperature is 1.28kWh and that the average energy consumption decreases with -9.1Wh/°C, dipping during a heatwave in the Summer to 1.05kWh.

With 5.3 washes, the number of washes peaked in the first week of January and reached a minimum in the middle of the summer holiday at 3.3. Both of these extremes also coincided with holiday events, which is exemplary of the effects events can have on the usage pattern throughout the year.

Within the weekly pattern, dishwasher usage differs per weekday with least usage on Friday. Main usage was identified to be directly after dinner time and just before bed time, with the highest peak on Monday after dinner and a more equally distributed use over the weekend.

5. How efficiently are dishwashers used in households?

When using the general energy efficiency threshold of 1.05kWh, 78% of dishwashers used by smart plug users classified as inefficient. The analysis of the large scale smart meter data shows that 16% of the households are classified as efficient and 84% are classified as inefficient.

The usage frequency threshold was set to be 4 washes per week, whereby the frequency of dishwasher usage was classified as inefficient for 69% of the smart plug users. The results of the efficiency classification for weekly number of washes for the large scale smart meter deployment resulted in 38% of the households to be classified as efficient and 62% as inefficient.

It was found that 29% of households use less energy than the set annual threshold of 218kWh, while the other 71% used more. With an average annual energy consumption of 312kWh this provides 94kWh of energy saving potential on average, a potential reduction of just over 30%. Only 7% of all households have efficient labelled dishwashers, make use of a low energy consuming washing program and wash less than 4 times a week, by that using 141kWh on average (24.5% below the threshold). Consuming more energy per wash (10%), washing more frequently (20%) or both (52%) results in an average of 9%, 33% and 61% overconsumption respectively.

5

Discussion

5.1 Discussion introduction

The first part of the discussion (5.1) focuses on the interpretation of the model performance results, pitfalls and validity of the used approach and the generalisability of the model. The second part (5.2) discusses the application of the model in real life. Hence, less focus is put on the applicability of the model in the first section of the discussion and more on its validity. In the second part, the discussion focuses more on the interpretation of results and what can be generalised from them about dishwasher usage in real life. In contrast to the previous chapters no separate conclusion to the discussion will be given as the next chapter (6. Conclusion) already functions as its concluding remarks.

5.2 Smart Meter Model Development

In this section the performance of the developed smart meter estimation model is discussed. The model was developed to detect energy usage, usage frequency and efficiency for dishwashers from the smart-meter signal, based on and validated with a data set of smart plug data. As these two main data sets are underlying all results, the data preparation is discussed first. The rest of this section is structured in line with the subcomponents of the meter estimation model results section. For each of these subsections an interpretation of the main findings is given. Because this section focuses on the development of the model, emphasis is placed on the validity of the research and hence less connection with existing literature can be generalised. Deployment of the model is then discussed in the next section, where more connection can be made to existing literature.

5.2.1 Data

The model was developed with a learning set consisting of the smart meter data from 100 households measured over three months. Data from smart plugs installed in the same 100 households was used to calibrate the model and analyse its performance. During the gathering and processing of the data various factors were found relating to the validity and generalisability of the data. These will be discussed subsequently for the smart plug (5.2.1) and the meter data training set (5.2.1).

Smart plugs

Essentially the plug data serves as a "ground truth" for the research, being the basis for the validation of the developed meter model. However, also this baseline data set is affected by aspects regarding its validity and generalisability. In particular the implications on the data gathering process, the choice of households and the data processing with the Basu algorithm will be discussed in more detail.

Data gathering
When discussing the applicability and quality of the plug data it is important to consider, that the data was initially gathered for a different experiment, hence the data was not gathered within the control of this research. In the previous experiment it was decided to sent the plugs to participants instead of install them for them in order to reduce operational burden and invasiveness. As a downside, human error or technical fault resulted in unavailable or faulty plug signal in several cases. Consumers not installing the plugs at all, installing the wrong appliance as well as misnaming or switching the plugs during operation to another appliance are additional factors that

may have occurred. Furthermore, the smart plugs provided to the consumers differed by brand, potentially impacting the functionality and resulting in variation in measurement accuracy. Finally, the plug measurements were not calibrated and tested on beforehand against user manual stated values for the energy consumption of each (or at least several) dishwasher's used washing programs.

Compared to a controlled environment, this field experiment data may include uncertainties and variances, that could not be accounted for within the frame of this research. To reduce this as much as possible, the data set was cleaned and filtered and outliers were visually inspected. Overall the chosen data gathering and processing approach led to an exclusion of 30 of the 130 households, of which data could not be well enough used for the analysis. The relatively small number of subjects (100 households) in the sample affects the diversity of what the model has been developed on and hence its generalisability, but nevertheless offers a larger data set than most previous monitoring research.

From the supplementary survey, conducted along with the smart plugs (see results shown in Appendix C.0.1 it can be concluded that the people in the field experiment were not a representative sample of consumers. The average household size of 3.3 lies above the Dutch average of about 2.2 people (see Table C.1). The sub-selection of households that filled out the survey (40%), showed certain similar characteristics such as type of machine and household size, indicating potential bias. The group seems skewed to more efficient and newer dishwashers with 67% of the households machine efficiency label of A+ or higher and 64% of dishwashers less than five years old, compared to an average replacement age of about 13 years (see 2.3.1) and A+ being categorised as efficient by Milieucentraal (see section 3.3.3). This phenomenon might occur when considering that costumers open and interested to partake in the experiment, might already be interested in energy saving management and have therefore taken steps to optimise the appliance efficiency by themselves. Since the data set is relatively small in statistical terms and potential bias could be identified, an important factor to take into account is that this potential bias from the plug validation data could have been carried over into the machine learning meter model, when deployed at large scale.

Data processing

The smart plug data was prepared with the feature extraction method of K. Basu to be able to compare the measurements with the smart meter results. This process introduces additional uncertainties into the the data quality, but also allowed for an additional filtering and cleaning to the data. Generally applying this process adds its own detection error, affecting the validity of the processed data.

To be able to calculate the energy per wash, a constant time interval was needed. The signal was received in a 10 seconds interval, however slight discrepancies occurred in the incoming signal. Therefore the data was resampled to exact 10 seconds, meaning that a redistribution of power between previous and the next timestamp was made. This resampling step slightly reduced the precision. However, considering a washing time of at least 60 minutes, would result in at least 360 timestamps per wash, making the loss of precision negligible.

To analyse individual washes, thresholds had to be set to be able to cut out the washes from the continuous signal. The rules were set arbitrarily, based on visual inspection. This could have lead to washes that were not considered or certain parts were cut off and not used in the analysis. To validate the load profiles and reduce the impact of false or varying activity, samples of power profiles were visually inspected in more detail for each month over a 3 month period for each household.

While the smart plugs supposedly detect every wash, visual inspection showed certain edge cases where the plug data is not perfectly labelled either. An example for these cases can be given by the case of consecutive washes where multiple washes are recognised as one single wash. As such situations show up as outliers in terms of energy consumption, these were visually inspected and manually relabelled. Another example could be major signal breakage. While most smaller signal breakages were not picked up on, visual inspection showed they often occur, although not affecting the count of number of washes and only having a very minimal effect on the overall measured energy. However, in cases of larger signal breakages these do show up as outliers and were excluded for energy calculations. A level of uncertainty remains as to whether all such cases were found and correctly labelled. This uncertainty may cause some additional noise in the data.

While such noise may have had only relatively small impact on the total energy per wash and detected number of washes, the effect on some features was more considerable. Visual inspection showed duration attributes for plug data are usually matched with a discrepancy of no more than 5 minutes and often only a minute or less. On the total average duration of a wash of 85 minutes, this causes a discrepancy of often no more than 5% and the energy usage lost at the start or end of a wash is also very minimal. On an average of 30 minutes of total heating periods, distributed over an average 2.7 heating moments (4.1), this could lead to discrepancies of more than 20% in some cases, although 5-10% in most appeared much more common.

The heating duration was therefore visually inspected in more detail. It was found that larger signal breakages can cause a larger heating period to be interpreted by the detection model as two short heating periods, but more often two separate distinct high power periods could be interpreted by the detection algorithm as one, without noticing a period of low power in between. Based on the visual inspection a minimum of 5 minutes was set to count a heating moment and a gap of more than 5 minutes was set as minimum separation to divide two heating moments as separate events. While some noise might remain regarding the measurement of the heating duration, the correlation of 0.90 shows the effect on the estimation model would only be minimal.

For the heating power, however a bias was identified. Because no power goes above the heating power measured by the plugs, there is no situation that causes overdetection of the heating power. Underdetection due to signal breakage therefore pulls the average value down. While the heating power should generally lie between 1.8 and 2.5 kW (literature 2.3.2), it was observed that maximum power often was mistakenly detected below that. This might give a slight misrepresentation of the average heating power, which might be more something like 2.2kWh instead of 2.0kWh on average. Despite this underdetection, it was shown that multiplication of heating duration with heating power has higher predictive value than heating duration alone, pushing the correlation from 0.90 to 0.92 (even 0.94 in case of averaging measurements per household).

Smart meter (training set)

The validity and generlisability of the meter training set was impacted by multiple factors regarding the sampeling, storing and processing of the data. The main factor affecting the validity of the smart plug data, is the accuracy of the feature extraction method of K. Basu. The generalisability of the data is highly dependant on the availability of smart plug data for those households, the model was developed on, creating similar challenges to be considered. The mentioned points will be discussed in more detail in the following:.

Data gathering
The smart meter training set was sampled for the same 100 households as the plugs by either the Toon smart thermostat via the smart meter's P1 port with 10 second intervals or from analogue meters, where a laser counter got installed when placing the Toon. This poses several potential validity problems related to the sampling and storage of the data. The data might for instance either not been sampled exactly well per 10 seconds interval or there might be signal breakages. Most of these problems are relatively minor and can be overcome with some resampling of the data. However, some households had very anomalous profiles. Analysis of the profiles of households with very bad detection quality showed three main problems: A very active household with many different appliances used at the same time, a very irregular profile due to many signal breakages or a disturbed but regular, low power profile.

The first problem can make it more difficult to detect dishwashers in the signal. However, these are perfectly normal profiles and hence were kept in. Improving on these type of profiles is one factor to improve the accuracy scores of the overall model. An example of such is developments to better identify the water heating phase(s) of a washing machine from those of a dishwasher (see 2.3.3).

The second can be caused by several sorts of faulty equipment, ranging from defects to the meter to internet connection. This problem seemed more common in households with analog meters, but within the scope of this research was not further addressed and households with very faulty signals were taken out of the sample.

Finally, the disturbed, but regular low power profile showed to be an effect of solar panels, installed on some households, providing much or all the power demand during the day. When the solar system is connected with the Toon as well, this data in theory could actually be matched to provide the full profile, however this posed too many technical obstacles to address within the scope of this research. Households with only small amounts of solar energy production might still draw enough load from the grid (in the winter months) to ensure dishwashers can be detected within the profile and would not cause outliers. Households that did show up as outliers due to installed solar panels were filtered out, to reduce noise in the training set. Due to the rapidly growing share of households with solar roofs and other forms of decentralised energy, it will be very important to address this issue by connecting the smart meter data with any data from any other energy sources in the household, such as generation from solar panels, small wind turbines, and charging/discharging of (electric car) batteries in future research.

Arguably the main factor affecting the validity of the smart meter data, is the accuracy of the feature extraction method of K. Basu. Detections of the dishwashers from the smart meter signal were provided by K. Basu based on the methodology as described in literature subsection 2.2.3. This method shows several potential advantages. In comparison to most other research, it had already been applied on ten seconds smart-meter data in a real-life setting. The method showed above 80% detection accuracy would be possible. As the model was developed for only 4 households there would be potential to further improve this with a larger data set.

As described in methodology subsection 3.3.1 the approach produces a set of features related to a dishwasher's power profile, resultant from its operating phases (see 2.3.2). These features were statistically analysed to gain deeper insight in separate components of the power profile. This was done for both the plug and meter data. The plug data was mainly used to gain insight in the actual functioning, which was then used by K. Basu to make improvements to the detection algorithm compared to the one published. The produced smart meter detection features were compared to see how well each feature is detected and to what extend this could be utilised to estimate the energy usage per wash.

The accuracy of detection of features affects the accuracy (RMSE) of the energy estimation (see section 5.2.3). The regression method adjusts to bias in the measurement by how the parameters are calibrated, but cannot reduce the noise. To reduce the noise, a

GMM was used (see appendix D.0.1). However, by improving the detection algorithm and hence detection quality, over time the added value of GMM reduced and simply averaging started to yield better results. The precision and recall of detection both affect the accuracy (RMSE) of usage frequency estimation. (see more in subsection 5.2.3).

In comparison to other NILM approaches, models developed on smart meter data have high applicability, because of the ubiquitous roll-out of smart meters. Systems developed on higher frequency data may be more precise but provide less opportunity to scale non intrusively as additional measurement devices will be required.

5.2.2 Energy per wash estimation

In order to estimate the energy consumption of dishwashers in households an energy estimation model was developed and performance indicators were analysed. An analysis of features was carried out comparing the smart plug detections with smart meter detection features as provided by K. Basu (see 2.2.3). Multiple regression approaches were used and compared (see results 4.2.1 and for more detail appendix D.0.1) to find the model with the lowest estimation error.

The three main findings to be discussed for the energy per wash estimation are the initial findings for the energy per wash from the plug analysis, the features affecting the energy usage per wash and the performance of the chosen estimation approach.

Energy per wash

From the smart plug analysis it was established that washes of the 100 plug users average at 1.22kWh with an error of the mean of 0.02kWh. Stamminger (2008) noted a typical profile with a consumption of 1.19 kWh over all European countries, based on the EuP survey results. However, instead of measurement of real-life usage this is an estimate based on lab consumption and was done back in the year 2008 [128]. Dupret and Zimmerman (2017) found 0.90kWh for the 60 households with dishwashers of the 107 households they installed plugs for their research in 2015. This actually seems very low. While care was taken to ensure the households give a good representation of French households, no additional information is given about the dishwasher specification in the sample and hence no 1-on-1 comparison can be made. While the considerably lower energy consumption might be caused by French households potentially using smaller dishwashers and/or more efficient machines and/or washing at lower temperatures, 0.90kWh seems surprisingly low (comparable to 9 couverts, A labelled or 11 couverts A++ labeled dishwashers using the eco program). Another potential reason explaining the very low value, might be that the energy consumption was multiplied by the share of households owning a dishwasher. However as the 59% ownership rate would result in 1.53kWh per wash, this might seem a bit high actually. While the penetration rate for all included appliances is given, no such multiplication is mentioned. A potential factor at play might be that the French research only used a sampling rate of 10 minutes. As a result the French research might actually have averaged out large energy peaks related to the water heating phases, potentially resulting in considerable lower measurements. As during this research a sampling interval of only 10 seconds was used, energy could be measured with considerable higher precision.

Van Holsteijn and Kemna (2008) [23] in their analysis assume a similar penetration rate of 60%, but an energy usage of 1.09kWh per wash for the period between 2015-2020 in their business as usual case and even slightly lower in their more eco conscious scenarios. This is based on the assumption that all dishwashers would be energy label A and 80% would be 12 couverts (1.06kWh) and 20% would be 9 couverts (0.75kWh), when washing with the standard eco program. They assumed a behavioural penalty of 9% in addition. When comparing this to the label calculations (see 2.4.2) the same 80/20 split of A labelled dishwashers would result in 1.12kWh. When including the 9% behavioural penalty this would actually result in just under 1.22kWh. Apparently Van Holsteijn and Kemna in 2008 were too optimistic (∼12%) on how much energy usage per standard wash would come down, but adjusting for that, the behavioural penalty (∼10%) might provide a very good adjustment compared to the standard energy per wash.

However, two additional factors have to be noted. The first regarding the actual machines part of the sample group, the second regarding the measurement period. While not completely covering all households, the sample of (41) participants who fully completed the survey, showed that at least 67% had an A+ dishwasher or higher (and 17% unknown). No households reported a dishwasher with less then 12 couverts, although 22% reported unknown. Assuming an A+, 13 couverts dishwasher, which seems to better represent the survey sample, would result in a standard wash of 1.06kWh. Even when the group of households that did not (fully) fill out the survey would have slightly less efficient machines, still some 10-15% difference between 1.22kWh measured and the standard wash would have to be explained. Given the young age of most dishwashers in the sample (64% below 5 years), reduced efficiency as a result of encrustation might have some effect for some of the older machines, but would play less of a role overall. While behaviour and chosen washing temperature in particular might explain much of the difference, ambient conditions related to the time of year (Winter months) might be another factor explaining some percentage points As results of large scale plug analysis showed, the average energy per wash fluctuates throughout the year in relation to the outside temperature.

Features of the wash

As shown in table 4.1 the energy intensity of about 80% can be attributed to the heating phases of a washing cycle. This is in line with the findings from [130] stating that energy is mainly used to heat the water for the rinse cycle and the drying phase (see section 2.3.3.

Water heating being the most relevant indicator to estimate the energy per wash is further established by a high correlation factor of 0.94 for plug detections and 0.82 for meter detections (see table 4.2.

For the meter data processing the methodology by Basu et al. was used. The difference between the smart plug and smart meter detection indicates that some estimation quality is lost due to the detection quality of the NILM detection algorithm. Two impacting factors for this are potential misdectections of other appliances and the precision of the detections. The latter can either be caused by heating moments remaining undetected or overdetecing by factoring in other appliances. The noise occurance in the Basu algorithm has been investigated with the present data set 4.1. From the publication of Basu it stems that the detection accuracy is over 80% [43].

From visual inspection it becomes clear that a bias towards underdetecting features is present. This relatively clear bias in one direction can be adjusted for by calibrating the regression parameters. However, when an error goes in both directions, i.e. detection system is over- and underdetecting, this cannot be adjusted for and results in estimation error.

Estimation approach

From the comparison of the estimation methods as described in section 4.2.1 it can be concluded that the best functioning estimation approach is Predict 3; the regression of the heating energy (created by multiplying heating duration and power, see section 4.2.1), combined with number of heating moments, and aggregated to the average.

For the interpretation of why this estimation approach has performed best, the following categories need to be considered: feature choice, regression approach and aggregation type.

Feature choice:
Based on the feature analysis the heating energy clearly stood out as the most important factor. The lowest estimation error was shown for the combination of heating energy and number of heating moments. Some features were already earlier discarded, such as average power for the washer as this was too low to be able to be detected

in a meaningful way (too high error compared to value and only very small additional influence). This in comparison to for instance a washing machine, where the drum spin does considerably show up more in the energy profile (see subsection 2.3.3).

Regression approach:
The chosen regression approaches are more precise than the heuristic approach because the parameters are calibrated to reduce the error. This does mean that a regression approach might risk overfitting. The data has been checked by cross validation (1000 runs, 50/50 split) to ensure a robust way of error analysis, which did show that the addition of some variables such as adding duration of the wash did not improve estimation performance. Resulting in Predict 3 as the simplest, best performing estimation method.

Aggregation type:
For Predict 1,2,3 the average did considerably better than the mode and the reason for that is probably that the average uses all data points while the mode only uses the most common value, so much less data points are available. A more sophisticated way to estimate the energy consumption of the most likely washing program is to use a clustering method such as K-Means or a Gaussian Mixture Model (GMM). As the GMM can split the data into clusters of similar washes, it provides the opportunity to look at the different used washing programs, including how often they are used, and calculate a weighted average. GMM clustering was used to aggregate the washes for each individual household to estimate both the weighted average and most common wash. This initially provided more precise results than the common average and mode, because it also filters out some miss-detections. However, after further improvement of the detection algorithm by K. Basu, simply averaging started to perform better than GMM clustering. The reason behind this can only be guessed. Probably the washes that are thrown out as misdetections might be washing programs that are not used that often and therefore not considered in the GMM clustering approach. With higher noise the GMM helps to filter this out, but with less noise risks to throw away useful information. More effort could have been made to investigate and further calibrate the GMM. However, as the average works as well or better, this is a simpler solution, which is easier to understand, less susceptible to overfitting and less computationally expensive. For further improvements, increasing the precision of the feature detection therefore seems a more fruitful route.

The only way to really find how well the calibrated model is able to estimate when actually deployed, is to test it with an unseen separate annotated ("learning") data set. There are some academic benchmark data sets that are often used to develop new algorithms with the ability to compare them to each other in terms of performance (see table 2.1). However as this research focuses not on the performance of a state of the art NILM detection algorithm, but real-world application, a data set that better reflects the deployment group was more fitting. Also the used data set in this research (100 households) is larger than any of these benchmark data sets, which often only contain several houses and use different measurement intervals. Some often cited data sets such as Loughborough University [52] for instance might have too low measurement rate for the application of this research, while another often cited research, the REDD database [72] has very high frequency, but contains only 19 days for 5 households.

Ideally a reflective sample of households would have been selected from the large scale Toon deployment data set and selected households would be fitted with smart plugs. Indeed this was considered as a next step, but was outside of the scope of this research to still consider.

Potentially the most fitting alternative with public data would have been the REFIT data set, an 8 seconds measurements data set for 20 households in the UK [69] or the more recent IDEAL data set with 39 plugged households at 1 seconds measurements in the UK as well [68]. As these data sets are not completely comparable to the case for the Netherlands, some adaptions would still have to be made. However, making such comparison and adaption for different countries could be a step for further research.

When pair-matching the smart-plug and smart-meter detections misdectections and missed detections by the meter (false positives and false negatives) would fall out. Hence, washes cannot be connected one-on-one and an aggregation method should be used to connect a group of washes. The data can be aggregated either per household over the full measurement period or per week or month for all households. Since the focus of this research is the estimation of energy consumption of individual households, the data was aggregated per household for the full measurement period.

The three months period was chosen based on a rule of thumb to have at least 30 observations to be able to do some statistic analysis. Even a household that only washes once every three days, hence 2-3 times a week, would wash about 30 times in a 3 months period. Potentially a next step to further improve the accuracy could be to test what a minimal measurement period would be and if it would be possible to either improve accuracy with longer measurement period or aggregate data in multiple shorter aggregation periods, such as a monthly average, to increase the number of data points per household.

The meter detection algorithm could potentially be improved. As detecting appliances is now no longer the end goal, but estimating their energy as accurate as possible as well, the accuracy with which subfeatures are detected becomes more important. Certain subfeatures can potentially have such large random detection errors that including them actually decreases the estimation accuracy and they better can be accounted for in the β_0 parameter. As certain subfeatures will correspond to larger areas under the power profile than others, some subfeatures will be more important drivers for accurate energy estimation than others. Therefore focusing on detection these features in particular with better precision could improve the accuracy of the energy estimation.

5.2.3 Usage frequency estimation

In order to estimate the frequency of usage of dishwashers in households a frequency estimation model was developed and performance indicators were analysed. The number of weekly washes according to the smart plugs was compared to the number of weekly smart meter detections as provided by K. Basu (see 2.2.3). Multiple regression approaches were used and compared (see results 4.2.2 and for more detail appendix D.0.2) to find the model with the lowest estimation error (NRSME). Key findings on the number of washes per week and the chosen regression approach are discussed below.

Weekly washes

With regards to the frequency of dishwasher washing cycles per week it was found that on average 5.2 washes per week occurred for the plug detection. The EU regulation uses 280 washes.

According to Van Holsteijn en Kemna this could potentially be an overestimation of about 23%, citing 4.1 washes per week or 214 washes per year. The EuP study in 10 different countries in Europe showed that an average of 4,06 washing cycles/week was declared, adding up to 203 washing cycles per year (50 weeks) . The Ecodesign directive assumed a usage of 220 times annually or 4.5 times per week taking 3 holiday weeks into account or 4.2 times per week not taking holiday weeks into account.

Dupret and Zimmerman 3.63 seems low. However, a comparison is made to Remodece+ study from 2008, where an average of 4.1 was found. This either indicated a reduction in weekly washes as Dupret and Zimmerman argue or simply a difference between sample groups. Also research in France.

In that respect the relative high 5.2 washes per week found in this analysis could be at least partially explained by family size in the sample. The average family size of households who filled out the survey was 3.3. In the VEWIN study used by Van Holsteijn en Kemna households of 3 people used the dishwasher on average 5.1 times per week and households with 4 even 6.4 times per week. While the number of weekly washes per person reduces with increasing household size, we can roughly estimate that 3.3 would average about 5.5 washes per week, so actually slightly higher than the number of washes found. However these results stem from survey research and hence are stated number of washes. In this research it was found households filling out the survey slightly overestimated their number of washes. Resulting in an overestimation of about half a wash per week. When assuming this same effect applicable to the VEWIN survey the results come even closer. Finally it has to be considered the measurments took place in the winter months. As results of large scale plug analysis showed, not only the energy per wash, but also the average number of weekly washes fluctuates throughout the year, being slightly higher in Winter.

Estimation approach

The comparison of the estimation methods for weekly washes (section 4.2.2) indicates, that the best functioning estimation approach is a hybrid using $\beta = 0$ with the average of general linear regression (Predict 3), showing an NRMSE of 27.2% with a 95% interval of ± 2.8 washes.

For the interpretation of why this estimation approach has performed best, the following categories need to be considered: regression approach, aggregation type and aggregation order.

Regression approach:
While the energy per wash estimation mainly focused on the usage of different features, for usage frequency only the weekly washes were available as variable. While other meta-features could be used such as family size, the aim was to be able to estimate this based on the smart meter detections to be able to compare this to factors such as family size afterwards. In order to do this different regression approaches were compared. As general linear regression (Predict 1) intersects with the vertical axis for $\beta_0 \neq 0$ this takes away the potential to predict 0 washes in a week, when nothing gets detected and instead would still result in some value. Hence, by doing a force $\beta_0 = 0$ regression (Predict 2) a week with zero detections on the meter indeed will also result in 0 washes in that week estimated. To adjust for the reduced accuracy of a force zero regression compared to the general linear regression, the force zero was adjusted for the distance to the mean according to general linear regression (Predict 3).

Aggregation type:
For the weekly washes average also did considerably better than mode. Similar as for the energy per wash this is probably as there are less data points used to calculating the mode than to calculate the average. In contrast to the energy per wash the GMM was not used here. The GMM could be used to compare the relative share of clusters of detections, hence usage of different washing programs. This function was no longer pursued once the energy per wash could be detected with higher accuracy using the average than the GMM's weighted average and the mode could be calculated with higher

accuracy than the energy usage of the GMM's most common used wash.

Aggregation order:
First regressing on all weekly number of washes and then aggregating that into one average weekly number of washes appears to create the best result. This can be the case because first regressing and then averaging makes use of the higher volume of data of all detection weeks (12 weeks for each household) instead of first averaging and then regressing, which results in regressing on only one data point per household.
Data in this case is number of washes in each week. For the meter this depends on accuracy of the detection algorithm. Similar factors apply as discussed under energy per wash. As the aim of the detection algorithm is to make as many correct estimations as possible, both over and under detections are made, with an average accuracy around 0.80.

The best performing regression approach (Predict 3) combines the force zero regression, but then adjusts for the distance from the mean. By doing this zero washes would still result in actually more than zero, namely the difference that the force zero regression has towards the mean. On the total average number of washes this results in a better overall average. However, in case an actual updated average would be given to users for every week, this would have to be adjusted to actually force zero detections really to be zero.

Optimising the detection algorithm to improve precision at the cost of recall, could be beneficial to ensure better energy per wash estimation. This could potentially also benefit improving reducing the error of weekly wash detection. However, as this would cause a bias towards underdetecting the regression would increase the estimated number of washes even more than currently already is done. Therefore, even in case of zero detected washes, multiple washes would get estimated. Instead therefore more emphasis could be placed on the recall, to ensure if there is a wash at least it is detected. This might cause some misdetections. When the regression algorithm adjusts for overdetection, potentiall also negative amount of washes could get estimated ($\beta < 0$). Simply setting that no negative numbers can be estimated would resolve this. Searching for an optimal balance of over or underdetection, might reduce error for the weekly wash estimation. As it might be more important to tell consumers over a period of time how often they washed, rather then to try to detect washes as precise as possible, the detection algorithm could therefore be optimized to reduce the error in weekly wash estimation. However, testing this was not in scope of this research.

5.2.4 Efficiency classification

In order to estimate the efficiency of dishwasher usage in households an efficiency classification method was developed and performance indicators were analysed. Both for energy per wash and for usage frequency a model was developed to establish a binary thresholds as a proxy to classify the estimation results as efficient or inefficient (see section 4.2.3). The efficiency classification of the estimated energy and weekly washes were compared to the efficiency classification of the plug measurements. Varying thresholds were analysed to find the maximum classification accuracy (precision, recall and F1 score).

The interpretation of the outcomes of the efficiency classification method can be based on the thresholds for energy per wash and usage frequency with the maximum accuracy. For the energy per wash a threshold of 1.05kWh was established, leading to 78% of plug users surpassing this threshold and thereby classifying the energy usage of their dishwasher as inefficient. For this threshold a classification accuracy of 0.89 (F1 score) was found for the meter estimation.

The usage frequency threshold was set to be 4 washes per week, whereby the frequency of dishwasher usage was classified as inefficient for 69% of plug users. For the meter estimation a classification accuracy of 0.89 was found, similar to the energy per wash estimation.

The validity for the efficiency classification depends on the estimation part and hence on the input data and models used, as already discussed in the previous two subsections. For the efficiency classification generalised set threshold values were used (for assumptions see 3.3.3). Choosing a general efficiency threshold does not take variation into account within the households, such as family size and dishwasher size for weekly usage and machine size, efficiency label, washing temperature but relies on the assumptions stemming from majorities and averages. When these features would be known, a household particular threshold can be chosen, which would allow for a more accurate assessment. A systemic and efficient connection could only be made, if all consumers would register their appliances in data bases, which could be implemented in app systems for efficiency assessments of households.

5.3 Smart Meter Model Deployment

In this section the deployment of the smart meter model is discussed. To be able to investigate dishwasher usage patterns in households non-intrusively and scalable, the development of the model was used to analyse a large set of smart meter profiles. The developed model was discussed separately in the previous section. In this section, the analysis of the smart meter profiles will be discussed. Starting with preparing the large-scale data set. The rest of this section is structured according to the same four subcomponents as the results section: consumption overview, category dependency, time dependency and efficiency analysis. For each of these subsections an interpretation of the main findings is given. Validity of the methodology is discussed as well as generalisability. Because this section shows the deployment of the model, emphasis is placed on the interpretation of these results and what we can generalise from this about dishwasher usage in real life.

5.3.1 Data

The developed algorithms were applied on smart meter data gathered for a full year for nearly 130.000 households in the Netherlands to investigate energy consumption, usage pattern and efficiency characteristics non-intrusively at large scale. A survey was conducted under a sub-sample of nearly 11.000 households to segment the results for household and machine characteristics. The assessment was accompanied by yearly temperature data and vacation days from 2018. The following paragraphs set a focus on the data gathering and processing of the large scale meter data set as well as the comparison to the smart plug data. The survey and other data will be addressed in their respective subsections.

The detection system of K. Basu was deployed on the full one-year data set to retrieve the dishwasher detections, which were then used to estimate energy per wash and weekly usage frequency for each household using the trained estimation models. One major factor to be taken into account for the data processing is that the training of the meter detections is based on the plug data. Thus, the plug data was set as the 'truth', which the meter detection algorithm is supposed to approach as accurately as possible. The previous section shows there are many reasons why the smart plug data is not "truth" either and bias might have been caused within the learning.

One of these aspects also shows when considering the seasonal variability. As the research indicates there is seasonal variability throughout the year for both energy consumption and usage frequency (see section 4.12 and 4.13). The training set was only gathered over a 3-month period in winter. The impact of seasonal variation on the data could therefore not be taken into account for the training of the algorithm, which clearly affects these averages and might also as a results have impacted the model (e.g. β_0 component).

The deployment of the detection algorithm over such a large database of smart meter profile showed to be computationally expensive. This strongly impacts the adaptability of the detection algorithm when deployed at scale. The algorithm was adapted over the time of the research but due to its computational expense was not re-run over the full year. As a result the data had to be adjusted for this adaptation, introducing a processing uncertainty into the outcomes. Ideally, the adapted algorithm should have been re-run over the full data set. However, with the adjustments made, the impact on the results were minimalised as much as possible.

When comparing smart plug and meter data it can be stated that smart plugs potentially give more precise results but are not as scalable as smart meter measurements due to their intrusive nature. This trade-off can be seen as a centre piece for the decision regarding the set-up of research in this field. While Toon users may not be a perfect representation of the complete Dutch population, compared to the plug data, the large-scale user group draws from a much wider household base, making results potentially much more generalisable. However, potential bias from the plug validation data could have been carried over into the machine learning meter model deployed at large scale.

To assess model accuracy, different models were compared and cross-validated, however the impact of a potential bias for the validation sample can only really be assessed by introducing a new set of randomly selected households with previously unseen matching plug and meter data, ideally for the period of at least one full year. This data is not easily available. Smart meter data retrieved via other sources than Toon and comparison with multiple plugged sample groups was out of scope of this research, but could be an avenue for further development of the model.

The data set is only focused on the Netherlands and within that only represents a subgroup of people who own a Toon device. This makes some of the results less generalisable to the Netherlands as a whole or even beyond. In particular for generating averages, it needs to be carefully assessed, whether such big unspecific data set should be preferred over a smaller but controlled set of households, representing an average. However when considering usage patterns such as temperature dependency, a demographic representation may be less important, but statistical soundness of the results is. Due to the size, the data set provides the opportunity to cut different subgroups and still provide large enough groups to derive such patterns.

5.3.2 Consumption overview

The developed models were used to detect dishwasher activity and estimate energy usage and usage frequency of dishwashers for each household. Summary statistics on average energy usage per wash, usage frequency per week, per year and the resulting total energy consumption were retrieved. The distribution for energy per wash and weekly usage was compared for all households, plugged households and households that filled out the smart meter survey.

The NILM system revealed an average energy consumption of 1.18 kWh/per wash, 4.6 washes per week, resulting in 240 washes and 285kWh per year.

Comparison of the energy per wash to the literature can be found in subsection 5.2.2, where the plug energy consumption was discussed. Compared to the plug results of 1.22 kWh per wash this value is slightly lower, but the difference is negligible. For the surveyed group of households a similar average of 1.22kWh was found. The reason for this may be that the plug measurements were taken in winter months but the meter measurements range over a full year (also see section 5.3.4).

The frequency of washes of the meter measurements/surveyed meter households lies 0.6/0.3 below the frequency of washes detected by the plugs. This might again be an effect of the measurement period, with more usage of the dishwasher in the winter months (also see section 5.3.4)). Furthermore, the potential difference could be caused by slight differences in average family size (3.3 for plug vs. 3.0 for surveyed meter households). Since there is no data available for all the meter households it remains unclear, what the average household size was for all households. It can be assumed, that the household size is somewhere in the range between 3.0 and 2.2 (Dutch national average).

When looking at the distribution of the results, the large scale meter detections showed to have less variance than the small-scale plug results, while more variability in the much larger group would be expected. It was shown the model generalizes towards an average value. This would result in good estimation of the population average for an unbiased estimator and for better estimation for values close to the average but results in less accuracy towards the outer regions of the distribution. This is particularly useful since the threshold for efficiency is also close to these average values where accuracy is most needed, and these average values are more relevant as overarching findings. However, this does indicate that exact energy consumption and usage frequency further away from the average becomes less certain for individual households. This could be potentially improved by specifically selecting households that are more on the outer edge of the spectrum for future data gathering and analysis. However, the regression method will remain to have a tendency towards averaging results. Additional data such as from surveying the user could help to provide more information to narrow down possibilities.

5.3.3 Consumption category dependency

For the assessment of the consumption category dependency of dishwasher use, the meter estimation summary statistics were segmented by categories retrieved from the survey. For the following categories main findings will be discussed: machine size, household size, efficiency label and temperature. The machine size affects both energy and frequency therefore estimated energy per wash and weekly washes were segmented by machine type. The household size impacts the amount of soiled dishes produced in a household and was therefore compared against estimated frequency. Additionally, to compare the estimations with survey results, stated frequency was compared as well. The efficiency label and washing temperature determine washing energy. These were used to categorise the energy consumption per wash.

Machine size

As shown in figure 3.7 the machine size affects both the energy per wash and the usage frequency. The majority of households in the survey (88%) stated to have a "Regular" machine. For these households an average energy consumption of 1.22 kWh/per wash and 5.0 washes per week was revealed. For households which indicated to have a "Table" machine or did not specify the machine type similar values were found. For the (3%) households indicating to have a "Compact" dishwasher an average consumption of 1.13kWh and 4.6 washes was detected. The survey results consolidate the approach chosen for the efficiency threshold model to focus on regular machine sizes. While "Table" dishwasher types are the smallest machine size (around 6 place settings), both the energy consumption and weekly usage appeared to be similar to the standard machine size. A potential explanation could be, that the survey was not clearly enough defining this type of dishwasher. This presumption is stemming from questions returned from individual households on this topic, therefore no differentiation can be made between regular, compact and non specified accounting together for a total of 97%.

Household size

Referring to figure 3.7 the household size affects the usage frequency. The assessment of the meter data set showed that a differentiation between family sizes and frequency can be made.

One person households were found to wash an average of 2.8 times per week, two people households wash 4.4 times on average. For every additional family member number of the weekly washes increases roughly 0.6 on average. This shows that for every additional family member only a marginal amount of extra dishes is produced, which could be explained by a basic set of pieces needed to prepare a meal, but every additional person at the table might only require some additional plate, cutlery and a glass. The stated washes reveal a similar pattern, although would indicate a slightly higher estimation of about 0.5 washes per week more than what is estimated. Similar results were found for the plug measurements as shown in **??**. While showing a larger interval, the review by van Holsteijn en Kemna shows a similar linear marginal increase. The weekly usage is dependent on household size and can be seen to be 2.2, 4.2, 5.1, 6.4 and 6+ for households ranging from 1 to 5+ people respectively [149]. As it was shown that the model generalises to an average value, is could be assumed that the model overestimated household sizes below the average and underestimated household sizes above the average. The average of 2.8 might therefore be an overestimation of the 2.2. On the basis of having this insight a post adjustment of the model could be done.

It also seems that the participating households did overestimate their frequency of usage in the meter survey (4.9 calculated average by households versus 5.4 average over data stated in the survey). This finding was manifested by similar findings of overestimation in the plug survey compared to the plug measurements (see **??**). This finding might have broader implications for appliance usage survey in general. The following factors could result in potential discrepancies:

- Detection accuracy: Potentially the higher detection error for weekly washes (27.2%) could play more a role compared to energy than what was established. Indeed due to the relative low number of households only a few households on either side of the spectrum can already change the accuracy number, as discussed in discussion subsection Y. As for energy the threshold could be adjusted to include the error margin, which has not been done, since the recall would be reduced accordingly.

- Survey range: In the survey people were asked to indicate their weekly usage as a range instead of a single number. The threshold of 4 washes would either have to align with the survey respondents who answered to use their dishwasher 1-3 times or those who answered to use their dishwasher 4-6 times a week. To provide a more conservative estimate the 4-6 times a week was chosen. However, in terms of classification accuracy a meter threshold of 5 washes actually would better match this range than the set threshold of 4 washes, resulting in a precision of 0.73, recall of 0.75 and F1 of 0.74. Single numbers or overlapping ranges could therefore have provided more precise survey results.

- Survey error: Similar errors to the survey apply as well. Additionally, it seems people overestimate the number of weekly washes. Potentially more considering what they perceive as a typical week, rather than estimating an average (see discussion subsection 5.3.3)

Efficiency label and washing temperature

Figure 3.7 shows that efficiency label and washing temperature affects the energy per wash. With 33% the A++ was the most common efficiency label, most (55%) households stated to use the temperature range between 50-75 degrees Celsius. This result is in line with findings from literature stating that the main washing temperature lies at an average of $59,3$°C [24]. This indicates, that a large share of the households might wash at higher temperatures than what would be seen as efficient. Another consumer survey on 200 households revealed that 52% of the interviewees with a dishwasher favoured washing temperatures of 65°C resulting in a higher energy consumption per washing cycle than stated by the respective energy labels [150]. Van Holsteijn en Kemna state there is evidence that the eco program is not the most used. [146]

The results show clearly that there is a dependency between average energy usage as detected by the algorithm in relation to the label and washing temperature. Estimated energy increases with decreasing label and increasing temperature. Furthermore, the temperature of the chosen washing program shows to impact the overall energy usage more than the label. This could be explained by the labels being developed for the eco program, while many households actually often use other (hotter, more energy intense) washing programs, as suggested by Van Holsteijn en Kemna.

However, the differentiation is not as pronounced as would be expected. Only a mean of 1.10 compared to 1.35 are registered between the most and least efficient. This could be explained by two factors:

- Detection accuracy: One factor could be that the energy consumption was not detected precise enough. From the energy estimation it is known that there is an energy estimation error of 8.8% (NRSME) or about 0.1kWh. This effect becomes more pronounced further away from the mean. Hence, especially the lower and higher values on average get estimated closer to the mean than they actually would be when measured with a smart plug. Potential further research could focus on different estimation model or after-processing to adjust for this.

- Survey error: Another impacting factor could be that the surveys were not filled out correctly. Potentially the survey may not have been filled out by the household member being most aware of the machine usage patterns (e.g. program choice). Furthermore, knowledge about the machine type and characteristics (e.g. label) might have not always been known correctly by the person filling out the survey. Additionally survey questions could have been misunderstood, or answers were not recognised or unclear.

5.3.4 Consumption time dependency

To analyse the effect of time as a variable for dishwasher usage patterns, dependencies on time of the day, week and year were investigated. Regarding the seasonality effect on the energy consumption, findings were related between the average outside temperature and the average energy per wash aggregated for each week of the year. To gain a better understanding for the weekhourly usage pattern, the chance for starting a wash throughout a day was assessed. In addition public holidays were matched with the average weekly usage frequency to determine the effects of events on the weekly washes.

Seasonality effect on energy

It was shown that the average energy usage varies throughout the year. The energy pattern shows a strong negative correlation (-0.94) with outside temperature. The average energy consumption per wash was highest at 1.28kWh in a freezing week in Winter and dipped during a heatwave in the Summer at 1.05kWh.

For refrigerators a similar effect has been found for ambient temperature[152]. While dishwashers might seem to operate in relative constant room temperature within the house, the most important aspect affecting their energy usage is the heating of water. The effect of outside temperature on the energy per wash could therefore be explained by a change in inlet water temperature. This might be colder in winter, hence more heating energy is needed to reach the desired water temperature for the chosen dishwasher program.

It has to be noted that no analysis has been done on the actual temperature of the inlet water to match this with the dishwasher's energy consumption. This is merely a correlation between the average outside temperature (from the Dutch meteorological institute, KNMI [161]) throughout the year and change in energy consumption by the dishwasher. While no further analysis was done on factors such as geographical spread and type of dwelling, it could be assumed that variation in inlet water temperature will differ by factors related to depth of pipes in the ground, types of pipes and insulation. Making some houses more sensitive to changes throughout the year than others.

Effect of events on weekly washes

With 5.3 washes, the number of washes peaked in the first week of January and reached a minimum in the middle of the summer holiday at 3.3. Both of these extremes also coincided with holiday events, which is exemplary for the effects events have been shown to have on the usage pattern throughout the year.

The seasonal effects could be related to people potentially making less use of the dishwasher in the summer since due to the warmer weather and increased daylight they are more outside and possibly the diet adapts to lighter meals. Events throughout the year can work in either direction, depending on the behaviour of large groups of people during these holidays and festivities. For instance Christmas caused a large spike in dishwasher usage, while the many people going abroad results in decreased dishwasher usage during the summer holiday. Furthermore, as the average was only calculated for household using their dishwasher that week (conditional vs including zero) it was shown that in summer not only less people are at home, but those at home also make less use of their dishwasher.

One important remark regarding the validity of this pattern, however, is that the NILM detection algorithm was adjusted throughout the year. While multiple small adjustments have been made throughout the year, in week 15 a major change was made to the sensitivity of the detection algorithm. This was done to increase the precision against the cost of recall. Hence less washes are misdetected, but more get missed. This resulted in an average reduction of 0.74 weekly washes per week per household. As it was too computationally expensive to rerun the full year of data with a single version of the NILM detection algorithm, the time line has been adjusted to match the time period before and after this change in week 15. For more on this change, see appendix E.1. While averaging over the surrounding weeks resolved to match both timelines, it cannot be guaranteed that some seasonality effect that might seem to be the effect of behaviour patterns, actually is the effect of a change in the detection algorithm. While this does not affect the relation between events from week to week, to ensure the complete annual pattern reflects the annual usage pattern, either multiple years of data or at least a rerun of the data with one and the same detection algorithm would be needed. However as both of these would be extremely computationally expensive and no additional year of data has been made available yet, this has not been considered within the scope of this research.

Generally, the power demand of a household fluctuates depending on the type and amount of appliances used in a specific time. Intraday fluctuation shows higher energy usage during daytime than at night. Typically an evening peak occurs due to intensified lighting and cooking activities. In addition, changes between different weekdays can be observed particularly between workdays (Monday through Friday) and weekend days (Saturday and Sunday). Over the year seasonal differences cause changes in the energy consumption due to differences in outside temperature, hours of solar light and behaviour between seasons. Total energy use in winter is higher than in the summer months. Autumn and Spring show comparable patterns. [51]. This research gives more detailed insight on how this actually is affected for a single appliance, which could be extended for other appliances as well, to better understand the household energy consumption profile bottom-up and how this potentially responds on events.

Weekhourly usage pattern

While events affect the washing pattern throughout the year, aspects of the daily and weekly rhythm can also be recognised throughout a week. Dishwasher activation peaks twice considerably in the evening, at 6/7 pm around dinner time and at 11pm before bed time. After a period of low activity at night, in the morning there is another but smaller peak and a slight peak after lunchtime can be seen, where people might be eating lunch at home.

A downward trend in activity was identified from Monday onward, with Saturday evening depicting the lowest dishwasher usage. This can be connected to that it might be more common to go out for dinners towards the end of the week rather than in the beginning. Saturday and Sunday also a more even pattern throughout the day can be noticed than during the workweek, as more people are at home. The morning peak also starts a bit later as people might be sleeping in.

An analysis on the demand shifting potential of appliances, using the EuP consumer survey transferred the outcomes into an estimated average curve for start time of dishwasher in Europe over the course of a day and the related power demand needed for operating a dishwasher per day and household (2.3.3). The results as shown in figure 2.6 show that the peak of washing activity lies between 7-9pm, with a smaller peak between 6-8 am. This timing seems to be slightly different for the analysed households. A possible explanation could be the average dinner time in Europe being later than in the Netherlands. Furthermore, the higher level of granularity of the meter detections compared to the EuP survey results, provides additional insight on smaller spikes and dips throughout the day establishing a clear probabilistic relationship between people's daily rhythm and usage of the dishwasher, which could be used in future energy demand estimation modelling.

5.3.5 Efficiency analysis

To find out how efficiently dishwashers are used in households, the estimated energy per wash and weekly usage for the nearly 11.000 households were compared to the established efficiency thresholds. Estimated usage was classified as efficient and inefficient and compared to the survey answers in order to be able to understand how well the developed NILM classification is able to detect inefficient usage. Finally the potential for improving energy efficiency in households was estimated.

Energy efficiency

Efficiency classification for energy per wash The analysis shows, that 16% of the households are classified as efficient and 84% are classified as inefficient with a precision of 0.73, recall of 0.89 and F1 of 0.80. This indicates, that the majority of the households are using more energy than the set threshold. The validity of the threshold has been discussed in discussion subsection 5.2.4. Most households using more energy per wash than this threshold could have multiple reasons.

- Washing program: The efficiency threshold was chosen for the energy consumption of the standard eco-program. However, the survey results showed (table

4.10), that 61% of the households stated that they use washing programs with a range of 50-75 °C), as discussed see 5.3.3. Van Holsteijn en Kemna state there is evidence that the eco-program is not the most used, which these findings confirm.

- Machine characteristics: The efficiency threshold was set for an A+ machine with 12 place settings. The efficiency label, see table A.1 indicated a 10% difference in standard energy consumption for each step in label. The survey showed, that 24% of the households have a machine with a label of A or lower (A-) (table 4.10). According to the label directive these machines henceforth use more energy per wash than what has been stated as efficient in the efficiency threshold model. The size of the machine and therefore the place settings may also have an impact on the energy per wash. However 88% of the households surveyed showed to have a regular machine and only 3% stated to use a compact machine (4.8), which makes this factor negligible.

As can be seen in figure 4.16 the average energy consumption per temperature range and label is plotted. This shows an even steeper increase within each label as a result of temperature than the difference in energy consumption between the labels. In order of magnitude it can therefore be concluded that the washing program is the biggest impact factor on the efficiency followed by the machine characteristics and the ambient temperature. This would also be in line with information gained from literature where the combination of the most used washing program and its nominal temperature are said to be two key parameters [147].

When considering the potential optimisation measures for the energy efficiency, one could conclude that an important factor for efficiency improvement is the use of lower washing temperature, which is a direct impact that the consumer can have on its own energy balance. The most difficult factor to change appears to be the water inlet temperature.

With regards to the accuracy of the efficiency classification several factor could play a role 5.3.3. In addition to the aforementioned factors, discrepancies between meter findings and survey results can also be caused by other factor:

- Ambient conditions: The average annual water inlet temperature could be lower than the temperature used for testing under the EU efficiency regulation (15°C). In the Netherlands for the year 2018 an annual average outside temperature of 11.4 °C was calculated. It can therefore be suspected, that the energy usage of dishwashers in the Netherlands might be slightly higher, having to overcome a potential lower inlet water temperature.

- Machine malfunction: Despite an efficient washing label or washing program malfunctions of machines e.g. due to high usage frequency and age could be reasons increasing the inefficiency of a dishwasher over time by encrustation. These factors are hard to analyse and have not been included in the research, but could potentially be analysed when detailing the load profile assessment per machine type and assessing for abnormalities e.g. in the heating cycles.

Frequency efficiency

The results of the efficiency classification for weekly number of washes resulted in 38% of the households to be classified as efficient and 62% as inefficient with a precision of 0.61, recall of 0.87 and F1 of 0.72.

This result indicates that only one third of the households is using their dishwasher an efficient number of times a week, according to the established threshold. The validity of the threshold has been discussed in discussion subsection 5.2.4. Some two third of households using their dishwasher more often may have varying reasons.

- Amount of dishes: Can be impacted by the number of family members and amount of dishes used per family member. While the number of family members is a given and efficient usage will differ by household size, the threshold is set for a household size of 3 family members. Indeed the average number of family members in the data set is just below 3. However, there are another 37% of households with more than 3 family members. Nevertheless, as shown in table 4.8 even households of 2 family members on average use their dishwasher 4.4 times a week. Much in line with other literature findings as discussed in the frequency of usage subsection.

- Capacity of the dishwasher: Too small dishwashers for larger families would result in the need to do more washes. However, this seems negligible since only 3% use compact dishwashers.

- Non-full loading. It was not possible to quantify the capacity usage, as this was not asked in the survey and can't be measured from the plug profile. However, in the plug survey, where this was asked, 13% of respondents answered to fill the dishwasher based on how much dishes are available and another 26% said to fill it up, but not too full. Much in line with the VEWIN surveys in the Netherlands, where some 13% of households said to fill their dishwasher usually only for three quarters [149]. According the Richter et all about 20% of washes are not fully loaded [150], while Van Holsteijn and Kemna estimate that inefficient placing of dishes reduces a common 12 ps dishwasher to only about 9 ps (a reduction of a quarter), particularly when consumers might need clean dishes [146].

As can be seen in figure 4.18 the average number of weekly washes per household size and machine size type is plotted. The average number of weekly washes by household indicate a stepwise increase with each additional household member. While the type (hence size) of dishwasher could potentially influence the number of weekly washes, the results do not show much variation.

Based on this, family size can at least partially explain why households might need more than 4 washing sessions per week. However, only about a third of families are larger than 3 family members. With even 2 person households often already using the dishwasher more than the set threshold, literature indicates about a quarter of usage could potentially be reduced by better loading. Finally, trying to reduce the amount of dishes produced per family member on a daily basis could potentially close the remaining 10-15%. Reuse of glasses and mugs or even the breakfast plate could potentially contribute to reduction of the overall daily dish production when making people more aware, without much actual loss of convenience. Similar to the energy consumption, for the accuracy of the efficiency classification several factors could play a role see 5.3.3.

Potential energy savings

It was found that 29% of households use less than the set annual threshold of 218kWh, while the other 71% used more. With an average annual energy consumption of 312kWh this provides 94kWh of energy saving potential on average, a potential reduction of just over 30%. Only 7% of all households have efficient labelled dishwashers, make use of a low energy consuming washing program and wash less than 4 times a week, by that using 141kWh on average (24.5% below the threshold). Another 10% of households either consume more energy per wash or wash more often than the set threshold, but are slightly underneath the annual threshold still overall. Consuming more energy per wash (10%), washing more frequently (20%) or both (52%) results in an average of 9%, 33% and 61% overconsumption respectively.

This shows that about 7 in 10 households annually use more energy than the set threshold and less than 1 in 10 households is efficient on all factors. The average energy saving of 30% seems very much in line with the 29.4% higher energy use under real-life conditions, compared to laboratory assumptions as mentioned by Stamminger. Although it has to be noted that a large share of that (17.2%) is attributed to the energy consumption for pre-rinsing dishes by hand and energy loss due to inefficient usage frequency is not taken into account. [24] Indeed the number of washes seems to be a bigger driver than the energy usage per wash. Which can be explained by the average wash using about 1kWh of energy, while the step size between different energy labels is only about 0.1kWh. Hence increasing from 5 to 6 weekly washes equals the same effect as moving from an A+ to a B labelled dishwasher. Pakula, who did a research on washing machines, suggests total energy savings up to 50% with better use of the machine capacity and the frequent use of low wash temperatures mentioned as two focus areas for behavioural improvement. [9] The average savings of 30% and as high as 61% for the 52% of households with both inefficient usage frequency and energy per wash, suggests that similar energy savings might be achievable for dishwasher usage as well.

While the EU Eco regulation calculates dishwasher energy consumption as total energy per cycle, based on the number of place settings [134], an attempt could be made to translate this into consumption per item. As an overall threshold of 1.05kWh per wash was set. As the majority of dishwashers are 12/13 place settings sized, such dishwasher would offer space to about 130 pieces (based on table 2.3), this would equate to using about 8 Wh/item. In the research of [150] dishwasher usage was compared with manually washing the dishes and here an average consumption of 25Wh/item was found. However, in that research the calculated energy usage includes energy consumption of manually pre-rinsing dishes, loading the dishwasher below full capacity and using a higher temperature washing programme, without specifying the share of energy loss on each of these. Furthermore, as the research was conducted in 2011, it can be assumed that dishwashers at the time would generally have a higher energy consumption, since average energy consumption has decreased over time (see figure 2.10).

Pre-rinsing has not been taken into account within the scope of this study, but when subtracting some 20% for the age of the dishwasher (in accordance with figure 2.10) and some 20% for the pre-rinsing (in accordance with the energy consumption expressed in the study for manual washing of the dishes) this would reduce the finding to about 15Wh/item. The average energy saving potential of 30% found in this study, would place the actual consumption indeed very close this 15Wh/item. While this approach is not as precise as the total energy consumption per washing cycle, loading and other factors might vary per washing cycle and dishes are not homogeneous, but also vary in size and shape, this does provide an indication of the average used energy and potential average saving per item nevertheless.

Finally, it has to be noted that these potential savings are an aggregation of many different underlying factors. As can be seen in figure 4.20, household specific factors such as family size change the actual specific efficiency threshold that would be applicable for that household. Likewise changing a very inefficient dishwasher may only be attractive at the end of its lifetimes. Therefore, while changing a dishwasher in theory is possible, a household's dishwasher and it's efficiency label should rather be assumed as a given. Particularly when considering other factors such as material consumption and energy usage of appliance production. Adjusting for these "fixed" variables has not been considered within the scope of this research, but the sensitivity to these factors could be investigated as further research.

6

Conclusion

6.1 Conclusion introduction

The thesis report concludes with main findings on each research sub-question, before answering the main research question. The conclusion is followed by the implications of this research and ends with recommendations for future usage and potential further development.

Findings are summarised for the research question and sub-questions:

How can energy consumption, usage pattern and efficiency characteristics of real-life dishwasher usage in households be detected and analysed with a smart meter based NILM system?

1. How is electric energy usage and efficiency for dishwashers characterised based on traditional analysis?

2. How can a smart meter based NILM system be developed to detect this?

3. How does the newly developed system perform in detecting these characteristics?

4. How much energy do dishwashers in households consume and how often are they used, depending on time, household and machine characteristics?

5. How efficiently are dishwashers used in households?

6.2 Main findings

In this chapter the main findings of the thesis will be summarised and final conclusions related to the posed research questions will be drawn.

1. *How is electric energy usage and efficiency for dishwashers characterised based on traditional analysis?*

   The load profile of a dishwasher follows a distinctive pattern, based on different steps in the washing cycle, such as water heating and rinsing. The average energy consumption depends on factors such as chosen washing program, efficiency label and size resulting in an energy consumption of about 1.1 kWh per wash. Usage frequency depends on family size, amount of dishes used and capacity utilisation of the dishwasher. The average annual number of washes in the Eco Design regulation (No 1016/2010) [137] is assumed to be 280.

   Energy consumption and usage patterns are both analysed by more traditional approaches, such as surveys and smart plugs. Efficiency characteristics as used in the Eco Design and Efficiency Labelling regulation are based on findings from traditional approaches, but do not take the potential of scalable and continuous analysis into account that NILM could offer, potentially reducing the sensitivity to survey bias and providing much larger research samples than would be affordable with smart plugs.

2. *How can a smart meter based NILM system be developed to detect energy consumption, usage pattern and efficiency characteristics?*

   The model developed in this research extends the NILM methodology developed by K. Basu (2017), which provided the possibility of detecting dishwasher usage on the smart meter power signal. This research used 100 households with plugs connected to the dishwasher and smart meter data sampled at 10 seconds intervals to develop a more granular level of detection. Energy consumption per wash was estimated, using linear regression on several of the detected features such as total heating period, heating power and number of heating moments. Energy consumption as calculated from the plug data was used to calibrate the regression parameters of the NILM algorithm.

   The detection model can either be set to ensure high certainty of detecting the right appliance (high precision) or ensuring that most possible dishwasher cycles are detected, risking mislabelling of activity (high recall). The model was set to an optimal balance between the two, ensuring a maximum accuracy (F1) score. Since a higher precision improves the quality of detection of energy per wash, but results in underestimating the number of washes another linear regression model was used to estimate the total number of washes.

In order to assess the efficiency of the dishwasher use, a binary threshold model was developed, serving as a proxy for efficient energy consumption per wash in kWh and efficient usage frequency in number of washes per week. In order to establish values for the energy efficiency threshold the EU Eco Design regulation was used. These regulations contain calculations for maximum energy consumption of dishwashers based on efficiency label and size, assuming a standard washing program. For usage frequency only an assumption of 280 washes per year is available, but no information is given on what usage frequency could be considered as efficient. Therefore weekly usage was estimated with a model based on number of family members and daily usage of dishes. Both efficiency thresholds were calibrated to improve accuracy based on the 100 validation households.

3. *How does the newly developed system perform in detecting these (energy consumption, usage pattern and efficiency) characteristics?*

   The developed algorithm was tested for its accuracy to estimate per household outcomes compared to measurement from the smart plug. The average estimation error (RMSE) was found to be 0.10kWh for energy per wash and 1.4 days/week for the usage frequency. This translates into a normalized estimation error (NRMSE) of 8.8% for energy per wash and 27.2% for the usage frequency. Estimated energy consumption and frequency of usage were compared to a specified efficiency benchmark. This binary efficiency classification provides accurate results for both, and a set threshold can be optimized to either high precision or recall, depending on requirements, leading to an energy efficiency classification accuracy (F1 score) of 91% and (F1 score) of 89% for classification of usage frequency efficiency.

4. *How much energy do dishwashers in households consume and how often are they used, depending on time, household and machine characteristics?*

   To gain insight into the characteristics of dishwasher usage in real-life the developed algorithms were applied on smart meter data gathered for a full year for more than 130.000 households to investigate energy consumption, usage pattern and efficiency characteristics at large scale.

   - How much energy do dishwashers in households consume on average per wash how frequently are they used per week and per year and what is the resulting total energy consumption?

In the small-scale smart plug research, an average energy consumption per wash of 1.22kWh was found. Dishwashers in this group are used 5.2 times per week on average. The deployment of the developed method on large-scale smart meter data resulted in an average energy consumption of 1.18kWh per wash, equalling an A label machine and the dishwasher is used 240 times per year, 40 times less than assumed for the EU efficiency label.

- How does the energy consumption of dishwashers in households depend on time of the day, week and year?

  Energy per wash and frequency of usage both show a seasonal dependency with peaks in winter and lows in summer. Energy consumption per wash changes in accordance with outside temperature, deviating by 0.23 kWh between the maximum and the minimum throughout the year on average. A relation between the frequency of dishwasher usage and the occurrence of events and holidays can be drawn with a minimum of 3.3 during Summer holiday and a maximum of 5.3 during Christmas time. Within the weekly pattern, dishwasher usage differs per weekday with least usage on Friday. Main usage was identified to be directly after dinner time and just before bed time, with the highest peak on Monday after dinner and more equally distributed use over the weekend.

- How does the energy consumption of dishwashers depend on household and machine characteristics?

  The energy consumption of dishwashers depends on household and machine characteristics. Number of people within the household appeared to be one of the main factors for weekly usage with 2.8 weekly washes for single households, 4.4 for two people household and increasing by 0.6 for each additional household member. The temperature of the chosen washing program impacts the energy consumption more than the efficiency label. For high efficiency label and low temperature resulting in average energy consumption of 1.13 kWh (for A+++, <30°C) and energy consumption for low efficiency label and high temperature of 1.35 kWh (for A-, <75°C).

5. *How efficiently are dishwashers used in households?*

   The analysis of the frequency and energy efficiency showed that 84% of households consume more energy per wash and 62% use their machine more often than needed. This results overall in an average improvement potential of 94 kWh (just over 30%) that could be saved on dishwasher usage annually per household by the 130.000 households in this research.

How can energy consumption, usage pattern and efficiency characteristics of real-life dishwasher usage in households be detected and analysed with a smart meter based NILM system?

This research shows that energy consumption assessment with NILM, based on data collected from the smart meter, can deliver real-life insights into the energy consumption, usage patterns and efficiency characteristics of dishwashers in a continuous, scalable manner and without much intervention of the user.

It does so by combining a NILM dectection system, with estimation of energy usage per wash and estimation of usage frequency. These findings are compared with Eco-design regulations and bottom-up assumptions on daily dish production. A binary classifier is used to detect if energy usage per wash and weekly usage frequency are below or above a chosen thresholds that serves as proxy to classify inefficiency and potential energy savings.

In this way this research extends NILM research and contributes to the field of energy systems analysis by providing both a methodology and more detailed information on household energy consumption, usage patterns and efficiency.

While this developed NILM methodology might be less accurate to measure consumption on a wash-by-wash basis than using plugs, it is much more scalable, less intrusive and can provide the user with insight over a more extended period of washes and can trigger a dialogue with the user to gain additional information.
Compared to the use of surveys the findings are not biased by user statements and data can be collected on a continuous basis. These overarching statistics for many households give a good indication of average values and can potentially be grouped by household demographic, user behaviour and machine characteristics, but care must be taken to ensure the algorithm is trained with a randomly sampled unbiased set of households.

6.3 Outlook

6.3.1 Implications

The conducted research has shown that tapping into large-scale smart meter data is possible via NILM application and delivers new insights into the real-life usage of household appliances. By utilising data from the already installed smart meter, the developed NILM system can detect dishwasher energy consumption and usage patterns from the household load profile non-intruisively and compare this to a set threshold serving as proxy for efficiency.

The outcomes of this research could be used to further develop a system to inform consumers on how to improve efficient energy usage of their dishwasher, usage profiling and benchmarking between consumer groups, energy waste check and demand side management. Generally it needs to be considered that from the detections of appliance usage the activity profile of a household can be inferred. While this provides multiple opportunities, as discussed in this paper, this poses privacy concerns as well. These need to be considered particularity in connection to commercial consumer facing products and regulation (General Data Protection Regulation)

As stated in the beginning of the research, long-term energy consumption assessments are important to find ways to improve energy efficient behaviour. Continuous performance monitoring is key to understand the changes of energy consumption and derive suitable measures and action. The results of the consumption overview provide a new source of data to analyse these parameters throughout the year and can be used to compare former and future studies for large-scale measurements over longer periods. It therefore contributes to the field of knowledge about how energy consumption of households is broken down and at what time of the day is used. The dynamic nature of the research furthermore provides the opportunity for the analysis of changes in consumer behaviour and habits, e.g. trends occurring as a result of increased working from home.

In this study dishwashers have been used as a use case, but the developed methodology can be extended to other appliances within the household to disaggregate the energy usage into individual appliances or appliance groups. If extended to multiple appliances, this research can contribute further to understand how each appliance specific usage pattern contributes to the overall energy use of a household, in other words break-down the energy bill. This can be used as a basis to deploy an energy monitoring and advice system for households and other buildings like offices and research facilities, helping them to improve their energy efficiency. The efficiency feedback and communications could potentially even be extended with use cases such as consumption analytics and consumption forecasting

The developed methodology can be seen as a first building block, to enable NILM diaggregated singal assessment to become part of many more use cases for demand side management (DMS) offering the ability to consumers to better understand their appliances and reduce inefficient usage.

The large-scale overarching continuous energy statistics can be used by energy companies and policymakers to understand usage patterns in relation to societal events, make better informed decisions and expedient grid development and for instance to forecast and match demand patterns with volatility caused by increased deployment of sustainable energy technologies. Hence, be able to improve matching energy supply with demand on the energy markets. Regulations on appliance efficiency and ecodesign as well as future policy incentives could be guided and evaluated by research similar to the present, using quantitative real-life data for the most effective leverage and impact of the measures. Using this new type of analytical measures could help inform future decisions, that previously would only rely on survey and plug results.

6.3.2 Further research

Throughout the research multiple aspects have been identified, which could be further developed, extended or detailed. One potential entry point for future research could be the aim of improving the model by optimising the detection model separately for detection and energy consumption parameters. Currently the detection model aims to detect as good as possible, based on the F1 accuracy score. However, as the end product is the energy consumption and number of washes, the NILM algorithm could be adjusted to optimize for these instead. Potentially this could mean a seperate slightly differently calibrated version of the model would be needed for the energy per wash (where precision for the estimation features is more important) vs the number of washes (where recall is more important). Alternatively other NILM detection algorithms or a hybrid with other existing NILM algorithms could be used to improve the detection quality.

Furthermore, it would be beneficial if more training data would be available to further develop and test the model. The availability of more training data could help improve the accuracy of the detection model, particularly for yet unseen circumstances. Extra training data could also help further validate the generalisability of the model. In order to do this either external data sets could be used, but these risk to be too different from the target group. Ideally therefore more households would receive plugs. For instance as temporary service for households where the survey results and meter estimates do not match up or for a target group specifically selected for their representative characteristics. Alternatively, due to high cost and logistic complexity of such an endeavour, a more scalable option would be to build a feedback button as part of the consumer facing application. Consumers could tap the button within the application on their smartphone every time they activate (and stop) their appliance, by that annotating their household profile. This could be a cost effective, scalable approach to improve recognition of appliances for each individual household. The advent of internet connected appliances (IoT) could further aid and automate these endeavours.

Another building block to create a more elaborate reference for the efficiency classification would be to base the classification on the survey results, allowing for a threshold development based on e.g. household size and machine type.

In addition to the surveys conducted along with the measurements, the aspects for the contextualisation of the energy measurements could be extended. When connecting the households location data to e.g. the Dutch Central Bureau of Statistics (CBS) data, information on rural/urban split, demographics and socio-economic factors such as income could be assessed on neighbourhood level. These inputs could create valuable insights into more stakeholder relevant energy information.

One aspect that has been mentioned above is the extension of the research for different appliances. In fact a large share of assessment has already been carried out for washing machines using the similar methodology and a start was made for dryers and refrigerators. However, to be able to completely describe all steps of the analytic pipeline, only a single appliance type (dishwashers) was chosen to be analysed in detail as a use case. With small adaptations the developed model can be further extended to other household appliances on the energy bill. A similar framework could then be used to analyse other major appliances and not only for electric appliances, but for gas consumption as well.

Finally, researching the impact of feedback on the usage behaviour in further detail is key to better tailor this feedback to the households. One possibility to optimise this could be by building on the existing tool with which the survey results for the meter data were retrieved and setting up trial runs with different messages and the respective consumer responses. As this requires a more design and behavioural scientific focus, this was not considered within the scope of this research, but the user interaction would be of essential importance to successfully deploy the developed system within a consumer facing context.

Bibliography

[1] Stephen Galsworthy and Kaustav Basu. It's raining patents for quby. *Quby. Outsmarting energy*, 26.9.2019.

[2] Netherlands Enterprise Agency. Patent 2020228: Detecting inefficient appliances, 2019.

[3] Annemarie C Kerkhof, Sanderine Nonhebel, and Henri C Moll. Relating the environmental impact of consumption to household expenditures: An input–output analysis. *Ecological Economics*, 68(4):1160–1170, 2009.

[4] Edgar G Hertwich and Glen P Peters. Carbon footprint of nations: A global, trade-linked analysis. *Environmental science & technology*, 43(16):6414–6420, 2009.

[5] Intergovernmental Panel on Climate Change (IPCC). Synthesis report: Contribution of working groups i. ii and iii to the fifth assessment report. *Proceedings of the IPCC*, pages 1–167, 2014.

[6] International Energy Agency (IEA). Energy efficiency indicators: Essentials for policy making. 2014.

[7] Katarzyna Grondys, Armenia Androniceanu, and Zdzisława Dacko-Pikiewicz. Energy management in the operation of enterprises in the light of the applicable provisions of the energy efficiency directive (2012/27/eu). *Energies*, 13(17):4338, 2020.

[8] Maria da Graça Carvalho. Eu energy and climate change strategy. *Energy*, 40(1):19–22, 2012.

[9] Christiane Pakula and Rainer Stamminger. Energy and water savings potential in automatic laundry washing processes. *Energy Efficiency*, 8(2):205–222, 2015.

[10] Marcos Pelenur. Household energy use: a study investigating viewpoints towards energy efficiency technologies and behaviour. *Energy Efficiency*, 11(7):1825–1846, 2018.

[11] Kees Vringer, Theo Aalbers, and Kornelis Blok. Household energy requirement and value patterns. *Energy Policy*, 35(1):553–566, 2007.

[12] Kees Vringer and Kornelis Blok. Long-term trends in direct and indirect household energy intensities: a factor in dematerialisation? *Energy Policy*, 28(10):713–727, 2000.

[13] Wokje Abrahamse and Linda Steg. How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? *Journal of economic psychology*, 30(5):711–720, 2009.

[14] Wokje Abrahamse and Linda Steg. Factors related to household energy use and intention to reduce it: The role of psychological and socio-demographic variables. *Human ecology review*, pages 30–40, 2011.

[15] Iana Vassileva, Fredrik Wallin, and Erik Dahlquist. Understanding energy consumption behavior for future demand response strategy development. *Energy*, 46(1):94–100, 2012.

[16] Kevin Burchell, Ruth Rettie, and Tom C Roberts. Householder engagement with energy consumption feedback: the role of community action and communications. *Energy Policy*, 88:178–186, 2016.

[17] Iana Vassileva, Erik Dahlquist, Fredrik Wallin, and Javier Campillo. Energy consumption feedback devices' impact evaluation on domestic energy use. *Applied Energy*, 106:314–320, 2013.

[18] Peter Palensky and Dietmar Dietrich. Demand side management: Demand response, intelligent energy systems, and smart loads. *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Industrial Informatics*, 7(3):381–388, 2011.

[19] Navigant Consulting. Market data: Advanced thermostats: Residential smart and communicating thermostats for single-family and multifamily buildings: Global market analysis and forecasts, 2018.

[20] Stefan Pfenninger, Adam Hawkes, and James Keirstead. Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews*, 33:74–86, 2014.

[21] Adeel Abbas Zaidi, Friederich Kupzog, Tehseen Zia, and Peter Palensky. Load recognition for automated demand response in microgrids. In *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society*, pages 2442–2447, Piscataway, N.J., 2010.

[22] Dietmar Bruckner, Jan Haase, Peter Palensky, and Gerhard Zucker. Latest trends in integrating building automation and smart grids. *Industrial Electronics Conference (IECON) Proceedings*, pages 6285–6290, 2012.

[23] Van Holsteijn and Kemna BV. Basisdocument: Elektrische apparatuur in nederlandse huishoudens (1995 - 2020), 2008.

[24] Rainer Stamminger. Preparatory studies for eco-design requirements of eups lot 14: Domestic washing machines and dishwashers. *Part I-present situation Task*, 3, 2007.

[25] Millieu Centraal. Koelkasten en vriezers: Koop- en bespaartips, 2020.

[26] ECN, Energie-Nederland, and Netbeheer Nederland. Energietrends 2016.

[27] Bodil M Larsen and Runa Nesbakken. Household electricity end-use consumption: results from econometric and engineering models. *Energy Economics*, 26(2):179–200, 2004.

[28] Jean Paul Zimmermann. End-use metering campaign in 400 households in sweden assessment of the potential electricity savings, 2009.

[29] Ad Almeida et al. Final report remodece-residential monitoring to decrease energy use and carbon emissions in europe (ieea program funded project). *University of Coimbra, Dep. of Electrical Engineering, Pólo II*, 3030:290, 2008.

[30] Department for Environment, Food, and Rural Affairs (Defra). Uk household electricity survey: A study of domestic electrical product usage, 2012.

[31] Maarten R Staats. Using wet appliances for demand side management: current state assessment and potential exploration. Master's thesis, DNV, 2015.

[32] Thomas Picon, Mohamed Nait Meziane, Philippe Ravier, Guy Lamarque, Clarisse Novello, Jean-Charles Le Bunetel, and Yves Raingeaud. Controlled on/off loads library (cooll), a public dataset of high-sampled electrical signals for appliance identification. *arXiv preprint arXiv:1611.05803*, 2016.

[33] Iman Mansouri, Marcus Newborough, and Douglas Probert. Energy consumption in uk households: Impact of domestic electrical appliances. *Applied Energy*, 54(3):211–285, 1996.

[34] Christina O Gibson, Linda R Murphy, Gary J Stanko, and Gary M Shapiro. *Interaction of survey questions as it relates to interview-respondent bias.* US Census Bureau [custodian], 1978.

[35] Edith Molenbroek, Paul Waide, Matthew Smith, Heleen Groenenberg, Sophie Attali, Corinna Fischer, Juraj Krivošik, Paula Fonseca, Bruno Santos, and João Fong. Evaluation of the energy labelling directive and specific aspects of the ecodesign directive: Ener/c3/2012-523: Final technical report, 2014.

[36] Muriel Dupret and Jean-Paul Zimmermann. Electricity consumption of cold appliances, washing machines, dish washers, tumble driers and air conditioners. on site-monitoring campaign in 100 ousehold. analysis of the evolution of the consumption over the last 20 years. pages 1501–1509, 2017.

[37] Fritz J Roethlisberger, William J Dickson, and Harold A Wright. *Management and the worker: An account of a research program conducted by the Western electric Company, Hawthorne Works, Chicago*. Harvard Univ. Press, Cambridge, Mass., 16. printing edition, 1975.

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

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

[40] José A Hoyo-Montaño, Carlos A Pereyda-Pierre, Jesús M Tarín-Fontes, and Jesús N Leon-Ortega. Overview of non-intrusive load monitoring: A way to energy wise consumption. In *2016 13th International Conference on Power Electronics (CIEP)*, pages 221–226. IEEE, 2016.

[41] Kanghang He, Lina Stankovic, Jing Liao, and Vladimir Stankovic. Non-intrusive load disaggregation using graph signal processing. *IEEE Transactions on Smart Grid*, 9(3):1739–1747, 2016.

[42] Shikha Singh and Angshul Majumdar. Deep sparse coding for non–intrusive load monitoring. *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Smart Grid*, 9(5):4669–4678, 2018.

[43] Kaustav Basu, Vincent Debusschere, Seddik Bacha, Ahmad Hably, Danny van Delft, and Geert Jan Dirven. A generic data driven approach for low sampling load disaggregation. *Sustainable Energy, Grids and Networks*, 9:118–127, 2017.

[44] Antonio Ruano, Alvaro Hernandez, Jesus Ureña, Maria Ruano, and Juan Garcia. Nilm techniques for intelligent home energy management and ambient assisted living: A review. *Energies*, 12(11):2203, 2019.

[45] Fernando D Garcia, Wesley A Souza, Ivando S Diniz, and Fernando P Marafão. Nilm-based approach for energy efficiency assessment of household appliances. *Energy Informatics*, 3(1):1–21, 2020.

[46] Haroon Rashid, Pushpendra Singh, Vladimir Stankovic, and Lina Stankovic. Can non-intrusive load monitoring be used for identifying an appliance's anomalous behaviour? *Applied energy*, 238:796–805, 2019.

[47] Xiangyu Kong, Shijian Zhu, Xianxu Huo, Shupeng Li, Ye Li, and Siqiong Zhang. A household energy efficiency index assessment method based on non-intrusive load monitoring data. *Applied Sciences*, 10(11):3820, 2020.

[48] Stichting Energieonderzoek Centrum Nederland. Het elektriciteitsverbruikpatroon in nederland nader geanalyseerd, 1988.

[49] João Gouveia. *Residential Sector Energy Consumption at the Spotlight: From Data to Knowledge*. PhD thesis, 2017.

[50] Jukka V Paatero and Peter D Lund. A model for generating household electricity load profiles. *International Journal of Energy Research*, 30(5):273–290, 2006.

[51] Joakim Widén. Improved photovoltaic self-consumption with appliance scheduling in 200 single-family buildings. *Applied Energy*, 126:199–212, 2014.

[52] Ian Richardson, Murray Thomson, David Infield, and Conor Clifford. Domestic electricity use: A high-resolution energy demand model. *Energy and buildings*, 42(10):1878–1887, 2010.

[53] Antonio Ridi, Christophe Gisler, and Jean Hennebert. A survey on intrusive load monitoring for appliance recognition. In *2014 22nd international conference on pattern recognition*, pages 3702–3707. IEEE, 2014.

[54] Fraunhofer. Nonintrusive load monitoring - entwicklung eines gerätespezifischen energiemanagements für gewerbebetriebe, 2020.

[55] Zico Kolter and Tommi Jaakkola. Approximate inference in additive factorial hmms with application to energy disaggregation. In *Artificial intelligence and statistics*, pages 1472–1482, 2012.

[56] Jorge Revuelta Herrero, Álvaro Lozano Murciego, Alberto López Barriuso, Daniel La Hernández de Iglesia, Gabriel Villarrubia González, Juan Manuel Corchado Rodríguez, and Rita Carreira. Non intrusive load monitoring (nilm): A state of the art. In *Trends in Cyber-Physical Multi-Agent Systems. The PAAMS Collection - 15th International Conference, PAAMS 2017*, volume 619, pages 125–138. 2018.

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

[58] Christoforos Nalmpantis and Dimitris Vrakas. Machine learning approaches for non-intrusive load monitoring: from qualitative to quantitative comparation. *Artificial Intelligence Review*, 52(1):217–243, 2019.

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

[60] Oliver Parson, Siddhartha Ghosh, Mark Weal, and Alex Rogers. Non-intrusive load monitoring using prior models of general appliance types. In *Proceedings of the Twenty-Sixth Conference on Artificial Intelligence (AAAI)*, pages 356–362. AAAI Press, Toronto, Ontario, Canada, 2012.

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

[62] Energy management using non-intrusive load monitoring techniques – state-of-the-art and future research directions. *Sustainable Cities and Society*, 62:102411, 2020.

[63] Ting Su and Jennifer G Dy. In search of deterministic methods for initializing k-means and gaussian mixture clustering. *Intelligent Data Analysis*, 11(4):319–338, 2007.

[64] Antonio Ruano, Alvaro Hernandez, Jesus Ureña, Maria Ruano, and Juan Garcia. Nilm techniques for intelligent home energy management and ambient assisted living: A review. *Energies*, 12(11):2203, 2019.

[65] Giuseppe Tina and Alex Amenta. Consumption awareness for energy savings: Nialm algorithm efficiency evaluation. In *Renewable Energy Congress (IREC), 2014 5th International*, pages 1–6. Institute of Electrical and Electronics Engineers (IEEE), 2014.

[66] Leen de Baets, Chris Develder, Tom Dhaene, and Dirk Deschrijver. Detection of unidentified appliances in non-intrusive load monitoring using siamese neural networks. *International Journal of Electrical Power & Energy Systems*, 104:645–653, 2019.

[67] Zuyi Li and Mengqi Lu. Dataset for paper "a hybrid event detection approach for non-intrusive load monitoring", 2019.

[68] Martin Pullinger, Jonathan Kilgour, Nigel Goddard, Niklas Berliner, Lynda Webb, Myroslava Dzikovska, Heather Lovell, Janek Mann, Charles Sutton,

Janette Webb, et al. The ideal household energy dataset, electricity, gas, contextual sensor data and survey data for 255 uk homes. *Scientific Data*, 8(1):1–18, 2021.

[69] David Murray, Lina Stankovic, and Vladimir Stankovic. An electrical load measurements dataset of united kingdom households from a two-year longitudinal study. *Scientific data*, 4(1):1–12, 2017.

[70] Andreas Reinhardt, Olivia Morar, Silvia Santini, Sebastian Zöller, and Ralf Steinmetz. Cbfr: Bloom filter routing with gradual forgetting for tree-structured wireless sensor networks with mobile nodes. In *2012 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pages 1–9, 2012.

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

[72] J Zico Kolter and Matthew J Johnson. Redd: A public data set for energy disaggregation research. In *Workshop on data mining applications in sustainability (SIGKDD), San Diego, CA*, volume 25, pages 59–62, 2011.

[73] Jack Kelly and William Knottenbelt. The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes. *Scientific data*, 2(1):1–14, 2015.

[74] Stephen Makonin, Fred Popowich, Lyn Bartram, Bob Gill, and Ivan V Bajić. Ampds: A public dataset for load disaggregation and eco-feedback research. In *2013 IEEE electrical power & energy conference*, pages 1–6. IEEE, 2013.

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

[76] Adrian Filip et al. Blued: A fully labeled public dataset for event-based nonintrusive load monitoring research. In *2nd workshop on data mining applications in sustainability (SustKDD)*, volume 2012, 2011.

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

[78] Akshay SN Uttama Nambi, Antonio Reyes Lua, and Venkatesha R Prasad. Loced: Location-aware energy disaggregation framework. In *Proceedings of the 2nd acm international conference on embedded systems for energy-efficient built environments*, pages 45–54, 2015.

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

[80] Matthias Kahl, Anwar Ul Haq, Thomas Kriechbaumer, and Hans-Arno Jacobsen. Whited-a worldwide household and industry transient energy data set. In *3rd International Workshop on Non-Intrusive Load Monitoring*, pages 1–4, 2016.

[81] Richard Jones, Alejandro Rodriguez-Silva, and Stephen Makonin. Increasing the accuracy and speed of universal non-intrusive load monitoring (unilm) using a novel real- time steady-state block filter. pages 1–5, 2020.

[82] Na Sadeghianpourhamami, Joeri Ruyssinck, Dirk Deschrijver, Tom Dhaene, and Chris Develder. Comprehensive feature selection for appliance classification in nilm. *Energy and Buildings*, 151:98–106, 2017.

[83] Hui Liu. Non-intrusive load monitoring. *Theory, Technologies and Applications.*, 2020.

[84] Bruna M. Mulinari, Daniel P. de Campos, Clayton H. da Costa, Hellen C. Ancelmo, André E. Lazzaretti, Elder Oroski, Carlos R. E. Lima, Douglas P. B. Renaux, Fabiana Pottker, and Robson R. Linhares. A new set of steady-state and transient features for power signature analysis based on v-i trajectory. In *2019 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)*, pages 1–6, 2019.

[85] Zhenyu Wang and Guilin Zheng. Residential appliances identification and monitoring by a nonintrusive method. *IEEE transactions on Smart Grid*, 3(1):80–92, 2011.

[86] Benjamin Wild, Karim Said Barsim, and Bin Yang. A new unsupervised event detector for non-intrusive load monitoring. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 73–77, 2015.

[87] Anthony Faustine, Nerey Henry Mvungi, Shubi Felix Kaijage, and Michael Kisangiri. A survey on non-intrusive load monitoring methodies and techniques for energy disaggregation problem. *CoRR*, abs/1703.00785, 2017.

[88] Marwa Hamdi, Hassani Messaoud, and Nasreddine Bouguila. A new approach of electrical appliance identification in residential buildings. *Electric Power Systems Research*, 178:106037, 2020.

[89] Sarra Houidi, François Auger, Houda Ben Attia Sethom, Dominique Fourer, and Laurence Miègeville. Multivariate event detection methods for non-intrusive load monitoring in smart homes and residential buildings. *Energy and Buildings*, 208:109624, 2020.

[90] Bo Liu, Wenpeng Luan, and Yixin Yu. Dynamic time warping based non-intrusive load transient identification. *Applied energy*, 195:634–645, 2017.

[91] Luis Rueda, Alben Cardenas, Sousso Kelouwani, and Kodjo Agbossou. Transient event classification based on wavelet neuronal network and matched filters. In *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, pages 832–837, 2018.

[92] Zhuang Zheng, Hainan Chen, and Xiaowei Luo. A supervised event-based non-intrusive load monitoring for non-linear appliances. *Sustainability*, 10(4):1001, 2018.

[93] Milad Afzalan, Farrokh Jazizadeh, and Jue Wang. Self-configuring event detection in electricity monitoring for human-building interaction. *Energy and Buildings*, 187:95–109, 2019.

[94] Kaustav Basu, Vincent Debusschere, Seddik Bacha, Ujjwal Maulik, and Sanghamitra Bondyopadhyay. Nonintrusive load monitoring: A temporal multilabel classification approach. *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Industrial Informatics*, 11(1):262–270, 2015.

[95] Michael Baranski and Jürgen Voss. Non-intrusive appliance load monitoring based on an optical sensor. In *2003 IEEE Bologna PowerTech*, pages 267–274. Institute of Electrical and Electronics Engineers (IEEE), 2004.

[96] Jian Liang, Simon Ng, Gail Kendall, and John Cheng. Load signature study—part i: Basic concept, structure, and methodology. *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Power Delivery*, 25(2):551–560, 2010.

[97] Shinkichi Inagaki, Tsukasa Egami, Tatsuya Suzuki, Hisahide Nakamura, and Koichi Ito. Nonintrusive appliance load monitoring based on integer programming. *Electrical Engineering in Japan*, 174(2):18–25, 2011.

[98] Lucas Pereira and Nuno Nunes. Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—a review. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, 8(6):e1265, 2018.

[99] Ying-Xun Lai, Chin-Feng Lai, Yueh-Min Huang, and Han-Chieh Chao. Multi-appliance recognition system with hybrid svm/gmm classifier in ubiquitous smart home. *Information Sciences*, 230:39–55, 2013.

[100] Hsueh-Hsien Chang, Po-Ching Chien, Lung-Shu Lin, and Nanming Chen. Feature extraction of non-intrusive load-monitoring system using genetic algorithm in smart meters. In *IEEE 8th International Conference on e-Business Engineering (ICEBE), 2011*, pages 299–304, Piscataway, NJ, 2011. Institute of Electrical and Electronics Engineers (IEEE).

[101] Daisy H. Green, Steven R. Shaw, Peter Lindahl, Thomas J. Kane, John S. Donnal, and Steven B. Leeb. A multiscale framework for nonintrusive load identification. *IEEE Transactions on Industrial Informatics*, 16(2):992–1002, 2020.

[102] Antonio Ruzzelli, C Nicolas, Anton Schoofs, and Gregory O'Hare. Real-time recognition and profiling of appliances through a single electricity sensor. In *7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2010*, pages 1–9, Piscataway, NJ, 2010. Institute of Electrical and Electronics Engineers (IEEE).

[103] Tehseen Zia, Dietmar Bruckner, and Adeel Zaidi. A hidden markov model based procedure for identifying household electric loads. In *IECON 2011*, pages 3218–3223, Piscataway, NJ, 2011. Institute of Electrical and Electronics Engineers (IEEE).

[104] Hyungsul Kim, Manish Marwah, Martin Arlitt, Geoff Lyon, and Jiawei Han. Unsupervised disaggregation of low frequency power measurements. In Bing Liu, Huan Liu, Christopher Wade Clifton, Takashi Washio, and Chandrika Kamath, editors, *Proceedings of the 2011 SIAM International Conference on Data Mining*, pages 747–758, Philadelphia, Pennsylvania, 2011. Society for Industrial and Applied Mathematics.

[105] Takekazu Kato, Hyun Sang Cho, Dongwook Lee, Tetsuo Toyomura, and Tatsuya Yamazaki. Appliance recognition from electric current signals for information-energy integrated network in home environments. In *Ambient Assistive Health and Wellness Management in the Heart of the City*, volume 5597 of *SpringerLink: Springer e-Books*, pages 150–157. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[106] Gu-yuan Lin, Shih-chiang Lee, Jane Yung-jen Hsu, and Wan-rong Jih. Applying power meters for appliance recognition on the electric panel. In *The 5th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2010*, pages 2254–2259, Piscataway, NJ, 2010. Institute of Electrical and Electronics Engineers (IEEE).

[107] Alan Marchiori, Douglas Hakkarinen, Qi Han, and Lieko Earle. Circuit-level load monitoring for household energy management. *Institute of Electrical and Electronics Engineers (IEEE) Pervasive Computing*, 10(1):40–48, 2011.

[108] Jing Liao, Georgia Elafoudi, Lina Stankovic, and Vladimir Stankovic. Power disaggregation for low-sampling rate data. In *2nd International non-intrusive appliance load monitoring workshop, Austin, TX*, volume 1, page F1, 2014.

[109] Mario Berges, Ethan Goldman, H Scott Matthews, Lucio Soibelman, and Kyle Anderson. User-centered nonintrusive electricity load monitoring for residential buildings. *Journal of computing in civil engineering*, 25(6):471, 2011.

[110] C. Laughman, Kwangduk Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong. Power signature analysis. *Institute of Electrical and Electronics Engineers (IEEE) Power and Energy Magazine*, 1(2):56–63, 2003.

[111] Roberto Bonfigli, Andrea Felicetti, Emanuele Principi, Marco Fagiani, Stefano Squartini, and Francesco Piazza. Denoising autoencoders for non-intrusive load monitoring: improvements and comparative evaluation. *Energy and Buildings*, 158:1461–1474, 2018.

[112] Roberto Bonfigli and Stefano Squartini. *Machine learning approaches to non-intrusive load monitoring*. Springer, 2020.

[113] Hui Liu. Hidden markov models based appliance. In *Non-intrusive Load Monitoring*, pages 163–190. Springer, 2020.

[114] Hui Liu. Deep learning based appliance identification. In *Non-intrusive Load Monitoring*, pages 191–214. Springer, 2020.

[115] Peng Xiao and Samuel Cheng. Neural network for nilm based on operational state change classification, 2019.

[116] Pedro Paulo Marques do Nascimento. Applications of deep learning techniques on nilm. *Diss. Universidade Federal do Rio de Janeiro*, 2016.

[117] Michael Devlin and Barry P Hayes. Non-intrusive load monitoring using electricity smart meter data: A deep learning approach. In *2019 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2019.

[118] Michele D'Incecco, Stefano Squartini, and Mingjun Zhong. Transfer learning for non-intrusive load monitoring. *IEEE Transactions on Smart Grid*, 11(2):1419–1429, 2019.

[119] Weicong Kong, Zhaoyang Dong, Yan Xu, and David Hill. An enhanced bootstrap filtering method for non-intrusive load monitoring. In *2016 Electrical and Electronics Engi- neers (IEEE) Power and Energy Society General Meeting (PESGM)*, pages 1–5, 2016.

[120] Yi-Wen Chen and Wei-Yu Chiu. A framework for a consumer-end energy management system in smart grid. In *2015 IEEE 4th Global Conference on Consumer Electronics (GCCE)*, pages 101–103. Institute of Electrical and Electronics Engineers(IEEE), 2015.

[121] Fengji Luo, Gianluca Ranzi, Weicong Kong, Zhao Yang Dong, Shu Wang, and Junhua Zhao. Non-intrusive energy saving appliance recommender system for smart grid residential users. *The Institution of Engineering and Technology (IET) Generation, Transmission & Distribution*, 11(7):1786–1793, 2017.

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

[123] Peter Bruce, Andrew Bruce, and Peter Gedeck. *Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python*. O'Reilly Media, 2020.

[124] Rene Kemna, Martijn van Elburg, William Li, and Rob van Holsteijn. Basisonderzoek elektriciteitsverbruik kleinverbruikers 1990 t/m 2000 (multiple reports), 2005.

[125] Group for Efficient Appliance (GEA). Dishwashers. long term efficiency targets. a technical and economic analysis, 1995.

[126] European Commission. Directorate-General for Energy. *Doing More with Less: Green Paper on Energy Efficiency*. Luxembourg: Office for Official Publications of the European Communities, 2005.

[127] Cleaning Institute. Understanding dishwashers, 2021.

[128] Rainer Stamminger, Gereon Broil, Christiane Pakula, Heiko Jungbecker, Maria Braun, Ina Rüdenauer, and Christoph Wendker. Synergy potential of smart appliances. *Report of the Smart-A project*, pages 1949–3053, 2008.

[129] Emir Lasic, Rainer Stamminger, Christian Nitsch, and Arnd Kessler. Construction of a virtual washing machine. *Tenside Surfactants Detergents*, 52(3):193–200, 2015.

[130] Rainer Stamminger. Quby power analysis: Dishwashers, non-published, 2017.

[131] Sean Barker, Sandeep Kalra, David Irwin, and Prashant Shenoy. Empirical characterization and modeling of electrical loads in smart homes. In *2013 International Green Computing Conference Proceedings*, pages 1–10. Institute of Electrical and Electronics Engineers (IEEE), 2013.

[132] Manisa Pipattanasomporn, Murat Kuzlu, Saifur Rahman, and Yonael Teklu. Load profiles of selected major household appliances and their demand response opportunities. *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Smart Grid*, 5(2):742–750, 2014.

[133] European Parliament and Council of the European Union. Directive (eu) 2018/2002 of the european parliament and of the council of 11 december 2018 amending directive 2012/27/eu on energy efficiency. *EU on energy efficiency*, 2018.

[134] Gregor Erbach. Briefing: Understanding energy efficiency, 2015. European Parliamentary Research Service.

[135] Kornelis Blok and Evert Nieuwlaar. *Introduction to Energy Analysis*. Routledge, 2017.

[136] European Commission. Directive 2009/125/ec of the european parliament and of the council of 21 october 2009 establishing a framework for the setting of ecodesign requirements for energy-related products. *Official Journal of the European Union*, 285, 2009.

[137] European Commission. Commission regulation (eu) no 1016/2010 of 10 november 2010 implementing directive 2009/125/ec of the european parliament and of the council with regard to ecodesign requirements for household dishwashers text with eea relevance. *Official Journal of the European Union*, pages 31—40, 2010.

[138] The European Commission. Commission delegated regulation (eu) 2019/2022 of 1 october 2019 laying down ecodesign requirements for household dishwashers pursuant to directive 2009/125/ec of the european parliament and of the council amending commission regulation (ec) no 1275/2008 and repealing commission regulation (eu) no 1016/2010. *Official Journal of the European Union, L 315*, 267:267–284, 2019.

[139] European Commission. Regulation (eu) 2017/1369 of the european parliament and of the council of 4 july 2017 setting a framework for energy labelling and repealing directive 2010/30/eu. *Official Journal of the European Union, L198*, 60:1–24, 2017.

[140] European Commission. Commission delegated regulation (eu) no 1059/2010 of 28 september 2010 supplementing directive 2010/30/eu of the european parliament

and of the council with regard to energy labelling of household dishwashers text with eea relevance. *Official Journal of the European Union*, 53:1–16, 2010.

[141] European Commission. Commission delegated regulation (eu) 2019/2017 of 11 march 2019 supplementing regulation (eu) 2017/1369 of the european parliament and of the council with regard to energy labelling of household dishwashers and repealing commission delegated regulation (eu) no 1059/2010. *Official Journal of the European Union*, 62:134–154, 2019.

[142] European Commission. Directive 2012/27/eu of the european parliament and of the council of 25 october 2012 on energy efficiency, amending directives 2009/125/ec and 2010/30/eu and repealing directives 2004/8/ec and 2006/32. *Official Journal of the European Union*, 315:1–56, 2012.

[143] The European Commission. Directive 2009/72/ec of the european parliament and of the council of 13 july 2009 concerning common rules for the internal market in electricity and repealing directive 2003/54/ec. *Official Journal of the European Union*, 211:55–93, 2009.

[144] European Committee for Electrotechnical Standardization. Household electric dishwashers - measurement method for performance characteristics, 2018.

[145] Rainer Stamminger. Modelling resource consumption for laundry and dish treatment in individual households for various consumer segments. *Energy efficiency*, 4(4):559–569, 2011.

[146] Van Holsteijn, Kemna BV, Vlaamse Instelling voor Technologisch Onderzoek, Viegand Maagoe AS, and Energie GmbH Wuppertal Institut für Klima, Umwelt. Cold appliances, washing machines, dishwashers, washer-dryers, lighting, set-top boxes and pumps, 2014.

[147] Angelika Schmitz and Rainer Stamminger. Usage behaviour and related energy consumption of european consumers for washing and drying. *Energy Efficiency*, 7(6):937–954, 2014.

[148] Angelika Schmitz, Farnaz Alborzi, and Rainer Stamminger. Large washing machines are not used efficiently in europe. *Tenside Surfactants Detergents*, 53(3):227–234, 2016.

[149] Henk Foekema and Lisanne van Thiel. Watergebruik thuis 2016. *Amsterdam: TNS NIPO, commissioned by VEWIN*, 2016.

[150] Christian Paul Richter. Usage of dishwashers: observation of consumer habits in the domestic environment. *International journal of consumer studies*, 35(2):180–186, 2011.

[151] Jasmin Geppert and Rainer Stamminger. Do consumers act in a sustainable way using their refrigerator? the influence of consumer real life behaviour on the energy consumption of cooling appliances. *International Journal of Consumer Studies*, 34(2):219–227, 2010.

[152] Jasmin Geppert and Rainer Stamminger. Analysis of effecting factors on domestic refrigerators' energy consumption in use. *Energy Conversion and Management*, 76:794–800, 2013.

[153] Öko Institut. Energieverbrauch von spülmaschinen, 2012.

[154] Rainer Stamminger, Alexandra Barth, and Susanne Dörr. Old washing machines wash less efficiently and consume more resources. *Hauswirtschaft und Wissenschaft*, 3:2005, 2005.

[155] European Parliament and L119 Council of the European Union. General data protection regulation, 2016.

[156] Erika L. Moen, Catherine J. Fricano-Kugler, Bryan W. Luikart, and A. James O'Malley. Analyzing clustered data: Why and how to account for multiple observations nested within a study participant? *PloS one*, 11(1):e0146721, 2016.

[157] Wojciech Kwedlo. A new random approach for initialization of the multiple restart em algorithm for gaussian model-based clustering. *Pattern Analysis and Applications*, 18(4):757–770, 2015.

[158] Matthew Stewart. The actual difference between statistics and machine learning. *Towards Data Science. Retrieved from*, 2019.

[159] Danilo Bzdok, Naomi Altman, and Martin Krzywinski. Points of significance: statistics versus machine learning, 2018.

[160] Joel Grus. *Data science from scratch: first principles with python*. O'Reilly Media, 2019.

[161] Koninklijk Nederlands Meteorologisch Instituut (KNMI). Daggegevens van het weer in nederland, 2020.

[162] European Commission. In focus: The improved eu energy label – paving way for more innovative and energy efficient products, 2021.

[163] Consumentenbond. Vaatwassers vergelijken, 2017.

[164] J Martin Bland and Douglas G Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476):307–310, 1986.

[165] Government NL. school holidays, 2018.

7

Appendix

Appendix A

Energy efficiency label

This section provides additional detail on relevant considerations connected to the EU efficiency label for dishwashers [140] and the retrieved methodology, that go beyond or have not been covered in the main body of the thesis.

A.0.1 Recent developments EU energy efficiency label

While the analysis in this research is based on data from 2017 and 2018 it is important to note that in March 2019 the EU commission finalised the format and appearance of the new Energy Efficiency Label for dishwashers. The new regulation was introduced to stores by 1 March 2021, following the Commission Delegated Regulation (EU) 2019/2017 [141] and thereby repealed Commission Delegated Regulation (EU) No 1059/2010 [140]. While this means this regulation is not applicable for the dishwashers analysed in this research, the framework could be extended to be applicable for newer machines as well, potentially by asking consumers if their machine was bought after March 2021.

In the updated regulation, labels return to the "A-G" energy efficiency scale for products (instead of labelling A+, A++, A+++). [162]. Along with launching the new labels, the commission Regulation (EU) No 1016/2010 on ecodesign requirements for household dishwashers [137] was repealed by Commission Regulation (EU) 2019/2022 [138]. In the frame of the update, changes were made to the energy calculations. The Energy Efficiency Index (EEI) is no longer based on the annual energy consumption, but now expressed as eco program energy consumption (EPEC), divided by standard program energy consumption (SPEC), both in kWh/cycle:

$$EEI = (\frac{EPEC}{SPEC}) * 100 \tag{A.1}$$

The EPEC calculation was simplified down to only considering the eco program energy consumption per wash instead of assuming an annual 280 washes. Complexity was further reduced by taking out stand-by mode energy consumption. The power consumption in off mode P_0 and power consumption in standby mode P_{sm} are now calculated separately.

The benchmark for the best available technology on the market for household dishwashers with 13 place settings was lowered to an energy consumption of 0,55 kWh/cycle. It is to be considered that about 2-3% of the reduction is corresponding to eliminating the low power modes from the EEI equation.

In accordance adaptions were made to the calculation of the SPEC to now calculate the kWh/cycle instead of per year:

1. for household dishwashers with rated capacity $ps \geq 10$ and width $> 50cm$:

$$SPEC = 0.025 * ps + 1.350 \tag{A.2}$$

2. for household dishwashers with rated capacity $ps \leq 9$ and width $\leq 50cm$

$$SPEC = 0.090 * ps + 0.450 \tag{A.3}$$

The corresponding efficiency table is presented below:

Table A.1: *Energy efficiency classes [137]*

Energy efficiency class	Energy Efficiency Index
A (most efficient)	EEI $<$ 32
B	32 $\leq$ EEI $<$ 38
C	38$\leq$ EEI $<$ 44
D	44 $\leq$ EEI $<$ 50
E	50 $\leq$ EEI $<$ 56
F	56 $\leq$ EEI $<$ 62
G (least efficient)	EEI $\leq$ 62

A.0.2 Energy efficiency thresholds

As described in section 3.3.3 for both energy consumption and usage frequency, efficiency thresholds were defined. When rated capacity and energy label are known, maximum energy usage of the standard program can be calculated with the energy efficiency equations given in [140]. The two tables below give an overview on the maximum thresholds for "Normal" sized (width $>$ 50cm) and "Compact" sized (width $\leq$ 50cm) machines, for different rated capacities and energy labels (for the pre-2021 regulation as used in the thesis).

Table A.2: *Efficiency threshold of machines with width > 50cm*

	Label							
Capacity	D	C	B	A	A+	A++	A+++	A+++-10%
10 ps	1.6	1.44	1.28	1.14	1.01	0.9	0.8	0.72
11 ps	1.63	1.46	1.3	1.15	1.02	0.91	0.81	0.73
12 ps	1.65	1.49	1.32	1.17	1.04	0.92	0.83	0.74
13 ps	1.68	1.51	1.34	1.19	1.06	0.94	0.84	0.75
14 ps	1.7	1.53	1.36	1.21	1.07	0.95	0.85	0.77
15 ps	1.73	1.55	1.38	1.22	1.09	0.97	0.86	0.78

Table A.3: *Efficiency threshold of machines with width ≤ 50cm*

	Label							
Capacity	D	C	B	A	A+	A++	A+++	A+++-10%
6 ps	0.99	0.89	0.79	0.7	0.62	0.55	0.5	0.45
7 ps	1.08	0.97	0.86	0.77	0.68	0.6	0.54	0.49
8 ps	1.17	1.05	0.94	0.83	0.74	0.66	0.59	0.53
9 ps	1.26	1.13	1.01	0.89	0.79	0.71	0.63	0.57
10 ps	1.35	1.22	1.08	0.96	0.85	0.76	0.68	0.61
11 ps	1.44	1.3	1.15	1.02	0.91	0.81	0.72	0.65

A.0.3 Dishwasher Energy Labels of new machines (2017)

As mentioned in section 3.3.3, the efficiency threshold characteristics for the energy consumption of a dishwasher were based on normal sized dishwasher (60cm), with 12 or 13 place settings and an A+ machine with eco program. To back this findings up an analysis of 500 available models on Dutch consumer website Consumentenbond was conducted to gain a better understanding of machines on the market.

The analysis was performed for new dishwashers available on the market in the year for which the household plug data was analysed (2017). While this does not represent an overview of penetration of different dishwashers in use in households in the Netherlands, the results show the latest machine sample on the market at the time. The following graphic shows the count of dishwashers per rated capacity(ps) and efficiency label for a total of 500 dishwasher scraped from website of Dutch consumer organisation Consumentenbond [163]. It appears that most new washing machines available (as tested by Consumentenbond) are 12 to 14 ps capacity and mainly A+ and A++ efficiency label. This further confirms the chosen general threshold, which was based on an A+ machine with 12/13 ps, resulting in a standard efficiency threshold of 1.05 kWh.

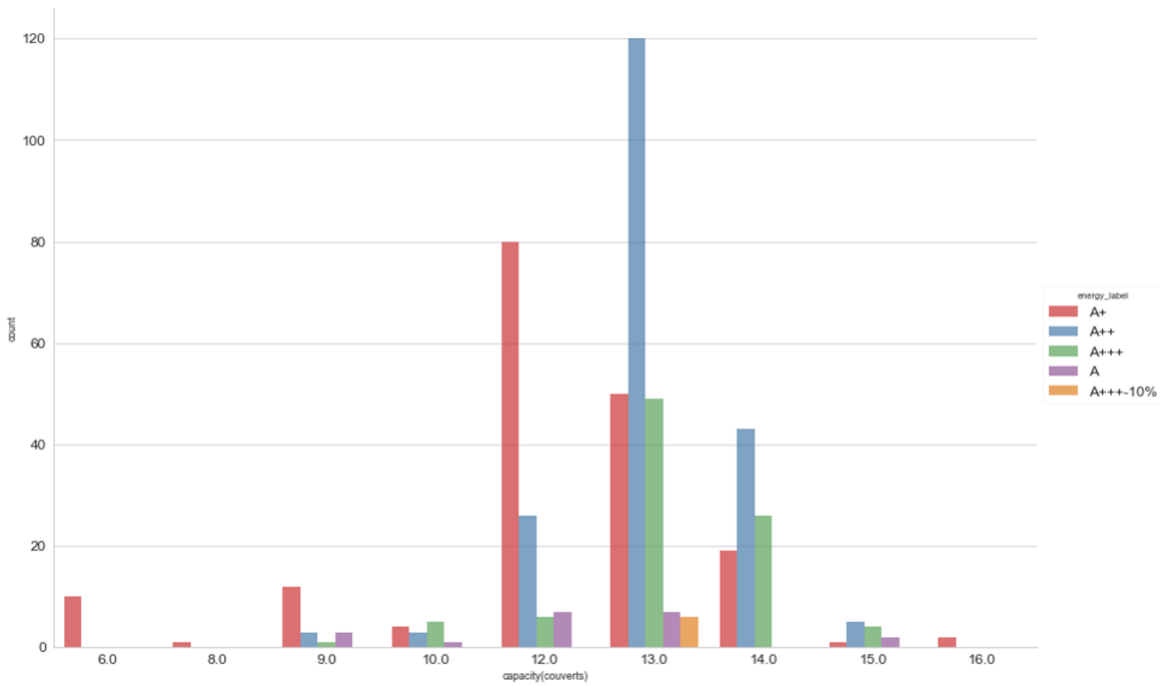

Figure A.1: *Count of new Dutch dishwasher by place setting and energy label (500 dishwasher scraped from Consumentenbond website [163]).*

Appendix B

Algorithms

As described in section 3.2.4 three main algorithm groups of relevance for this research are clustering, regression and classification. The generic overview on these algorithm categories is given in the respective section. The appendix includes further details on the algortihms, encountered problems and application.

B.0.1 Main algorithms

Clustering

A comparison of different methods for clustering was made for the application on three problems encountered with the data handling: Identification of the most likely estimate (MLE) for nested design of the observations, facilitation of pair-matching and noise filtering. Following these problems are further described:

Nested observations
When collecting data of multiple washes for the same household, multiple measurements i are sampled for the same study subject j. This results in a relatively much higher number of observations K than the number of subjects N. These observations could be treated as individual observations, which could be described by the following linear regression [156]:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \mu_{ij} \tag{B.1}$$

However, two main issues arise causing a form of statistical dependence:

- Measurements taken from the same household are not independent
- The number of observations varies per household as the sampling period is the same for all households, but the frequency of usage and thus number of observed washes per time period, varies per household.

In addition problems arise in relation to misdetections and how to threat these. This affects the pair-matching between smart plug and smart meter data and also causes noise.

Pair-matching

Individual observations of washes are not always detected by the smart meter and some washes detected by the smart meter are actually misdetections, therefore it is not always possible to "pair-match" every smart meter detection with a corresponding smart plug detection. When applying an aggregation the number of observations per subject can be reduced to one MLE (set) for the smart meter and one MLE (set) for the smart plug, which then can be pair-matched.

Noise filtering

The detection data can be noisy (especially for smart meter detections) and it could therefore potentially help to filter out misdetections to reduce noise in the analysis.

Clustering methods

To solve the above mentioned problems, different aggregation methods were compared. A simple approach is to aggregate the data to a summary measure (e.g. mean or mode), which can be described by the following equation, where the bar above the variables ($\bar{Y}$, $\bar{X}$, $\bar{\mu}$) represents the aggregated variables [156]:

$$\bar{Y}_i = \beta_0 + \beta_1 \bar{X}_i + \bar{\mu}_i \tag{B.2}$$

However, aggregation to a single summary measure can not represent the potential multi-modal character of detection data, e.g. different washing programs. Furthermore, multidimensional noise filtering cannot easily be performed by this method. A potential way to group the data into multiple multidimensional subsets of similar data points is with the use of a clustering algorithm like K-Means or a Gaussian Mixture Model (GMM).

Both the K-means and GMM method divide the data set into K number of clusters, where each cluster has its own centroid mean value. Each data point is grouped with the cluster centroid to which they have the closest distance, after which the new mean value is updated.

The K-means algorithm makes a hard assignment to which centroid a data point belongs, attributing each data point to a single centroid it is closest to. The GMM algorithm instead assigns a Gaussian distribution in each dimension to every centroid, with not only a mean value, but standard deviation as well. This multivariate Gaussian distribution forms a point cloud to which each data point is assigned with some probability. While the K-means might be slightly easier to interpret and can run much quicker, particularly with high dimensional data, clusters are spherical and in combination with hard assignment can lead to miss grouping. The GMM on the other hand, due to its probabilistic nature, is more flexible in shape and additionally a threshold can be set to discard every point that doesn't belong to either cluster with a minimum probability. The relative low dimensionality of the data, importance to group individual data points correctly and need for noise filtering made the GMM algorithm the clustering algorithm of choice.

The algorithm can be run with multiple numbers of centroid initialisations to see which K number of clusters performs best on the given data set, using the Akaike or Bayesian Information Criterion (AIK or BIC) as performance measure. Furthermore, since the centroids are assigned randomly, there is a risk the algorithm could get stuck in a local minimum. Hence, r random initialisations can be performed. To initialise the GMM algorithm, the K-means can be used as a quick starter. Hence, combining both to improve performance. [63] [157]

Regression

In section 3.2.4, the considered regression analyisis and calculation was described:

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i \tag{B.3}$$

Where Y_i is the dependent variable representing observation i part of N observations, X_i is a corresponding covariate, and β_0 and β_1 are unknown regression coefficients representing the linear line's intercept parameter and the slope coefficients of the covariate, respectively. The covariate X^n and regression coefficients β_n can be extended for multivariate linear regression comprised of n variables. The error, μ_i is the distance of a measured data point Y_i compared to the expected value on the linear line $\hat{Y}_i$. A line that best represents the pattern in the underlying data points has a low overall error. The aim of this method is therefore to find the β parameters for the linear line that minimises the root mean squared error for the measured data points:

$$RMSE = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(\hat{Y} - Y_i)^2} \tag{B.4}$$

Difference between statistics and machine learning
In this research the regression was not used to infer the relationship between measured variable Y and corresponding measured covariates X^n, but was used as supervised learning, to estimate unknown values $\hat{Y}_i$ based on a set of n measured variable(s) X_i^n. Hence, for the application in machine learning the linear regression model is composed of two steps. First the model is trained on beforehand to obtain the underlying relationship, described by the set values for the β parameters that minimises the RMSE.

Secondly, when the β parameters are set, these can be used with the X_i^n values of new observations to estimate the unknown Y_i values of these new observations. [158] [159]

Classification

A machine learning classification algorithm generally is trained based on a data set with observations with known classifications and a set of sub-features and learns how to estimate the classification of unseen data, based on observed sub-features. Several common machine learning classifiers, including Logistic Regression, Naive Bayes, K-Nearest Neighbour and Support Vector Machines (SVM) [160] were considered. However, it was decided not to deploy these.

Classification algorithms are particularly useful when the threshold(s) for the sub-feature (set) is/are unknown, but the class of the training observations is known. In case of the detection data, the classification was unknown and would first have to be labelled. To do this a pre-defined threshold had to be set for the training data, hence a threshold would have to be available already regardless. Therefore, this same threshold was used post regression analysis as a simple binary check, to assess if the regression result is either higher or lower than the set threshold. While this might not pick up on subtle differences in the sub-feature set in the same way a well-trained classifier potentially could, this approach only requires training of the regression algorithm and no further training. Several of the classification algorithms, particularly SVMs need much larger training sets, while the training data set in this research was confined to the relatively small number of households. Quick analysis therefore showed much better results for this simple "regression + threshold" approach, than any of the classification algorithms.

B.0.2 Validation

Error estimation

As explained in section 3.2.4 the NRMSE (Normalized Root Mean Squared Error) served as the comparison metric for the analysis of the error estimation. A selection of algorithms with the lowest NRMSE were further compared based on a visual inspection of their performance. This was done with three different plots:

- **Distribution plot**
  Used to compare the meter detections and several estimation algorithms, to see how well the distribution of meter detections and estimations aligns with the distribution according to the plug detections.
- **Regression plot**
  The meter estimations are plotted against plug detections. The regression plot

shows the many different cross-validated regression lines through these detection points, and the average regression line. The estimation model would estimate the points to be on this line, based on the meter input values. When all meter detections would equal plug detections, they would align on a zero difference line. The regression line and its 95% interval are compared to the zero difference line to give a visual indication of how well the model performs and how the errors are distributed.

- **Difference plot**
  A different way of comparing the errors is the difference plot, where the estimated detections minus the plug detections are plotted against the plug detections. When the mean difference is zero, the average of the meter estimations equals the average of the plug detections. The 95% interval shows the margins of estimation error. [164]

Appendix C

Smart plug analysis

To gain a better understanding of the user group for which the smart plug data was collected, additional data was gathered. This appendix section describes the household and machine characteristics and the energy consumption and usage frequency.

C.0.1 Data Overview

Plug survey and user manuals

A survey was send out to better understand household and machine characteristics of the sample as well as how the dishwashers are used within the household. Part of the survey was to ask for model type numbers. The model number could be used to retrieve the dishwasher's user manual online for further information.

Smart Plug Survey
To gather more information on household and machine characteristics of the smart plug data set, I sent a survey to all households who had participated in the earlier demand side management trial by Eneco. Information gathered from this survey were machine characteristics: efficiency label, size, age, brand and model type number. Images and descriptions were added, to explain where the model number could be found. Furthermore, family size and usage characteristics, such as stated number of weekly washes, number of different programs, program most used and second most used, what type of program and how frequently it is used, were also asked.

Technical specifications such as size and efficiency label (if not answered in the survey) and energy usage for the different programs were manually retrieved from online user manuals. In many cases the energy consumption was provided as a range. If so, the minimum, maximum were both noted.

Of the 100 households 69 responded of which 41 fully answered, including the model number. For 26 households the energy consumption for at least one program was retrievable. In total for 40 washing programs the energy consumption was retrieved, the most frequent used program for 25 households and second most frequent used for 15 households. Reasons for not being able to retrieve information could stem from a wrong type number being provided, the manual not being online available, no washing program that matched the description was found or energy consumption was not provided in the manual.

Appliance sample overview
An overview of the sample of plug users who filled out the survey is presented in Table C.1. For the different categories (brand, machine size, efficiency label, machine age and household size) the respective count of households and relative share are given in the left two columns. The right two columns provide additional information on the ability to retrieve energy consumption data from the user manuals.

As can be seen in the brand section, two brands namely Siemens and Bosch where predominantly found in the households participating in the survey. A total of 32 of the 41 households stated a standard sized dishwasher of 60cm. A place-setting of 12 to 13 couverts appears to be the most common capacity. However, for 9 households it was not possible to retrieve this data. The most common energy label is the A++ label. The efficiency label of 7 dishwasher stays unknown.

Over 60% of the dishwashers analysed in the survey were less than 5 years old. Of the 63% of dishwashers for which the energy consumption was retrievable the majority were relatively young dishwashers. For dishwashers above 5 years the number of retrievable energy consumption information decreases significantly as either user manual or energy reporting in the user manual was less often available online. If information was not retrievable, the washing program energy consumption could not be compared. However, other specifications stated in the survey could still be considered in further assessment.

Table C.1: *Overview of type of dishwashers and households in survey*

		Total in survey		Energy retrievable	
		count [a]	**share**	**count** [b]	**share**
Brand	Siemens	16	39%	9	22%
	Bosch	9	22%	6	15%
	Miele	3	7%	3	7%
	AEG	2	5%	1	2%
	Bauknecht	2	5%	2	5%
	IKEA	2	5%	0	0%
	Pelgrim	2	5%	2	5%
	Boretti	1	2%	1	2%
	Indesit	1	2%	0	0%
	Neff	1	2%	0	0%
	Samsung	1	2%	1	2%
	Whirlpool	1	2%	1	2%
Size	60 cm - 12 couverts	11	27%	10	24%
	60 cm - 13 couverts	12	29%	12	29%
	60 cm - 14 couverts	3	7%	3	7%
	60 cm - 15 couverts	1	2%	1	2%
	60 cm - unknown	5	12%	0	0%
	Unknown	9	22%	0	0%
Label	A+++	4	10%	3	7%
	A++	15	37%	12	29%
	A+	8	20%	8	20%
	A	7	17%	3	7%
	Unknown	7	17%	0	0%
Age	0 to 2 years	13	32%	12	29%
	2 to 5 years	13	32%	8	20%
	5 to 10 years	7	17%	3	7%
	10 to 15 years	7	17%	3	7%
	15+ years	1	2%	0	0%
Household size	One	1	2%	1	2%
	Two	11	27%	7	17%
	Three	10	24%	3	7%
	Four	13	32%	10	24%
	Five	6	15%	5	12%
Total		41	100%	26	63%

[a] Number of households in survey with respective dishwasher [b] Number of households in survey for which respective dishwasher manual information was found on energy consumption of the stated program

Usage summary

From the smart plug data a number of different consumption statistics can be generated. The detected washes can be aggregated in several different ways. Either as all washes individual, regardless of household and dishwasher they belong to. As average per household, for the 100 households considered in the sample or based on clusters for most likely washes per household. Table C.2 shows the summary statistics for energy based on these 3 methods.

Table C.2: *Summary statistics: Energy consumption in kWh per wash under different aggregation methods for all smart plug users*

	Energy consumption [a]					
	count [b]	**mean**	**std**	**min**	**median**	**max**
All detected washes	5691	1.211	0.27	0.44	1.18	2.29
Averaged	100	1.215	0.20	0.86	1.18	1.69
Clusters	171	1.215	0.24	0.76	1.17	1.95

[a] Energy consumption in kWh per wash [b] Number of detected washes. Individual washes, average of washes per household and clusters of comparable washes per household consecutively

The mean energy consumption of all three aggregation methods is almost similar at 1.21-1.22 kWh. However, the clustering provides the opportunity to distinguish different washing programs.

A consumption overview including weekly washes, number of different programs and the usage frequency and energy consumption per program is provided in table C.3.
On average the 100 households presented in table C.3 use their dishwasher 5.2 times per week. According to the clustering algorithm people use 1.7 programs on average, as 59 households use a second program and 13 households even use a third. The frequency of weekly usage of different programs, the utilisation, is reported in conditional share. Thus, the frequency of usage if used. Of all households the algorithm detects that program used as the primary program is used 67% of the time. If more than one program is detected by the algorithm, the conditional share of the second most frequently used program is 41%. In addition 13 households are detected to be using a third program with a conditional share of 28%.

The unconditional share can be calculated by multiplying the number of household that use the respective program, with the frequency of usage, divided by all 100 households. The unconditional share for cluster program 1 hence is 67%, for program cluster 2 it is 24% and for program cluster 3 is 3.5%. This does not fully add up to a 100%

Table C.3: *Summary statistics for dishwasher: Weekly utilisation of dishwasher and energy consumption per wash kWh for all smart plug users*

	Consumption overview					
	count [a]	**mean**	**std**	**min**	**median**	**max**
Number of weekly washes	100	5.2	2.1	1.0	5.6	11.3
Number of different programs	100	1.7	0.7	1.0	2.0	3.0
Utilisation program 1 [b] [c]	100	67%	24%	33%	53%	100%
Utilisation program 2 (if used)	59	41%	8%	21%	41%	50%
Utilisation program 3 (if used)	13	28%	6%	20%	32%	33%
Energy usage program 1 [d]	100	1.23	0.25	0.83	1.17	1.95
Energy usage program 2	59	1.18	0.23	0.76	1.16	1.80
Energy usage program 3	13	1.30	0.21	1.05	1.29	1.73

[a] Number of households [b] Program 1 refers to program cluster most often used within a household, program 2 and 3 refer to the second and third most often used program respectively. Detected program clusters are based on GMM estimation; actual values might be slightly different. [c] The mean value shows what share of washes the program is used, if used; if someone does not use a program 3, it is not counted, instead of counted as zero [d] Energy consumption in kWh per wash

as no more than 3 program clusters were considered per household and some washes (5.5%) are not categorised.

The mean energy consumption of program cluster 2 appears to require 0.05 kWh less energy per wash than program cluster 1 and 0.12 kWh less energy per wash than program cluster 3. These findings could lead to the assumption that the most used program is not the most energy efficient program, less energy intense programs are used as either the first or second most often used program, while more energy intense programs are used less often and fall into the program cluster 3.

C.0.2 Consumption analysis

The plug data and the information retrieved from the plug survey were compared into how real-life usage compares to stated information.

Frequency of usage

As shown in the last section, the smart plug detection data reveals information on program utilisation, energy consumption and frequency of washes. This data is combined with data from the survey to compare stated versus revealed usage. To gain an understanding of the usage frequency of the dishwasher and whether this matches with people's perceived usage frequency, a comparison of these two was made. The number of weekly washes detected by the plugs was directly compared to the weekly washes stated in the survey. The survey results for the 41 households that answered the survey were matched with their respective smart plug data, as shown in table C.3.

Table C.4: *Summary statistics for dishwasher: Weekly utilisation of dishwasher for all smart plug users who answered consumer survey*

	Stated in survey			**Detected by plugs**		
	count [a]	**mean**	**std**	**count**	**mean**	**std**
Number of weekly washes	41	5.8	2.2	41	5.1	2.0
Number of different programs	41	1.8	0.6	41	1.7	0.6
Utilisation program 1 [b]	41	88%	15%	41	68%	24%
Utilisation program 2 (if used)	28	27%	16%	24	42%	7%

[a] Number of households [b] Program 1 refers to program cluster most often used within a household, program 2 and 3 refer to the second and third most often used program respectively. Detected program clusters are based on GMM estimation; actual values might be slightly different. The mean value shows what share of washes the program is used, if used; if someone does not use a program 3, it is not counted, instead of counted as zero

Comparing the average number of weekly washes stated in the survey (5.8) with the number detected by the smart plugs (5.1), it can be seen that these numbers diverge by 0.7 washes per week. On average the survey participants seem to slightly overstate their washing frequency.

The survey participants stated a utilisation of 88% on average for the primary program, basically saying 9 in 10 washes are using the same program, while the detection algorithm detects 68% or about 7 in 10. Not all 28 stated second programs were detected by the plugs, which detected a second program for 24 of the 41 households. Furthermore, the conditional share deviates by 15%. This seems to indicate that not only do households overestimate how often they wash, but also how often they use the exact same program. Apparently there might be some variation, potentially stemming also from different users within the household.

Energy consumption per wash

The energy consumption detected by the plugs can be compared to the energy usage according to the user manual. Table C.5 summarises the average energy consumption split up into the primary used program cluster 1 and secondary program cluster 2. Initially these were matched on utilisation, but it was found that households not always really use the program the most that they state they use the most. Therefore, in table C.5 for every household the programs were not matched on utilisation, but on energy instead.

Table C.5: *Summary statistics for dishwasher: Energy usage per wash according to user manual matched by closest proximity with energy per wash detected by smart plug*

	Program 1			Program 2		
Energy consumption [a]	**count** [b]	**mean**	**std**	**count** [c]	**mean**	**std**
Smart plug detection	22	1.20	0.22	12	1.15	0.17
User manual stated average	22	1.21	0.30	12	1.09	0.19
User manual stated min	22	1.13	0.30	12	1.04	0.19
User manual stated max	22	1.29	0.35	12	1.13	0.24

The average energy usage of program 1 in table C.4 is 1.20kWh according to the plugs and 1.21 kWh stated on average in the user manual. On average the energy usage stays under the maximum threshold. For program 2, there is slightly more deviation. At 1.15kWh the plug detected energy consumption falls 0.06kWh above the average and 0.02kWh above the average upper limit of 1.13 kWh. Closer inspection however showed that this is caused by one single outlier with a relatively high energy consumption. Potentially the user might have given a wrong description of the program, hence resulting in a mismatch. The plug detected energy consumption for all others programs falls nicely within the spectrum of the user manual. Hence, for the investigated sample the energy usage in real life as measured with the smart plugs on average does not exceed the energy consumption according to the user manual, when considering the specific program used. This indicates that, at least for the investigated sample, dishwashers in real-life wash efficiently in accordance with the stated washing program.

Appendix D

Model results

This section in the appendix focuses on the results that stem from the detailed analysis and testing of the different potential features and algorithm for the development of the energy and usage frequency estimations. The most relevant results have been lined out in the main thesis in section 4.2.

D.0.1 Energy per wash estimation

Features

This section presents the statistics for different features, where detections were first averaged per household. Table D.1 shows this for the smart plug detected consumption variables. Table D.2, presents the same parameter statistics, gathered by the smart meter detection model.

As described in chapter 4.1 for the energy consumption, a larger standard deviation is observed with all plug detections than in the case of per household averaged detections. This is also the case for the individual variables as the more extreme observations are averaged out.

The summary statistics for the smart meter detections do not include a row with energy usage per wash, as this is not detected on the smart meter. The share of water heating is shown, however, as a would be percentage, comparing the detected water heating with the energy usage as detected by the plugs.

Table D.1: *Summary statistics of potential energy consumption variables for washes detected by the smart plugs averaged per household*

	Plug detected consumption variables					
Averaged	**count**	**mean**	**std**	**min**	**median**	**max**
Energy usage of total wash [kWh]	100	1.22	0.21	0.86	1.18	1.69
Energy usage of water heating [kWh]	100	1.01	0.20	0.65	0.97	1.54
Share of energy for water heating	100	82%	5%	66%	83%	93%
Duration of total wash [min]	100	85	23	29	85	132
Duration of water heating [min]	100	31	6	20	29	52
Heating power [kW]	100	2.0	0.1	1.6	2.0	2.2
Number of heating moments	100	2.7	0.6	2	2.5	4.1

Table D.2: *Summary statistic of potential energy consumption variables for the smart meter detection averaged for households*

	Meter detected consumption variables					
Averaged	**count**	**mean**	**std**	**min**	**median**	**max**
Energy usage of water heating [kWh]	100	0.94	0.14	0.65	0.92	1.51
Share of energy for water heating	100	78%	8%	59%	78%	100%
Duration of total wash [min]	100	87	15	39	87	128
Duration of water heating [min]	100	28	4	20	28	48
Heating power [kW]	100	2.0	0.1	1.7	2.0	2.2
Number of heating moments	100	2.5	0.4	2.0	2.5	3.8

Estimation algorithm

As described in section 4.2.1 the average and mode were calculated and tested for different regression approaches. The NRSME results for the different estimations models can be seen in table D.3.

Table D.3: *NRMSE [%] for different energy estimation methods for average and mode energy consumption per dishwasher wash*

	Average		**Mode**	
	mean	std	mean	std
Heuristic calculation				
Mean heating energy % of total energy	9.6	0.7	19	2.0
Heating energy + mean offset *(Predict 1)*	9.2	0.6	19.9	3.1
Regression model with				
Number of heating moments	14.4	1.0	17.9	1.5
Total heating duration	10.1	0.7	16	1.8
Heating energy	9.4	0.6	14.4	1.4
All regressors *(Predict 2)*	9.1	0.7	14.9	1.7
Heating energy and moments *(Predict 3)*	8.8	0.7	14.4	1.5

NRMSE mean and standard deviation over 1000 iterations of 50/50 split cross validation on plug data and given in percentage as normalised versus the mean plug detected value. In the rows different energy estimation methods are given, in the columns the used aggregation methods Average and Mode respectively. The numbered energy estimation methods correspond to Predict 1, 2 and 3 for Average aggregation.

By comparison, when regressing energy with the heating energy as detected by the plugs, this results in a NRMSE of 5.7%. Adding the number of heating moments as additional information, results in an NRMSE of 5.4%, or RMSE = 0.066 kWh per wash. This shows that the best performing meter estimation (Predict 3: 8.8% NRMSE) performs nearly as well.

Clustering algorithm (GMM)

The section below describes the development and usage of the GMM clustering algorithm. This algorithm was dveloped and tested and initially showed superior results. However, as the results from the NILM detection improved over time, the noise canceling function of the GMM algorihm became less relevant. Over time simply using an average value performed better. However, since the GMM algorithm does provide multiple beneifts, including recognition of multiple washing programs and how commonly they are relatively used, considerations during the development of the GMM algorithm are outlined below.

As described in section 4.2.1, the performance of the different estimation methods was assessed and the best perfoming method (Predict 3) was further analysed. A Gaussian Mixture Model (GMM) clustering algorithm as described in section B.0.1 was used to investigate the estimation performance of the clustering method.

The GMM clustering algorithm was based on total energy consumption per wash and the subfeatures duration of the water heating, number of heating moments and total duration for each individual wash.

The GMM method allows for the detection of different washing programs as used in C. In addition it filters out the noise of misdetections, which is especially relevant for detection algorithm with a lot of noise detection. In C existing smart plug data was clustered by the GMM method. In 4.2 smart meter estimation data is clustered by the algorithm.

The energy consumption related to a cluster is the energy of the program. The resulting number of different clusters from the GMM can be interpreted as the number of different commonly used washing programs per household. The frequency of occurrence of each cluster can be translated into the frequency of usage of the washing program. The plug cluster and meter cluster could be pair-matched in three different ways. The simplest is to only select the cluster with the highest likelyhood. This would represent the most common used washing program only. Alternatively each cluster can be matched with the cluster with the closest likelyhood or a weighted average can be taken. Since it can happen that the meter and plug data don't result in the same number of clusters and clusters would have to be dropped out, only the most likely cluster and the weighted average were tested.

Because the GMM clustering algorithm randomly initialises it can have slightly different outcomes every time it is applied. Because of the random initialisation a single run outcome can not be solely relied on. To make the outcomes more stable the algorithm was run 30 times. Around that number the algorithm converged to a stable mean. A 2nd stage GMM was additionally considered where a GMM clustering algorithm was ran over the outcomes of the 30 run GMM. This has as advantage that for every household a specific overview is created. The disadvantage is that the second stage GMM again is subject to the effect of random initialisation. Although this effect should be much lower because the variance between clusters of detections coming out of the 30 run GMM should be again lower than the differentiation between meter detections.

In table D.4 the NRSME the GMM single run, 30 runs, and second stage GMM is depicted. The outcomes of these three different GMM clustering approaches were aggregated on the (weighted) average and the mode (highest confidence).

A key feature of the GMM method is that it shows the probability, its confidence, in the existence of a cluster. Based on the amount of total data points, in this case the number of all washes per household that are grouped together in a cluster. A cluster supposedly is constituting one washing program. The weighted average takes into consideration the weight of each cluster i.e. how often a washing program is used. The highest confidence only selects the cluster (washing program) with the highest probability, i.e. the washing program that is used the most.

Table D.4: *NRMSE [%] of energy estimation of energy per dishwasher wash using Predict 3 with different GMM aggregation methods*

	Weighted average		**Highest confidence**	
	mean	std	mean	std
GMM runs				
Single run GMM	9.1	0.7	14.6	1.3
30 runs GMM	9.5	0.6	15.5	1.1
2nd stage GMM	9.6	0.7	15.3	1.1

NRMSE mean and standard deviation over 1000 iterations of 50/50 split cross validation on plug data and given in percentage as normalised versus the mean plug detected value. For all estimations the regression with heating energy and moments (3) is used. In the rows different runs of the GMM algorithm are given, in the columns the two GMM aggregation methods are given that relate to mean and mode respectively..

As can be seen in table D.4 the average error of estimation for the weighted average of all washes is lower than the estimation error for the program with the highest confidence. Thus, the algorithm performs better to estimate the energy consumption of all washes together than to estimate the energy consumption of a specific washing program, even when this is the washing program that supposedly is used the most.

Based on the quality of the smart meter estimation variables the GMM appears not able to find a specific cluster that matches up as precisely with the energy usage of the most frequent used program of the smart-plugs, as it is able to find the weighted average.

Due to random initialisation the single run GMM seems to perform better than the average of multiple runs. The deviation as a result of random initialisation only decreased from 0.6 to 0.7% thus the random selection of the data in cross validation causes more deviation than the random initialisation. However, the weighted average of GMM clustering with (9.1 mean, 0.7 std) performs worse than simply taking the average (mean 8.8, std 0.7) see table D.3.

D.0.2 Frequency per week estimation

Estimation algorithm

In addition to the energy per wash estimation, the frequency per week estimation was assessed in section 4.2.2. To assess the accuracy of the different estimation methods the normalised root mean squared error (NRMSE) can be compared for several estimation methods as presented in table D.5. The different estimation algorithms were run for a 1000 iterations of 50/50 split cross validation, resulting in a mean NRMSE.

The estimation algorithms were run for several aggregation methods. Either looking at all detections separately, looking at the average weekly washes per household over the 12 week period and the most common (mode) of weekly washes per household. Furthermore it was tested whether it is better to first regress and then aggregate or first aggregate and then regress.

For a suitable estimation method ideally a low estimation error and a a low standard deviation occurs. Predict 3 with average of regression showed to perform best, with a NRMSE of 27.2% had the lowest estimation error.

Table D.5: *NRMSE [%] for different combinations of weekly dishwasher detection methods*

	Meter		**Predict 1**		**Predict 2**		**Predict 3**	
	mean	std	mean	std	mean	std	mean	std
All detections	43.9	1.0	35.1	1.0	40.3	1.1	39.2	1.3
Average of regression	36.5	2.5	27.6	2.3	28.7	2.8	27.2	2.7
Regression on average			27.5	2.6	28.4	2.8	28.3	2.8
Mode of regression	42.3	3.3	34.1	3.2	37.8	3.2	36.9	3.1
Regression on mode			35.2	3.2	38.4	3.3	37.8	3.4

NRMSE mean and standard deviation over 1000 iterations of 50/50 split cross validation on plug data and given in percentage as normalised versus the mean plug detected value. In the rows different aggregation methods are given, in the columns the different estimation methods.

Appendix E

Smart meter analysis

This section in the appendix is comprised of additional findings and backgroud analysis that has been done to complement the smart meter methodology and results.

E.0.1 Additional Smart Meter Analysis

Figure E.1 shows two ways of calculating the weekly average. One includes all households, hence also those with no detections, thus who are assumed not to be home that week. The second only includes the households where at least a single wash was detected that week, to ensure only households where family members are home are included. This was done to differentiate between the effect of reduction in washes due to less washing, versus more households on holiday.

As has been pointed out in section 3.4.2 and can be seen in figure E.1 in week 15 a large change happened. This happened as a result of a change in the NILM detection algorithm by K. Basu. A major changes was made to the sensitivity of the detection algorithm reducing the the amount of misdetected washes, but increasing the amount of missed washes. This resulted in an average reduction of 0.74 weekly washes per week per household. As it was computationally to expensive to rerun the the full year of data an adjustment has been made to align for the change. The time line therefore was adjusted to match the time period before and after this change in week 15.

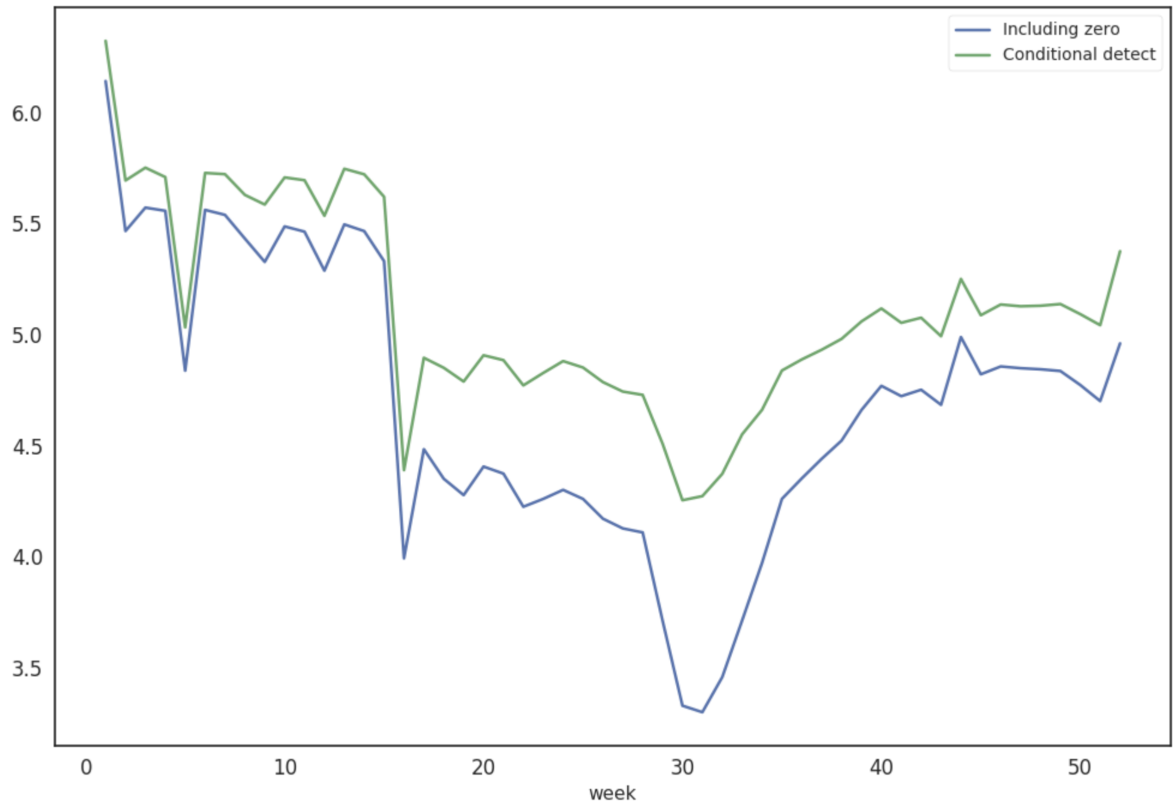

Figure E.1: *Estimated average weekly usage frequency in washes over the year including households with zero detections or conditonal that at least one wash is detected*

E.0.2 Holiday Netherlands 2018

In figure 4.13 the fluctuations in the average washes per week per household over the year were related to public holidays and other events. The table below shows the comparison data for Dutch holidays in 2018, which was used in the graphic. [165]

Table E.1: *Holidays in the Netherlands in 2018 [165]*

Schoolvakanties 2018	Regio	Van	Tot	Week
Kerstvakantie 2017	Heel Nederland	23.12.2017	07.01.2018	52 t/m 1
Voorjaarsvakantie 2018	Noord en Midden	24.02.2018	04.01.1900	9
Voorjaarsvakantie 2018	Zuid	17.02.2018	25.02.2018	8
Meivakantie 2018	Heel Nederland	28.04.2018	06.01.1900	18
Zomervakantie 2018	Noord	21.07.2018	02.09.2018	30 t/m 35
Zomervakantie 2018	Midden	14.07.2018	26.08.2018	29 t/m 34
Zomervakantie 2018	Zuid	07.07.2018	19.08.2018	28 t/m 33
Bouwvak 2018	Noord	06.08.2018	24.08.2018	32 t/m 34
Bouwvak 2018	Midden	30.07.2018	17.08.2018	31 t/m 33
Bouwvak 2018	Zuid	23.07.2018	10.08.2018	30 t/m 32
Herfstvakantie 2018	Noord en Midden	20.10.2018	28.10.2018	43
Herfstvakantie 2018	Zuid	13.10.2018	21.10.2018	42
Kerstvakantie 2018	Heel Nederland	22.12.2018	06.01.2019	52 t/m 1

E.0.3 Efficiency analysis

In section 4.3.4 the distribution of estimated energy consumption per wash by efficiency label and washing temperature was presented, excluding the full length of the 1st and 4th quartiles (whiskers). The full boxplots including the whiskers are shown in the following figures for the estimated energy per wash (see figure E.2) and per number of weekly washes E.3.

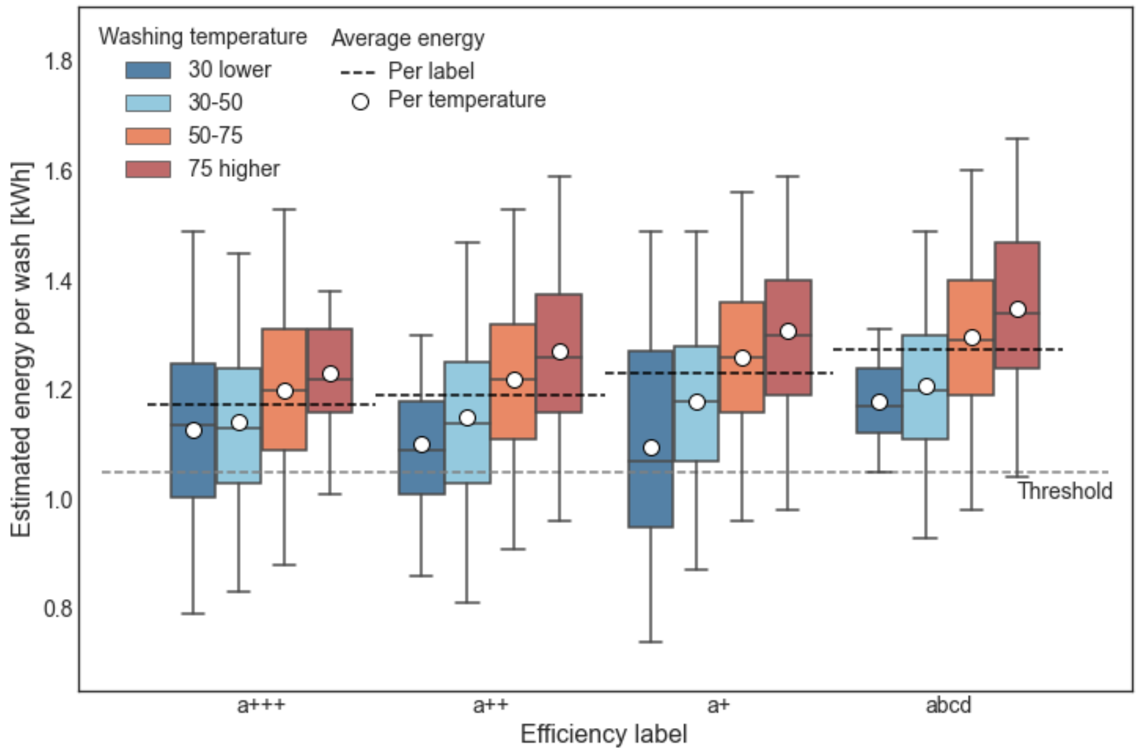

Figure E.2: *Distribution (boxplot) of estimated energy consumption per wash for efficiency label and by washing temperature*

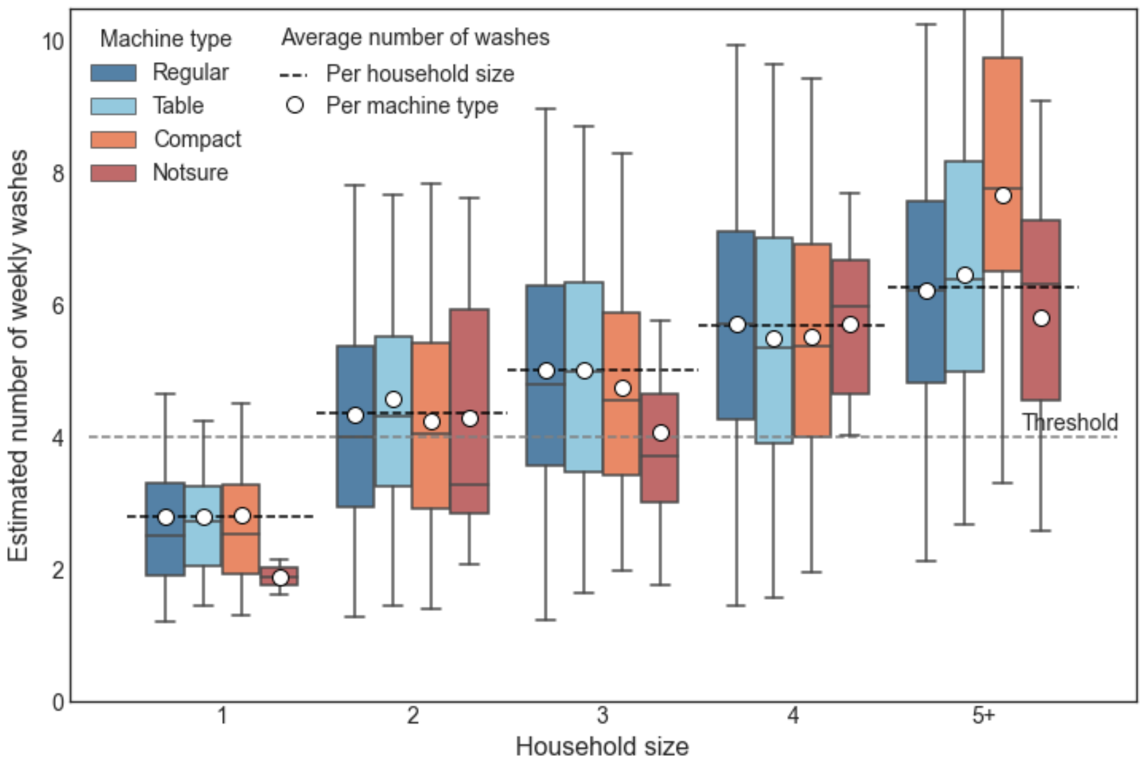

Figure E.3: *Distribution (boxplot) of estimated number of weekly washes for different household sizes and by dishwasher machine type*