Technical and Socio-economic Evaluation of Energy Disaggregation Using Non-Intrusive Load Monitoring (NILM) for the Characterization of Residential Appliances' Profiles: the Netherlands Case Study.

Federica Blasioli
federica.blasioli@tecnico.ulisboa.pt
Instituto Superior Técnico, Universidade de Lisboa, Portugal

January 2021

Abstract* — In a context where 97% of the European buildings are energy inefficient, numerous countries are investing on smart meters rollouts, and a severe intensification of the electricity demand from the residential sector is expected, energy disaggregation could serve as a valuable resource for leveraging advanced meter infrastructures and creating opportunities for innovative services. In this thesis, the Non-Intrusive Load Monitoring (NILM) technology has been evaluated for the residential use-case. The Netherlands was set as reference scenario for the two energy disaggregation models aimed at the recognition of household's appliances fingerprints. An event-based approach was designed using a cost-effective system that allowed the identification of the fridge signature and its duty cycle, bringing the potential to alert consumers about misfunctioning and inefficient devices. Still, compared to the event-less approach using unsupervised learning algorithms, the proposed model showed less accurate results and appliances classification. Currently, the concerns for data security and the existing technological limits are preventing the NILM technology to be deployed with high range of accuracy on a larger scale. Finally, despite a promising market potential, NILM technology was evaluated still too expensive for direct investments of residential consumers. With utilities being the ones deploying NILM solutions in the households, a broader implementation could be possible instead. Additionally, a detailed appliance-specific breakdown together with a customized feedback system, could help energy consumers to save up to 5% of their energy consumption and increase their awareness on energy conservation.

***Keywords* — Energy Disaggregation; Non-Intrusive Load Monitoring; Appliances fingerprint recognition; NILM Economic Impact; Energy feedbacks; Residential energy savings.**

I. INTRODUCTION

In the EU, buildings are the largest energy consumers accounting for 40% of the total energy consumption [1]. Recent studies indicate that 97% of the European building stock is currently inefficient while at the same time 75-90% of those buildings will remain in use until 2050 [2]. In this context, aggravated by an expected severe intensification of the electricity demand of the residential sector, especially for cooling, appliances and plug loads such as electric vehicles, the implementation of an algorithm that can disaggregate the specific electric loads in a household, being able to provide real-time insights to the tenants and detailed tailored information about the specific appliance's usage, could contribute not only to a better energy management and smart meter leveraging but also to foster the consumers' engagement in the cause. Inspired by numerous progress in the field of artificial intelligence, machine learning, networks and data communication, research efforts in this area have lately been focusing on the concept of *Non-Intrusive Load Monitoring* (NILM). This method consists of performing a breakdown of the total energy consumption data, measured at house level by adding just one sensor measurement to the smart meter. In fact, every appliance is characterized by a signature, representing its specific consumption profile that can be detected from the overall electricity consumption of a house in a process called *disaggregation*. If broadly implemented, NILM could serve as a valuable resource for understanding the variability of energy consumption and creating an opportunity for innovative services to increase awareness on the issue of excess energy consumption, finally both reducing the large impact of energy utilization in buildings and expanding grid flexibility capacity.

This thesis aims to provide a comprehensive understanding of the NILM technology from a technical, economic and social point of view.

A. Literature Review: NILM Process

The concept of NILM was firstly introduced in 1985 by George W. Hart from the Massachusetts Institute of Technology [3]. The basic assumption behind his idea is that every change in the total electrical load of a household is the consequence of an electric device changing its state. In almost four decades since NILM was firstly introduced, a large amount of literature has been produced over the topic, proposing different processes to obtain disaggregation through the NILM methodology.

The overall NILM process is constituted of five main modules: data acquisition, event detection, feature extraction, load identification and energy estimation. However, it is important to clarify the difference between event-based and event-less approaches for energy disaggregation. While the event-based approach requires the detection and the labelling of every appliance transition from the aggregated load, the event-less approach, does not rely on event detection and tries to match each sample of the total consumption with the consumption of one or different combinations of specific devices [4]. The latter approach applies statistical and probabilistic machine-learning methods without the necessity of any labelled information. Thus, the whole process is more straightforward, but at the same time more computationally intensive [5]. On the other hand, the event-based approach requires less computational power because the feature extraction is only performed when an event is identified and not on the whole dataset, which also leads to a higher possibility of error propagation. Overall, each module of the process is needed for the event-based approaches while for the event-less approaches two modules are not included in the process: the event detection and the feature extraction.

II. Energy disaggregation modelling

A technical analysis on NILM is performed using two different energy disaggregation models. The first one includes the execution of the NILMTK for REDD and DRED datasets to understand the performances of two of the most common event-less algorithms already publicly implemented within this toolkit. The second model aims at the development of an event-based algorithm able to identify the signatures of common appliances in a Dutch household, by analysing the differences in power in the total electricity consumption.

A. Methodology

Publicly available on GitHub, NILM-toolkit (NILMTK) was used as a tool to understand the quality of performance of two common NILM algorithms: Combinatorial Optimization (CO) and Factorial Hidden Markov Models (FHMM). The analysis was first performed using the standard REDD dataset, with which the toolkit was designed, and then executed with the Dutch DRED dataset in order to tailor it to the target area of this thesis. In a second phase, a system to collect the data from multiple households, already provided with smart meters, was developed and the data was visualized in Grafana and saved in SQL databases. Additionally, an algorithm was developed and implemented to perform energy disaggregation on the collected Dutch household's data, without having any data about the loads of individual appliances. The home conceived code is also available on GitHub and it was written using the programming language Python [6].

B. NILMTK

The NILMTK is an open-source software available online on GitHub with detailed documentation that guides both the users and the developers through each step of the disaggregation process. It was designed to enable the analysis of different datasets and algorithms and to create a uniform and simple interface to perform those algorithms and implement others.

1) Data Acquisition: The first analysed dataset was REDD which contains data for six different homes in the USA in 2011. Despite the fact that the dataset only provides several weeks of power data at the frequency of 1 Hz, it includes both active power at the meter-device level and apparent power for the whole-house measurements. The disaggregation was only performed on one house (building 1) which is composed of two site meters describing the total electricity consumption per circuit, and eighteen monitored devices.

The second dataset, the DRED, is constituted of five months data of a single household in the Netherlands in 2015. The sampling frequency of the data is 1 Hz, the apparent power is measured both at the appliance level and at the site meter, and there is no measurement of active power. It is composed of 12 monitored devices and one site meter for the overall consumption. Despite the longer range of measurements, the structure of the data is simpler than that of the REDD dataset.

2) Load Identification: In this study, the two most common algorithms for energy disaggregation in NILMTK were selected: CO and FHMM. Those both belong to the supervised learning category in which training with previously labelled data is necessary to perform disaggregation. The main reason why the authors decided to include these common and well-known algorithms is to provide an established approach to enable a uniform comparison for novel algorithms' results [7].

The Combinatorial Optimization (CO) algorithm tries to estimate the state of each appliance, minimizing the difference between the household total power demand measured at a specific time and the summatory of all the appliances' estimated power demand at the same time. The CO finds the best combination of these appliances' states selecting the ones closer to the aggregate signal. Before that though, every time slice and consequently every state is assumed to be an independent cluster, increasing the possibility of error propagation [7].

The Hidden Markov Model (HMM) and its variations are the most commonly used methods for NILM algorithms. The approach is based on the Markov Model according to which a number of finite states of a process can be described by a final state, or output. If this is applied to energy disaggregation, the output would be the actual aggregated power measured at time *t*. Since the total consumption in every instant is described as the change of states of many different devices in the household and not only one, a combined model of many HMM is necessary to perform NILM. The Factorial Hidden Markov Model (FHMM) is composed of k-HMMs, where k is the number of appliances that evolve simultaneously and independently over time. As the number of introduced parameters increases, FHMM is more complex than CO because of the additional state transition and temporal information that were missing in the previous model [8][9].

3) Results and Performance Metrics: The final part of the NILM process involves analysing the obtained results, evaluating the performances of the algorithms using different metrics and getting an estimation of the energy consumed by the appliances. Due to the high computational capacity of the selected algorithms, the disaggregation was only performed for the five most consuming appliances (top-5) of each dataset.

Besides, particular attention was given to the training phase. For the REDD dataset, which is composed of only a few weeks of data, the NILMTK proposed to train just 32% of the entire dataset. At first, the same percentage was also applied for training the DRED dataset, however it appeared to perform better when 42% of the data was set as training, and thus this percentage was used for the evaluation phase.

In *Figure 1* the disaggregation of the fridge, being the most consuming appliance in the DRED dataset is shown for one typical day of September. The upper graph represents the ground truth (GT), which is the actual consumption profile of the fridge. The other two graphs are respectively the results for the CO and FHMM algorithms. As it can be observed, the FHMM performs better than the former one, being able to recognize, despite some noises, all the events. The CO instead, clustering different groups of appliances' states, results in more disturbance on the lower and most frequent powers, not returning a very sharp fridge profile.

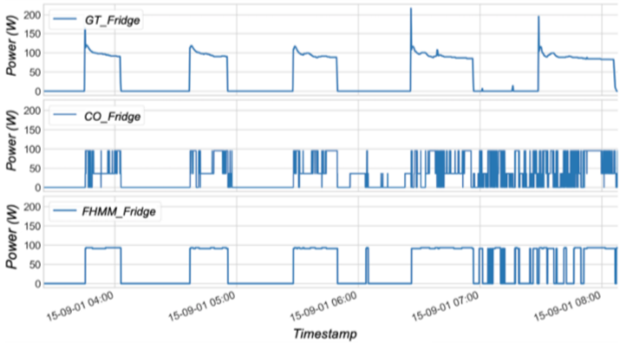

Figure 1 DRED. Fridge disaggregation

The evaluation of results is a very challenging phase of NILM because there is no uniform way to measure and compare the accuracy of the disaggregation algorithms. The most common approach is to use the root mean square error (RMSE). This metric represents how convergent the predicted values are to the actual measurements. The lower this error, the more accurate the estimated values will be. *Figure 2* shows the RMSE obtained for the DRED and REDD datasets performing the CO and FHMM disaggregation algorithms for the five most consuming appliances. It is worth noticing that the major devices in the two datasets are not the same, except for the fridge and the microwave.

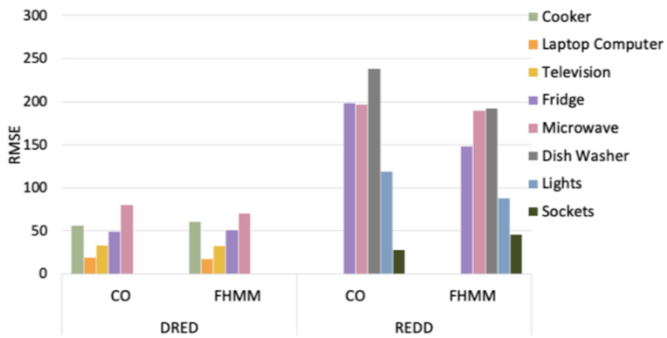

Figure 2 RMSE for the energy disaggregation of the top-5 appliances of the DRED and REDD datasets

Overall, the REDD dataset performed worse than the Dutch dataset for both the algorithms, likely caused by the shorter time period of the dataset. Moreover, it is interesting to notice that the most intermittent appliances such as television, computer and sockets received the lowest errors. This is due to their irregular but recurrent usage which allowed the algorithms to averagely detect their mean consumption. Finally, it is important to clarify that the RMSE is not a normalized measurement. Thus, the comparison between the results for different appliances is influenced by this factor.

In *Table 1* two other evaluation metrics calculated for the two datasets are presented. The F_{score} is a number that ranges from 0 up to 1, which indicates the perfect accuracy of state detection. The F_{score} metric is used to indicate how good the algorithm is in classifying the operating states of an appliance, but it does not provide any information about the power consumption of the appliance [10]. The fraction of energy assigned correctly is computed as the percentage of the predicted energy that was accurately recognized over the total actual consumption.

Table 1 F_{score} and fraction of energy assigned correctly metrics results for DRED and REDD datasets

	DRED		REDD	
	CO	**FHMM**	**CO**	**FHMM**
F_{score}	0,172	0,250	0,275	0,326
Fraction of energy assigned correctly (%)	66,1	70,0	71,5	88,6

The evaluations obtained from these two metrics alter the RMSE results in *Figure 2*. The REDD dataset, which received the worst convergence of estimated values around the true values, is the one with the highest percentage of energy assigned correctly and with the highest F_{score}. This means that even though the predictions of the DRED dataset are more concentrated around the true values, less true positive events are actually recognized, as pointed out by the low F_{score}. On the opposite, the REDD dataset correctly recognized a high number of events, including a higher number of true positives (higher F_{score}). Finally, the FHMM resulted to be the algorithm with the better performances for all the metrics considered in this chapter. The results for CO are likely worsened by the lower power consumption appliances which are most prone to be mismatched during the optimization process. Besides, the FHMM algorithm includes temporal information of the state transitions which are not considered for the CO algorithm.

C. Signature Recognition Approach

This event-based approach is intended to perform disaggregation with low-frequency data (0,1 Hz) in order to recognize the basic appliances in a household. The signature recognition approach consists of defining the main characteristics for each of those devices' consumption profiles and evaluating their changes over time. For this method, experimental measurements were carried out in several buildings in the Netherlands.

1) System Design and Data Acquisition: The hardware necessary for this system is very simple and requires low investments. It consists of a smart meter already installed in the households, a router with wi-fi connection to collect the data and to send them to the server and a Raspberry PI with an SD card. The Influx DB is an open-source time-series database handling the timestamp received every 10 seconds and the loads and demand information necessary to execute the algorithm, coded in Python language. Consequently, both live data and disaggregated data are visualized with Grafana dashboards to monitor and analyse the results of the process. The schema proposed in the following *Figure 3* shows the entire procedure of the approach.

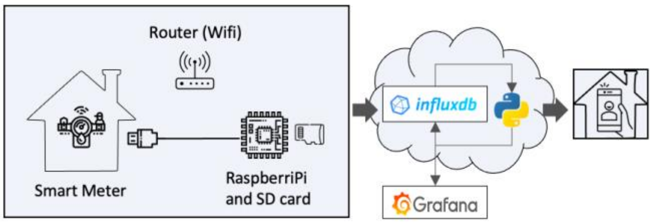

Figure 3 Diagram of System Design

The data has been collected with low-frequency sampling rate of 0,1 Hz, from standard detached houses in the Netherlands used for a four people family. The houses have PV production and three phase powers system. In the specific, one of the houses has been considered as a reference for the development of the algorithm due to the copious availability of information regarding the appliance's specifications and the devices' connection to each circuit-level of the household. However, the measurements for the individual consumption of each device are not available. The data has been collected since April 2020 at a non-constant rate due to an unstable connection. Nevertheless, the signal became stable from August 2020.

2) Event-based disaggregation. Fridge case-study: The case study for the signature recognition approach was selected to be the fridge (with a freezer included). In fact, it belongs both to the type of devices which are always on throughout the day, and to the continuously variable devices. In the latter case, the consumption varies either when the fridge's door is opened by the consumer, letting the cold air exit the device, or according to the number of items contained in the fridge itself. During the night, the energy consumption of the fridge is undisturbed by the consumer's actions, and therefore its signature is more easily detectable.

The characteristics of a load signature are here referred as those periodical features which are representative of the operational profile of a device and therefore allow its unambiguous recognition. For simplification, as it can be observed from *Figure 4*, some letters have been attributed to these typical characteristics. The symbols which directly refer to the active power from *phase III*, connected to the fridge, are *B*, *L* and *G*. They represent respectively the baseload (*B*) (power always on), the net load (*L*) (maximum active power during the fridge cycle minus the baseload) and the gap (*G*) in power consumption that the fridge shows in every cycle between its maximum power and the instant just before decreasing its power to the minimum. The ΔP or load differential indicates the positive or negative changes in the active power, hence if the device is changing its state. Furthermore, it is also very important to consider the fridge cycle in terms of time and frequency: *d* is representing the time when the compressor is in operation to cool down the fridge environment, while *D* is the remaining time before the next operative state. It is possible to associate these two values to the concept of Duty Cycle, or more simply how long the device is on (*d*) within a full cycle.

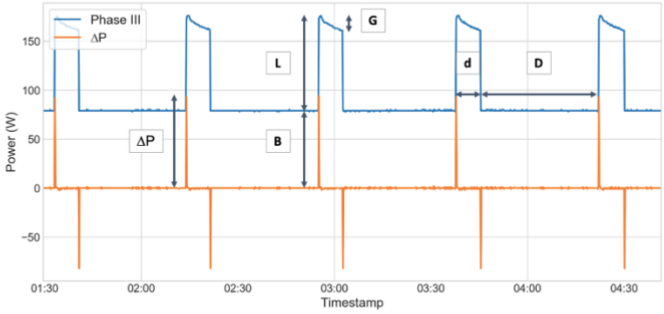

Figure 4 Typical fridge load during the night and representation of its main characteristics: ΔP, L, B, G, d, D

From an efficiency point of view, the aim is to keep the Duty Cycle value as low as possible. In this case study, the Duty Cycle was found to fluctuate between around 17% and 36%. An experiment has been executed where the fridge's door has been kept open for ten minutes in a September's afternoon. In this case, the profile presented a ΔP with a sharp increase (*L*), a large gap in power (*G*) between the maximum power and the mean power during the on cycle (*d*), before the negative ΔP. This translates into a noticeable increase of the Duty Cycle up to 36% compared to the average of 22% of the cycles on the same day. From this analysis, it can be inferred that the characteristics most affected by the door opening event are a gap of power *G* during the positive ΔP and the duration of the cooling cycle *d*. For this reason, their variation over a certain range percentage can be considered as the trigger to detect when the fridge's door was opened or for the recognition of an inefficient device with a very high and constant Duty Cycle.

3) Results and Performance Metrics: The signature recognition method that was developed in this thesis does not use any learning algorithm. Instead, it was inspired by some expert heuristics' methods for event detection, in which some threshold ranges of power variation are pre-defined for a specific appliance. In a similar way, the fridge's signature was identified after the initialization of its main features as it can be seen in *Figure 5*. *Phase III* represents the total electricity consumption on the circuit where the fridge was plugged, and the orange line is the predicted load of the fridge throughout that day. Despite the missing availability of a measurement that could precisely show the consumption of this device, it is evident that the prediction of the load signature was rather accurate for the detection of the fridge.

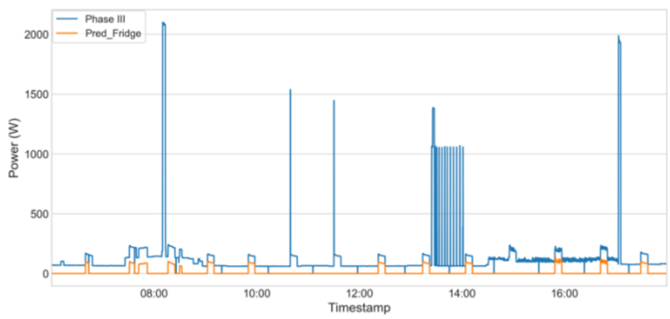

Figure 5 Fridge disaggregation during a typical day in September

The algorithm developed is able to recognize the occurring ΔPs, to detect an event when it happens and to classify it according to the power differences. However, this classification does not recognize all the different appliances since the algorithm was modelled over the specific characteristics of the fridge. It is also worth mentioning that since this method does not use a learning algorithm, the code developed does not learn by itself about the different appliances and does not improve its performances with the larger availability of data. Also, every appliance has to be initialized with a specific number of variables within certain ranges of power to be recognized. The multi-state and continuously variable appliances represent a limitation for this algorithm due to the need of tracking the device state variations over time. However, for the case study of the fridge, the signature recognition algorithm proved to be able to detect unusual behaviours and could thus be used as a trigger to alert the inhabitants of possible misbehaviours such as a door which is not perfectly closed, and which would consequently alter the occurring Duty Cycles.

Due to the unavailability of the individual consumption data for each appliance, especially for the fridge, most of the common metrics for event-based approaches cannot be used in this analysis because there is no possibility of comparison with a ground truth measurement. Despite this, it was possible to calculate the RMSE comparing the sum of the ΔPs obtained with the aggregated consumption in *phase III* which resulted to be 180. This value is comparable to the results obtained using the NILMTK approach for the REDD dataset with both CO and FHMM algorithms presented in *Figure 2*. Finally, the error profile given by the difference between the aggregate *phase III* consumption and disaggregated fridge consumption, averaged 99,6 W per timestamp.

D. *Challenges and Limitations*

Despite having been studied for decades, energy disaggregation still shows limited implementation due to a series of challenges that need to be overcome to allow a broader employment of this technology.

Firstly, the creation of a dataset is a very complex and time-consuming process which makes of data acquisition one of the most problematic steps of energy disaggregation. At the data collection phase, the noises, network malfunctions, data compression and the difference in the type of power meters installed in each household can generate mismatching data, reduced accuracy and loss of raw data. The problem encountered in this thesis for the signature approach method was that the individual loads were not available. In this case, knowing the circuit level data and the breakdown of the appliances connected to each of the three phases was a great simplification which allowed the recognition of the fridge load.

Once data is collected, the challenge lies on the load identification strategies to disaggregate the total electricity consumption. In case of supervised learning approaches, the labelling of training data requires human effort and intervention. Thus, in order to create a tailored signature dataset to train the algorithms, every appliance has to be manually switched on or off to be identified, creating a time-consuming process and fallible results. The signature algorithm proposed in this thesis did not include a learning approach. While this simplified the code and facilitated its replicability, it also limited the potential of the results. The impossibility of improving the recognition with time and the manual labelling of the specific characteristics for every device are some of the consequences.

Finally, the performance of the NILM algorithms is strictly dependent the presence of standardized frameworks that promote unified interfaces, data formats, and performance metrics influences the possibility of comparison, analysis and further research of these algorithms. The lack of individual loads in the signature recognition approach prevented the obtained results to be adequately compared with the NILMTK results since most of the performance metrics adopted are based on the comparison of the disaggregated loads with the ground truth measurements of the single devices.

III. Economic and Social impact

In this section of the thesis, energy disaggregation is analysed from an economic and social perspective. After the technical analysis of NILM which emphasised both the advantages and the limitations of the technology, this chapter explains how energy disaggregation can be implemented and which could be its impact on the consumers.

A. *Methodology*

The socio-economic analysis was performed by consulting literature reviews and market reports. Due to the early stages in the implementation of this technology, only little information is present regarding the potential business models that could be adopted to develop energy disaggregation. Thus, a study was conducted to evaluate the impact of NILM in the American, European and Dutch markets considering the possible economic and energy savings. Also, there are many uncertainties regarding the percentage of obtainable savings from disaggregated energy feedbacks. Thereby, a review of different studies was conducted to specify the different types of feedbacks that can be received by the consumers to increase the savings and their level of engagement on energy conservation.

B. *Economic Impact*

The evaluation of the potential economic impact of an emerging technology requires the analysis of different factors. In the specific, for energy disaggregation there are many different applications, stakeholders and markets that need to be considered.

1) Customer segmentation and NILM applications: There are numerous potential use-cases for energy disaggregation and consequent benefits of appliance-specific over whole-house data applications. These applications can be classified according to the two customer segments: electricity consumers (residential, commercial and industrial), and utilities and service companies.

The first type of potential customers are the electricity consumers who in this thesis will be focused on residential applications. These are tenants and homeowners who are using electricity without having a clear idea on which appliances are the most consuming ones and responsible of their high energy bills. These people can benefit from appliance-specific data to optimize their energy usage and reduce costs [11]. The most commonly mentioned application for energy disaggregation is the possibility of offering these consumers itemized energy bills showing the individual appliances consumption. This would not only lead to a higher energy literacy but also to an increase in the awareness of the distribution of energy consumption per appliance. Another application is to detect anomalies in the electricity consumption. In fact, NILM is able to recognize if an appliance's profile is not correctly shaped, identifying thus malfunctioning or inefficient devices. The fridge case study can serve as an example for this application.

The second customer segment refers to both public and private utility and service companies. A NILM application with a high economic potential for utilities targets Demand Response (DR). This mechanism involves modifying the consumption of electrical loads (consumers demand) in response to the needs of the electricity grid [12].The deferrable or flexible consumption devices identified by energy disaggregation can provide flexibility in the electricity demand which can be used to adjust it to the supply. Moreover, utilities and service companies have to deal with the expensive issue of segmenting their diversified customers in order to personalize energy efficiency recommendations. Engaging clients with online platforms by showing them disaggregated energy information could help utilities to cheaply obtain more data to perform this customer segmentation.

Furthermore, energy disaggregation offers numerous other benefits to policy makers, researchers and developer who share a common concern for energy saving and energy management. In fact, many companies are entering this market trying to deliver innovative and enhanced features in the automotive, home energy management, and security fields [11]. Differently, governments necessitate tools that easily and cheaply would allow a global understanding of the public behaviours before defining new and effective energy policies.

In conclusion, a greater amount of disaggregated data collected would help to analyse energy usage behaviour, to forecast the demand response mechanism potential, to better manage energy in grids, communities and buildings, and to have a greater understanding of electricity loads and their flexibility.

2) Product Implementation and Business Models: There are several approaches to create a business model for the services offered by energy disaggregation. However, it is firstly important to understand the possible ways of implementing a NILM product. Generally, the NILM technology can use two type of solutions to acquire the disaggregated energy consumption data [13][14]:

- Hardware solutions. These are sensor based and involve the development of both hardware and software. A specific device can be plugged to the smart meter to collect the required sampling frequencies in order to perform whole-home monitoring (plug-in devices).
- Software-as-a-Service (SaaS) solutions. These are generally cloud based and only involve the development of the software. The whole-home electricity data must be already available at an adequate sampling frequency and can then be disaggregated through NILM algorithms. The computations can be performed either in the cloud where the data are collected or locally at the smart meter.

Finally, the NILM business models can be analysed and classified into two different categories according to the customer segment that the company delivering disaggregation is targeting and to the selling strategy of the product [13]:

- B2C for energy Consumers (Residential, Commercial and Industrial): delivering a plug-and-play device.
- B2B2C for Utilities and Service Companies: partnering with utilities that are directly providing and installing the smart meters to the customers.

The most common approach is to deliver the disaggregation services directly to the consumers, mainly residential (B2C). However, in recent years, utilities demonstrated to be interested in the services provided by NILM algorithms and an increasing number of partnerships have been stipulated with companies offering energy disaggregation (B2B2C).

3) Market Potential: According to a recent study, the global Home Energy Management System (HEMS) market was valued at $7,7 billion in 2019 and it is expected to reach a value of $26,9 billion by 2025 [15]. This sector includes all the technologies for monitoring, managing and controlling the house. However, the rapid growth foreseen for this market can be considered as an indication of the potential offered by the NILM technology, justified by the introduction of several disaggregation companies among the 24 innovators in the sector considered in the study [15]. Regarding the energy disaggregation field, the US market shows to be the most consolidated due to the high implementation of NILM technologies based on partnerships of disaggregation companies with utilities, as seen with *Uplight*, *Bidglely* and *Innowatts*, already offering NILM services to millions of customers all over the United States. In comparison to the North American market, the European market is still segmented, and companies primarily focus on residential applications. However, due to its competitive energy market, the high electricity prices, the numerous energy efficiency policies and the strict regulation targets, the EU holds a great market potential for energy disaggregation [13].

There are a limited number of studies which propose a cost analysis for the implementation of energy disaggregation. At the moment, the installation of a NILM hardware solution for residential customers who are already equipped with a smart meter is approximated between $200 and $500 [16]. According to the results in this thesis in which savings are obtained with direct and personalized appliance-specific feedbacks, an average of 5% energy savings per household can be expected [14]. This scenario is set as reference to estimate the potential cost benefits of the technology.

The analysis that was conducted in this thesis compares three different markets: United States, Europe and the Netherlands. The percentage of smart meters rollout for each of those markets is expected to be 80%, 77% and 95% respectively by 2024 [17][18][19]. The same percentages have been assumed as the correspondent potential for NILM technology adoption. The total cost of an energy disaggregation hardware solution is overestimated to be 500€ per household. The study showed that in the US, if 80% of the households would benefit from energy disaggregation this could result in 11 TWh of electricity saved nationally [16]. In Europe, where the price of electricity is more than double the price of electricity in the US, the savings would reach 45 TWh of electricity which leads to 9,6 billion euros annual savings. In the case of the Netherlands, the implementation of energy disaggregation would lead to 1 TWh energy savings, corresponding to 179 million euros savings.

However, if the high investment for the NILM hardware is done by the individual residential consumers, the savings that are obtained are relatively small, even when assuming the value of 5% energy savings used for the analysis. Despite those high energy savings, the return of investment for a Dutch family would be 20 years, for a European family it would be less than 12 years, and, in the US, the disaggregation investment would be covered in 9 years.

C. Social Impact

The implementation of a NILM technology able to provide an accurate and detailed appliance-specific breakdown of the energy consumption to consumers is alone not enough to achieve quantifiable energy efficiency improvements. The energy disaggregation data obtained needs to be analysed and converted into actionable feedbacks for the consumers in order to achieve greater energy savings [20]. This transition can be accomplished by incorporating the understanding of the psychological and behavioural characteristics of energy consumers to the technological challenges.

1) *Behavioural Change:* The sustainable utilization of electricity and the reduction of its consumption, defined as energy conservation, constitute an especially difficult challenge for the residential sector. In fact, due to the intangible and ephemeral nature of electricity, energy conservation is not perceived as a coherent field of action by households' inhabitants. For the consumers, their diversified sets of actions and behaviours are difficult to be associated quantifiably with their final energy bills. In this context, recognized in the late 1960s, Environmental Psychology is an interdisciplinary field which studies the interaction between individuals and their surroundings. In the energy disaggregation framework, some of these theories can support the adoption of NILM technologies and the integration of appliance-specific data with effective approaches to provide tailored feedbacks. By doing so, consumers are supported to naturally shift to more sustainable behaviours.

In order to understand which are the different phases that an energy consumer has to face for a successful behavioural change, the *heuristic model of environmentally relevant behaviour* is presented in a modified version which integrates the area of feedbacks [21], together with the *Transtheoretical Model of Change (TMC)*, cited in several publications on energy efficiency [22][23]. According to this simplified version of TMC tailored for the context of residential energy consumption [23], there are three stages which an individual faces when engaging with the process of behavioural change: raise awareness, inform complex actions and maintain sustainable routines. All these steps should be continuously supported by a feedback mechanism which would act as a reinforcement during the behavioural process development, identifying the constraints that people face, and offering information at a time and place which is close to the decision of the consumers (real-time feedbacks).

2) *Energy Feedbacks:* The process of providing people with information about their actions or behaviours is known as feedback. This well-studied mechanism has been proved to be an effective strategy, if adequately delivered, to increase the awareness of energy consumers and consequently lead to energy savings. In this context, the expression eco-feedback refers to a technology that aims at providing feedbacks on individual or group of behaviours in order to reduce the environmental impact and influence energy savings. Based on the environmental psychology theories, this technology is specifically designed to strategically present information to energy consumers considering the complexity of human behaviour and leveraging the availability of sensing technology and interactive displays [24]. In order to be more effective, the eco-feedback technology includes different interventions that can be used individually or in combination, acting as strategies to increase the people's motivation towards pro environmental behaviours. These can be classified as antecedent interventions, occurring prior to the action, and as consequent interventions happening after the action occurred [25].

The antecedent interventions are information, goal setting, modelling and commitment [24][25]. The first and most widely diffused way of intervention is simply presenting information. However, this type of intervention alone results in marginal effects and rarely leads to action. Studies suggest that its effectiveness depends on the individual's values and other motivations of the consumers. Thus, information should be coupled with other kinds of interventions [25]. The second type of antecedent strategy is goal setting which consists in defining an objective to be achieved in a future scenario. Nevertheless, the new goals can be in conflict with pre-existing ones and thus, this strategy is very often assisted by energy programs and other strategies such as information and feedbacks to give people the possibility of checking their own progress. Modelling is another antecedent intervention, and it consists of observing peers as models to perform and demonstrate sustainable behaviours. This is a tool which is based on the concept of learning in a social environment where people apprehend how others behave in certain contexts and are motivated by their actions. Finally, commitment is the promise to behave in a specific way, stimulating the consistency, the attitude and the social influence of an individual. This kind of intervention has long lasting impact on the behaviour, even when the pledge is no longer active.

The second type of eco-feedback strategies are consequent interventions [24][25]. Those are used post-behaviour to affect future decisions, to help the establishment of a routine or to evaluate its effectiveness. There are only two kinds of consequent interventions. The first is the implementation of rewards and penalties, often connected to the goal setting strategy. Some usual techniques are the gamification which involves small rewards such as points when the goal has been reached, and tax reduction. However, this is a short-term strategy to instigate behavioural change and it has been shown that the electricity usage often returns back to the baseline after the reward is removed, differently from commitments. The last consequent intervention is the feedback itself. In this part of the analysis the feedback is to be more specifically considered as the response to the different behaviours and actions that energy consumers perform quotidianly.

The effectiveness of the feedbacks largely depends on how, when and to whom those feedbacks are delivered [26]. In fact, there are many ways in which it is possible to receive feedbacks about the own electricity consumption. *Table 2* distinguishes between the two main types of feedbacks, indirect and direct feedbacks, and their relative subcategories, while further describing their characteristics in accordance with four dimensions: frequency, medium, type of information and type of interaction [27][28].

The indirect feedbacks are the ones received after the electricity has been consumed. Those are traditionally delivered each month by energy retailers and generally show the total electricity consumption of the household and the historic records of the previous years. Contrarily to the standard billing which provides the customers with non-engaging information, the enhanced billing relies on third parties to deliver extra services such as statistical data, energy efficiency advices and social comparisons with similar households. Using this method, the bills could also include some antecedent interventions like goal setting and information. However, both the indirect feedbacks rely on a passive interaction with the consumers. The direct feedbacks on the other hand, are delivered by software platforms, interfaces, or smart hardware located in the house (IHD), allowing the users to actively engage with real-time information on their energy consumption. In particular, the web connected devices feedbacks rely on the connection of the smart meter to the internet, enabling energy disaggregation technologies and allowing the consumers to control their smart devices and receive customized feedbacks, in a bidirectional communication [26][27].

Studies identified that the best hypothetical feedback should be a direct feedback system connected to the internet, delivered through an interactive display located in the household, such as an in-home display, where the users can actively access very frequently, on a daily basis. This display should be positioned in an area of the house accessible to every member of the family and provide a clear and appealing interface which should show the current electricity consumption, its comparison with historic data and with the consumption of other similar households, in addition to monetary and economic measures. Energy disaggregation constitutes another important dimension that should be present in a feedback mechanism in order to reduce the disconnection between consumer's actions and their consequent effects, while better identifying their patterns of consumptions and directly modifying their utilization to diminish the energy consumption [27].

3) *Energy Savings:* Numerous studies have been conducted over the years to understand how the different types of feedbacks and interventions affect human behaviour in the context of energy conservation. After having analysed the substantial variety of adoptable strategies, it is evident that the results obtained from these studies are as well variegated and not always representative of plausible scenarios. In fact, most of the analyses are carried out for small sample size and similar targeted groups, in a short-term scenario of only a few weeks or months. Many biases are also present in the studies such as for example when people undergoing the experiment are aware of the aim of the study and thus, have the tendency to behave more sustainably, called the Hawthorne effect.

Most reviews are not directly aimed at the evaluation of energy disaggregation but instead target all aspects of feedbacks. However, one of these meta-analysis which assessed several studies conducted from 1995 to 2010, defines energy disaggregation as the "holy grail of energy efficiency", attributing more than 12% of annual energy savings to direct feedbacks provided real-time with appliance specific information [14]. Another well esteemed review from 2008 compared different projects to identify the most effective feedback dimension in terms of energy savings. The energy savings resulted to usually range from 5% to 12%. In this case too, the most efficient feedbacks were those that provided frequent and detailed, appliance specific information [21]. Similarly, a recent meta-analysis on 70 studies collected over the last 50 years for the evaluation of the energy feedback systems was conducted by the European Union in 2019. This review suggested that an appropriate feedback mechanism could realistically contribute to reduce between 5% to 10% of a household's energy consumption.

Direct feedbacks with two-way communication interaction

Table 2 Feedbacks classification: indirect and direct

	INDIRECT FEEDBACK		**DIRECT FEEDBACK**	
	Standard Billing	**Enhanced Billing**	**In-Home-Display (IHD)**	**Web Connected Devices**
Frequency	Monthly or Bimonthly	Monthly or Personalized	Always-on	Always-on
Medium	Paper Bill	Paper or Electronic bill	IHD or Web	IHD, Web or Smart Meter
Type of Information	- Energy Consumption - Historic Comparison	- Energy Consumption - Historic and Social Comparison - Energy Efficiency Advice	- Energy Consumption - Historic and Social Comparison - Real-time Information	- Energy Consumption - Historic and Social Comparison - Real-time Information - Energy Disaggregation
Type of Interaction	Passive	Passive	Active	Active and two-way communication

(including energy disaggregation), are recommended and considered more effective than indirect ones.

To date, the only systematic review on the effectiveness of domestic, disaggregated electricity feedback examined twelve studies that were conducted up to 2015 [29]. The result of reduced energy consumption through disaggregation feedbacks ranges from 0,7% to 4,5% for the total population [29]. Additionally, this systematic review considers the "fine-grained" or true disaggregation obtained with high resolution data to not reduce the consumption to an extent which would justify the engineering efforts when compared to the statistically calculated "coarse-grained" disaggregation. In fact, one study conducted by *Bidgely* in 2013 and included in the review, reported a 6% of average energy savings through their true disaggregation, cloud-based NILM solution [30]. This result was not evaluated high enough compared to the estimated 4,5% for the fine-grained disaggregation to be considered necessary, even though no specific studies have ever compared the two technologies directly. Finally, the review questions if the energy disaggregation feedbacks enable greater saving than aggregated ones. Despite the studies may suggest that aggregate feedbacks are more effective, it is important to analyse the differences in the medium that was used for both cases. Overall, the comparison is not equal for the aggregated feedbacks to be considered as more effective than the disaggregated ones. Yet, the analysis suggested that users who were receiving appliance specific feedbacks had higher perception of their energy consumption. For this reason, the authors supported the idea that disaggregation would lead to more significative reductions.

Additionally, the most recent study on energy disaggregation feedbacks was conducted with over 700 people in Germany for six months during 2017 [31]. The study demonstrated that the overall electricity consumption was reduced by 5% with disaggregated feedbacks compared to a fifth control group who received aggregated feedbacks. This value confirms the results obtained from the systematic review.

In conclusion, the studies supported the hypothesis that customized and tailored feedbacks lead consumers to increased electricity savings. In particular, energy disaggregation used as a tool to reduce electricity consumption via feedbacks resulted to have a positive effect on the sense of control and awareness of the consumers and showed potential energy savings between 0,7% and 5%.

D. Challenges and Limitations

The economic and the social impacts of the NILM technology proved to have a great potential on energy consumers and other stakeholders if wisely implemented.

However, the deployment of energy disaggregation is still hampered by the technical challenges of the hardware and SaaS solutions. In fact, to employ energy disaggregation both the solutions still necessitate of an expensive hardware installed inside the household that provide high resolution data from the smart meter. In addition, these solutions, if not implemented inside the household, need to assure the data transmission on the cloud with appropriate protocols and personal data security procedures. Moreover, the prices of this innovative technology are still too high and the results too inaccurate for the energy consumers to personally invest on energy disaggregation. For this reason, at this state of the art, there should be more incentives and motivations from utilities and service companies to deliver NILM technology and foster its implementation. Finally, there is no availability of a market potential study on the NILM technology. The one performed in this thesis has many limitations linked to the missing data on the numbers of smart meters already installed in each of the three nations analysed in the study. Furthermore, the percentage of achievable energy savings used, and the estimated cost of the technology represent large approximations, positively and negatively biased respectively.

On the other side, the main limitation in the social analysis is in the lack of long-term studies comparing several feedback dimensions among large and diversified targeted groups in order to verify the actual effectiveness of the feedback mechanism. Besides, due to the lack of long-term studies there is little statistical information about the response-relapse effect on the decreasing utilization of the feedbacks over time, and the rebound effect according to which the total energy savings are cancelled in time by other triggered energy consumptions. An additional issue for these studies is the inevitable presence of biases caused for example by the small sample size of the experiments, and the publication biases of negative results less willing to be published.

IV. Conclusion

This thesis provided a detailed evaluation of the Non-Intrusive Load Monitoring (NILM) technology from a technical, economic and social prospective, specifically for the residential use-case. The Netherlands was set as reference scenario for the energy disaggregation modelling aimed at the recognition of the fingerprints of the most common household's appliances.

Even though the event-less approaches remain the most computationally intensive methods, those have the potential to achieve the best load identification results through unsupervised learning algorithms. Overall, the FHMM algorithm performed better on both datasets showing a more precise signature recognition of the top five consuming appliances. The REDD datasets obtained the highest percentage of true positives, while the DRED was found to have the lowest root mean square error (RMSE). This means that the algorithms recognized the highest number of exact events for the REDD dataset, and the highest concentration around accurate events for the DRED dataset.

The second energy disaggregation model was designed to test both the hardware installation and the potential of event-based approaches without using machine learning. Despite the economic and easily implementable system design, the algorithm, not including a learning approach, was not able to identify multi-state and continuously variable devices. However, the model precisely recognized the fridge signature and its duty cycles, showing its potential to be used as a trigger to send push notification to the consumers for misfunctioning and inefficient devices. Overall, machine learning based approaches are believed to obtain more accurate results and precise appliances classifications.

Moreover, the market potential of energy disaggregation and in general of the HEMS is forecasted to grow strongly in upcoming years. Nevertheless, the NILM solution resulted to be too expensive for the residential consumers to personally invest in this technology, having a return of investment in the Netherlands of 20 years. In order to be widely implemented, utilities could act as facilitators, since they are already in charge of the smart meter installations in the households. Utilities could promote and deploy the NILM technology on national scales and thus, help the disaggregation solutions to reduce the initial costs in a B2B2C business model.

Finally, energy feedbacks would provide the consumers with engaging and tailored recommendations that would lead them to create new pro environmental habits and act sustainably. A direct feedback system, frequently delivered through in-home displays and in combination with interventions such as goal setting, comparisons, information and appliance specific data, proved to increase energy awareness and savings. The numerous studies conducted throughout the years to understand the percentage of energy savings obtainable using these feedbacks are influenced by several biases, small sample sizes of the targeted groups and the short-term experiments. For this reason, it is still not certain that the identified 0,7% to 5% reduction of energy consumption achievable through energy disaggregation feedbacks is enough to be judged more effective than the one obtained from aggregated feedbacks.

REFERENCES

[1] European Commission, "In focus: Energy efficiency in buildings," *European Commission*, 2020. https://ec.europa.eu/info/news/focus-energy-efficiency-buildings-2020-feb-17_en (accessed May 01, 2020).

[2] BPIE, "97% of buildings in the EU need to be upgraded," *BPIE - Buildings Performance Institute Europe*, 2017. http://bpie.eu/publication/97-of-buildings-in-the-eu-need-to-be-upgraded/.

[3] G. W. Hart, "Nonintrusive Appliance Load Monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992, doi: 10.1109/5.192069.

[4] A. L. de S. Pereira, "Hardware and software platforms to deploy and evaluate non-intrusive load monitoring systems," Universidade da Madeira, 2016.

[5] L. Pereira and N. Nunes, "Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—A review," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 6, pp. 1–17, 2018, doi: 10.1002/widm.1265.

[6] Arne Kaas, "MakeTheMeterSmarter.com – Open Source Disaggregation initiative," *GitHub*, 2018. https://makethemetersmarter.com/disaggregation/.

[7] N. Batra *et al.*, "NILMTK: An open source toolkit for non-intrusive load monitoring," *e-Energy 2014 - Proc. 5th ACM Int. Conf. Futur. Energy Syst.*, pp. 265–276, 2014, doi: 10.1145/2602044.2602051.

[8] A. Fiol, "Algorithms for Energy Disaggregation," p. 47, 2016, [Online]. Available: https://upcommons.upc.edu/bitstream/handle/2117/89937/108724.pdf.

[9] A. Zoha, A. Gluhak, M. Nati, and M. A. Imran, "Low-power appliance monitoring using Factorial Hidden Markov Models," *Proc. 2013 IEEE 8th Int. Conf. Intell. Sensors, Sens. Networks Inf. Process. Sens. Futur. ISSNIP 2013*, vol. 1, no. May 2014, pp. 527–532, 2013, doi: 10.1109/ISSNIP.2013.6529845.

[10] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A. E. Rizzoli, "A Hybrid Signature-based Iterative Disaggregation algorithm for Non-Intrusive Load Monitoring," *Appl. Energy*, vol. 185, no. December 2019, pp. 331–344, 2017, doi: 10.1016/j.apenergy.2016.10.040.

[11] D. Christensen, "NILM Applications for the Energy-Efficient Home," *NREL Natl. Renew. Energy Lab.*, no. June 2019, 2012.

[12] D. A. Kelly, "Disaggregating Smart Meter Readings using Device Signatures by Submitted in partial fulfilment of the requirements for the MSc Degree in Computing Science of Imperial College London," no. January 2011, 2016, doi: 10.13140/RG.2.1.1475.3043.

[13] Cleantech Group, "Energy Disaggregation and NILM – If You Can Measure It, You Can Improve It," *Cleantech Group*, 2019. https://www.cleantech.com/energy-disaggregation-and-nilm-if-you-can-measure-it-you-can-improve-it/ (accessed Dec. 19, 2020).

[14] K. Carrie Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity," *Energy Policy*, vol. 52, pp. 213–234, 2013, doi: 10.1016/j.enpol.2012.08.062.

[15] Frost & Sullivan, "Global Home Energy Management Systems (HEMS) Market," *Frost & Sullivan*, 2019. .

[16] M. Baechler and H. Hao, "Business Case for Nonintrusive Load Monitoring," *U.S. Dep. Energy*, no. May, 2016.

[17] F. Tounquet and C. Alaton, "Benchmarking smart metering deployment in EU," *Eur. Comm.*, no. December, 2019, doi: 10.2833/492070.

[18] European Commission, "Smart grids and meters," *European Commission*, 2019. https://ec.europa.eu/energy/node/83.

[19] Smart Energy International, "Smart meter penetration in North America will reach 81% by 2024," *Smart Energy International*, 2019. https://www.smart-energy.com/industry-sectors/smart-meters/smart-meter-penetration-in-north-america-will-reach-81-by-2024/ (accessed Dec. 19, 2020).

[20] A. Faustine, N. H. Mvungi, S. Kaijage, and K. Michael, "A survey on non-intrusive load monitoring methodies and techniques for energy disaggregation problem," *arXiv*, 2017.

[21] C. Fischer, "Feedback on household electricity consumption: A tool for saving energy?," *Energy Effic.*, vol. 1, no. 1, pp. 79–104, 2008, doi: 10.1007/s12053-008-9009-7.

[22] L. Stronks, "Raising energy awareness : outlines for concrete techniques for analyzing the energy usage of residential and business consumers based on smart meter data," Radboud University Nijmegen, 2018.

[23] Y. Riche, J. Dodge, and R. A. Metoyer, "Studying always-on electricity feedback in the home," *Conf. Hum. Factors Comput. Syst. - Proc.*, vol. 3, pp. 1995–1998, 2010, doi: 10.1145/1753326.1753628.

[24] J. Froehlich, L. Findlater, and J. Landay, "The design of eco-feedback technology," *Conf. Hum. Factors Comput. Syst. - Proc.*, vol. 3, no. January 2010, pp. 1999–2008, 2010, doi: 10.1145/1753326.1753629.

[25] R. Sokoloski, "Disaggregated Electricity Consumption: Using Appliance-Specific Feedback to Promote Energy Conservation," California State University San Marcos, 2015.

[26] P. Zangheri, T. Serrenho, and P. Bertoldi, "Energy savings from feedback systems: A meta-studies' review," *Energies*, vol. 12, no. 19, 2019, doi: 10.3390/en12193788.

[27] T. Serrenho, P. Zangheri, and P. Bertoldi, *Energy Feedback Systems: Evaluation of Meta-studies on energy savings through feedback*. European Commission, 2015.

[28] K. Ehrhardt-Martinez, K. A. Donnelly, and J. A. Laitner, "Advanced metering initiatives and residential feedback programs: A meta-review for household electricity-saving opportunities," *Am. Counc. an Energy Effic. Econ.*, 2010.

[29] J. Kelly and W. Knottenbelt, "Does disaggregated electricity feedback reduce domestic electricity consumption? A systematic review of the literature," no. May, 2016, [Online]. Available: http://arxiv.org/abs/1605.00962.

[30] A. Gupta and P. Chakravarty, "Impact of Energy Disaggregation on Consumer Behavior," *Bidgely*, 2013.

[31] A. Gerster and M. A. Andor, "Disaggregate Consumption Feedback and Energy Conservation," *CRC TR 224 Discuss. Pap. Ser.*, no. 182, 2020.