



Article

Load Flexibility Forecast for DR Using Non-Intrusive Load Monitoring in the Residential Sector

Alexandre Lucas ^{*}, Luca Jansen, Nikoleta Andreadou , Evangelos Kotsakis and Marcelo Masera

European Commission, Joint Research Centre (JRC), 21027 Ispra (VA), Italy

* Correspondence: Alexandre.Lucas@ec.europa.eu; Tel.: +351-961741327

Received: 4 June 2019; Accepted: 9 July 2019; Published: 16 July 2019



Abstract: Demand response services and energy communities are set to be vital in bringing citizens to the core of the energy transition. The success of load flexibility integration in the electricity market, provided by demand response services, will depend on a redesign or adaptation of the current regulatory framework, which so far only reaches large industrial electricity users. However, due to the high contribution of the residential sector to electricity consumption, there is huge potential when considering the aggregated load flexibility of this sector. Nevertheless, challenges remain in load flexibility estimation and attaining data integrity while respecting consumer privacy. This study presents a methodology to estimate such flexibility by integrating a non-intrusive load monitoring approach to load disaggregation algorithms in order to train a machine-learning model. We then apply a categorization of loads and develop flexibility criteria, targeting each load flexibility amplitude with a corresponding time. Two datasets, Residential Energy Disaggregation Dataset (REDD) and Refit, are used to simulate the flexibility for a specific household, applying it to a grid balancing event request. Two algorithms are used for load disaggregation, Combinatorial Optimization, and a Factorial Hidden Markov model, and the U.K. demand response Short Term Operating Reserve (STOR) program is used for market integration. Results show a maximum flexibility power of 200–245 W and 180–500 W for the REDD and Refit datasets, respectively. The accuracy metrics of the flexibility models are presented, and results are discussed considering market barriers.

Keywords: flexibility forecast; demand response; STOR; disaggregated loads; non-intrusive monitoring

1. Introduction

Balancing power in the grid is a progressively challenging task, both for long-term and short-term balance. The ever-increasing contribution of renewable energy generation coupled with the associated variability and unpredictability of supply is expected to further complicate this task by increasing the number and scale of sharp fluctuations in demand/supply mismatch. On June 16, 2019, an outage left 47.5 million people without electricity in Argentina and neighbouring countries, supporting the need for a more resilient, smart, and flexible grid, both at transmission and distribution level. The traditional mechanisms for managing these fluctuations by controllable, fossil-fuel-based power plants are becoming less effective as the proportion of such generation in the energy system is decreasing. Demand response (DR) is a promising method for balancing supply and demand in power systems, with a high share of variable renewable energy generation offering flexibility to the market. In many countries such as the United Kingdom, France, and northern European Union member states, proper regulatory frameworks have been developed or partially implemented, meaning demand response services are already available for large units as industries or commercial clients. DR services may be explored through explicit or implicit demand response activities. The first requires active participation of end users responding (either automatically or by giving permission to a third party) to requests from

within an existing framework agreement. Implicit demand response is that which occurs naturally due to a tariff framework, which encourages energy use at certain times and discourages energy use at others, for example for peak reduction. There is a focus herein on the forms of explicit demand response, as these are technically more difficult to achieve and is the main activity carried out by aggregators when signalled to participate in an event. Since typical consumers do not participate or trade at all in energy markets, they may subscribe the services of an aggregator to do so. Aggregators manage assets of varying characteristics, such as response time and size, and offer backup for those individual assets as part of the aggregation activity, increasing the overall reliability and reducing the risk of individual client participation. The aggregated load is used to bid into the market, acting as a single resource.

An aggregator can only present its services to the market successfully if their costumers show high levels of participation in DR programs. Furthermore, an aggregator relies on excellent data analytic processes, asset management, and bid strategies to see its bids being cleared, while maximising its revenues. Typical clients or assets may be steel factories, the concrete industry, or hospitals, where a set of chillers or furnaces may easily reach several units of MegaWatt. While not being widely spread, the aggregation activity is not novel, and hence, a growing interest in accessing new loads and diversifying assets and portfolios is arising. Smaller and medium customers have become the targets of interest for such applications, which coupled with larger assets, could form diversified and leaner portfolios to respond to different needs. Even though the estimation of flexibility focusing on specific assets can be found in the literature, an integrated approach is seldom seen. In this study we propose estimation of the flexibility of a typical client's set of assets, in this case a household. For this, we present an approach that combines non-intrusive load monitoring for data collection, which is used with disaggregation algorithms to train a model for load forecast. This forecast is then subjected to a flexibility criterion proposed in this study, which forecasts the flexibility based on categorization of assets, where the power amplitude variation is presented along with its allowed time correspondence. This linear behaviour between the variation of the loads and the time an asset can endure such variation (based on customers comfort) can be updated by the customer's feedback. The output flexibility of the studied household can then be grouped with other flexibilities by an aggregator and presented to the market.

1.1. Business Models

Several business models exist to explore DR services. Depending on the market design, the services to be provided may have different characteristics, but in general, DR Strategies can be clustered into 6 categories (one implicit and five explicit): Direct Load Control, Load Curtailment Requests, Demand Reduction Bidding, Ancillary Service Provision, Emergency Response, and Implicit DR.

Based on experience from DR markets in European Union countries [1], the main categories with respect to the different related income streams are described:

- **Explicit DR as stand-alone service:** In this business model, a DR aggregator bundles DR potentials from different clients, which as stand-alone potentials are too small to be offered to the various flexibility markets. The income streams originate from payments either from the Transmission or Distribution System Operator (TSO/DSO) or from the Balance Responsible Party (BRP), which are usually shared between the DR aggregator and the clients. In this business model the service of DR aggregation has no interlinkage to power supply or any other service for the client.
- **Explicit DR combined with energy efficiency services:** In this business model the DR aggregation service is embedded into a more comprehensive energy efficiency service (EES). This approach is sometimes referred to as "dual service" and requires clear and transparent definition of the roles of the energy service company (ESCO) and the DR aggregator.
- **Implicit DR service aiming at the optimal utilization of time-of-use (ToU) contracts:** This business model is enabled by the fact that electricity customers already have tariffs with different price levels depending on the time of consumption. As many customers will not be able to exploit the

full saving potential of ToU tariffs, an external service provider takes care of load shifts in the equipment of the client, in a way that the client receives maximum benefit from an (existing) ToU tariff, being remunerated either by fixed or performance-based fees.

- Implicit DR including power supply: This business model combines DR with the role of a retailer on the electricity market. In addition to its usual function of selling electricity to customers, the retailer has access to DR potential at the customers’ sites. From the retailer’s point of view, the access to DR potential represents value, as it may lead to savings both in wholesale prices and in balancing energy payments. The savings achieved are shared with the customers—usually through favourable tariffs.
- Microgrid management is a specific DR business model for those cases where the regulatory framework allows for microgrids. To this extent, a microgrid is, hence, defined as a group of interconnected loads and distributed energy resources that are part of a common infrastructure whose boundaries are electrically defined at a higher hierarchical level, such as a transformer or substation, which acts as a single controllable entity with respect to the main grid.

Considering the explicit DR as a standalone service, an aggregator has at its disposal several market participation options. From the several options of programs in the balance market, some are shown in Table 1 as programs explored by a leading aggregator in the United Kingdom [2]. These are Dynamic Firm Frequency Response (DFFR), Static Firm Frequency Response (SFFR), Constraint Management, Short Term Operating Reserve (STOR), and Capacity Market. The constraint management program was recently implemented in the United Kingdom (May 2019) as a distribution network operator (DNO) program, and focuses on energy communities, therefore the geographical location of the assets is fundamental. In this research we will focus on the short-term operating reserve (STOR) program for simulation, which due to the nature of the assets, could enable the bidding of the flexibility under study. The availability window for the STOR program depends on the season and may experience slight changes.

Table 1. Demand response programs example for the U.K. balance market.

| Type of Programs | DFFR | SFFR | Constraint Management | STOR | Capacity Market |
|-----------------------|---------------------------------|---------------------------|------------------------------|----------------------------------|--|
| Response Time | 1 s | 1–30 s | 15 min | 15 min | 4 h |
| Duration of the event | Continuous | 30 min fixed | 2–4 h | Max 2 h | 4 h |
| Payments | Availability | Availability | Availability and utilization | Availability and utilization | Availability |
| Types | Dynamic—Any Variance from 50 Hz | Static—50.2 Hz to 49.7 Hz | Static | Demand turn up/turn down | 1 = Existing Generation; 2 = DSR A = Proven; B = Un Proven |
| Availability Window | 24/7 | 24/7 | Summer/winter | 06:00 to 14:00 16:30 to 22:00 | 24/7 |

DFFR—Dynamic Firm Frequency Response; SFFR—Static Firm Frequency Response; STOR—Short Term Operating Reserve; DSR—Demand Side Response

1.2. Demand Response Trends and Research

The residential sector’s contribution is approximately 29% in terms of final electricity consumption, when compared to transport, agricultural and forestry, services, and industry [3]. This provides huge unlocked potential, enabling demand-side management for grid stability through the participation in DR programs. Several barriers have prevented this from happening. The disaggregated nature of the assets (households) and their relatively low size, in terms of providing a significant bid to the market, have posed some challenges. Figure 1 shows typical household consumption share by appliances in the

European Union 27 members (EU-27), based on a Joint Research Centre (JRC) report [3], which suggest that the majority of the electric energy could be flexible.

The development of standards and protocols (OpenADR, Saref), distributed ledger technologies, smart contracts, and the much anticipated 5G networks facilitate residential DR development. However, particular difficulty has been the load flexibility forecast of residential units. If wrongly predicted, this could mean financial losses to a potential aggregator, if this power is not delivered to the market as promised. This is especially true for explicit demand response. Furthermore, the associated suspiciousness of households', with the lack of privacy from load monitoring, may lead to low levels of participation in such programs. Reference projects developing pilots and developing DR initiatives in the residential sector in Europe often deal with the integration of buildings or a set of buildings as the flexible management unit [4,5].

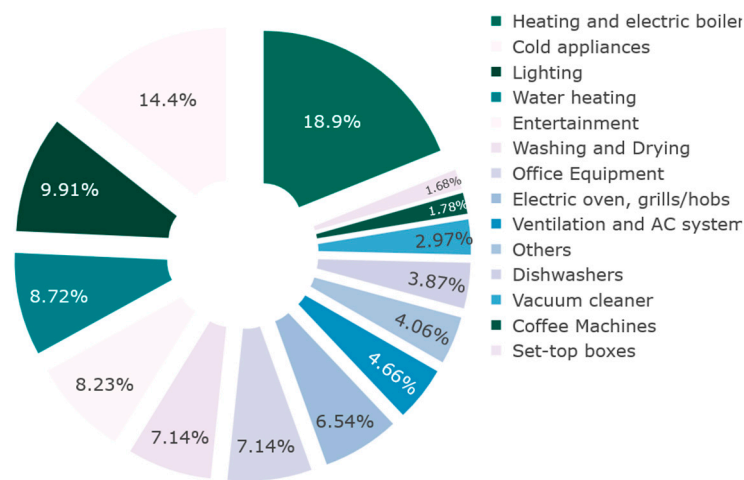


Figure 1. Residential electricity consumption breakdown in the European Union 27 (EU-27), own source based on [3].

A vast number of research and innovation projects involving DR pilots have demonstrated a growing interest in this topic. This is evidenced by multiple European H2020 projects being financed in the last 3 years, and a growing number of stakeholders being involved. Reference projects, among others, are Delta [6], which presents a decentralized architecture and introduces the virtual node concept, responsible for carrying out the complex tasks associated with managing DR assets, and aims at introducing small and medium consumers into DR programs making use of smart contracts. Another project, demand response in blocks of buildings (DR BoB) [4], aims to demonstrate the economic and environmental benefits of DR, focusing on buildings as manageable units in pilots across different countries. Demand response integration technologies (DRiVe) [7] also aims to deliver an integrated, interoperable, and secure DR management platform to be used by aggregators offering advanced forecasting techniques, optimization, fast-response capabilities, and enhanced user interaction, in compliance with Open Automated Demand Response (ADR). The Integrated demand response solution towards energy positive neighbourhoods (RESPOND) [8] project goal is to roll out a monitoring and control solution focused on providing DR services at a community or building level. The developed system is set to adjust indoor and outdoor conditions and comfort levels in real time, as well as detect energy conservation opportunities.

The demand response for energy cooperatives (FLEXCoop) [9] introduces a complete DR framework, providing a tool suite for aggregators and residential electricity consumers. The aim is to enable aggregators to exploit the flexibility of end users, giving them the possibility of participating actively in energy management. The project presents models for load and generation forecasting and features a visual and thermal comfort adaptive model to inform participants about flexibility criteria. In addition to this, the enabling new demand response advanced, market oriented and secure

technologies, solutions and business models (eDREAM) [10] project makes use of distributed ledger technologies and aims to contribute to the transformation of traditional energy market concepts, considering smart grid capabilities and novel decentralized and community-driven energy systems. The main goal is the exploration of local capacities, constraints, and Virtual Power Plants. The project investigates optimisation scenarios towards local and secure grid nodes stabilization. The ultimate vision is the delivery of a novel near real-time Closed Loop optimal blockchain based on a Demand Response ecosystem, enabling distribution system operators (DSOs) and aggregators to cooperate in an efficient and secure way.

The mentioned revision is not exhaustive and only constitutes a sample of H2020 projects focusing on DR. The large majority of the projects revised focus on hardware development and explore the communication possibilities and challenges of accomplishing DR interactions on a grand scale. Consumer acceptance seems to be somehow overlooked and is typically not part of the nature of such projects. Furthermore, the privacy or intrusiveness of the data collection procedure of individual appliances to train models is often not considered. These two aspects, however, will play a major role if a massive adoption of residential DR is to take place. Non-intrusive load monitoring is, hence, a solution with high potential to complement many of the projects being rolled-out.

1.3. Contribution

In this study, we develop a model to forecast the flexibility for a given household load diagram based on an expected load forecast. We use a machine-learning library, non-intrusive load monitoring tool kit (NILMTK) [11], and apply non-intrusive load monitoring data gathering and forecasting by training a model with historical data and applying two disaggregation algorithms. The training data is gathered data, providing a disaggregation per appliance, which can be collected using different approaches. Depending on the level of privacy desired, there are two common approaches followed: (i) using high frequency meters to read a house's consumption for a determined time window, which recognizes the individual signature of each appliance, which is the non-intrusive way; (ii) to take the measurements directly from the devices themselves, implying an invasive way of collecting data. The summed disaggregated data matches the load profile presented on the smart meter (aggregated). Once the model has been trained, the remaining load profiles are submitted as a test. We used two popular methods to run the prediction and disaggregation: the first was a Combinatorial Optimization, and the second a Factorial Hidden Markov model (FHMM).

After having obtained the disaggregated data, we proceeded with an analysis at a household level, carrying out categorization of the loads predicted, separating them into shiftable, flexible and shiftable, and flexible and non-shiftable groups. We then performed an evaluation of the amount of variation a load could provide based on a literature review, and possible time windows for such variations or shifts to occur, maintaining an estimated comfort level. The methodology is verified using two public datasets, the REFIT [12] Electrical Load Measurements dataset and the Reference Energy Disaggregation Data Set (REDD). The main study contribution is the automated procedure with the creation of a flexibility criteria and the application to a household with a set of assets.

2. Non-Intrusive Approach

Other than the cost factor, a crucial element to unlock demand response in the commercial and residential sectors is to acknowledge the sensitivity to privacy, and hence, to be able to gather information in a non-intrusive way. In effect, this can be accomplished with high frequency meters, which collect information of the electric circuits within a building at rates of around 15 kHz and are able to detect individual signatures corresponding to consumption patterns, distinguishing one appliance from the other. This approach is called Non-intrusive load monitoring (NILM) and is explained and defended by several authors [13] as being an advanced, simple, and low cost solution. Such data set composition can be seen, for example, in the REDD dataset used in this study. The Nilm toolkit (NilmTK) was developed to assist such a task and has been used in different studies [14,15].

Following this approach has several advantages. First, it allows the household's occupants to be informed about the energy use of each appliance, empowering them to take actions towards energy optimization. Second, this system can provide comparative information on an appliance's performance, for example by quantifying the savings when an old, inefficient appliance is replaced by a more efficient one. Third, if the NILM system is able to monitor the time of use and pattern of each appliance, implicit demand response signals can be advised by the system to shift the load to a time of day when electricity pricing is more convenient [13]. However, as mentioned before, collecting data sets is expensive, time consuming, and for a time step of 1 Hz (1 s) may result in large files that are difficult to store or process [14].

As part of the python NilmTK original library there are four available algorithms: Combinatorial Optimization [16], Factorial Hidden Markov Model [17], Hart Algorithm [18], and Maximum Likelihood Estimation [19]. The choice depends on factors such as the size of the data set, the acceptable training time, the completeness of the data, and accuracy. Moreover, the library is capable of working with third party algorithms, such as the Latent Bayesian Melding [20] and the Neural NILM [21]. In this study, we use two commonly used benchmark disaggregation algorithms: combinatorial optimisation (CO) [18] and Factorial Hidden Markov model (FHMM) [17]. The aim of the inclusion of these algorithms is not to present state-of-the-art disaggregation results, but instead to enable new approaches to be compared to well-studied benchmark approaches, without requiring the reimplementations of such algorithms.

2.1. Combinatorial Optimization

The CO algorithm estimates the optimal combination of appliance status, minimizing the difference between the sum of the predicted appliance's power amplitude and the observed aggregate power. This method belongs to the supervised disaggregation methods, because it requires appliance signatures to be available in order to run the model. One of the main applications of this method is to solve the typical "Traveling Salesman Problem" and the main steps can be summarized as follows:

- Get the aggregated data, x_t ;
- Form an appliance set: $[n_1, \dots, n_N]$;
- At a specific time, an appliance is in a single state, showing that the n^{th} appliance can be at the k -th state at time moment t :

$$\sum_{k=1}^{k=K} z_{t,k}^n = 1, z_{t,k}^n \in 0, 1 \quad (1)$$

- Power consumption by n -th appliance in k -th state is:

$$\theta_{t,k}^n = \sum_{k=1}^K z_{t,k}^n \cdot \mu_k^n \quad (2)$$

where μ^n is the power drawn from n -th appliance, μ_k^n is the power drawn by n -th appliance, which is in the k^{th} state.

- Overall consumption of all appliances:

$$\hat{x}_t = \sum_{n=1}^N \sum_{k=1}^K z_{t,k}^n \cdot \mu_k^n \quad (3)$$

- The error after the load assignment is:

$$e_t = \left| x_t - \sum_{n=1}^N \sum_{k=1}^K z_{t,k}^n \cdot \mu_k^n \right| \quad (4)$$

- Try to minimize the error by finding the optimal combination of appliances in different states, which gives this minimized error:

$$z_t = \operatorname{argmin}_{z_t} \left| x_t - \sum_{n=1}^N \sum_{k=1}^K z_{t,k}^n \mu_k^n \right| \quad (5)$$

This method becomes more complex when more possible combinations in the dataset are available. The error is difficult to minimize when the set of appliances available in the database is not complete.

2.2. Factorial Hidden Markov Model

This is a probabilistic model and takes into account the duration of operation of a specific appliance or asset. In addition to the time of usage, their real power consumption is also considered. The aggregated power from the loads is given as $\{Y_t\}$ with $t = \{1, \dots, T\}$ and is the visible variable, whereas the hidden variables $\{S_t\}$ are the states of the appliances that have led to observing Y . The hidden states are connected by a Markov transition structure, as explained in previous work [17]. The concept of the method is to analyze the observed state and figure out the hidden states that have resulted in this observation. Each device is modeled with a Markov model, which describes the states of the device and the connections between them, and the probabilities that one state will lead to the other. The overall observation is the addition of the different states of each device. The model becomes more complex when the number of appliances involved is increased and when unknown appliances are included in the overall consumption [17,22].

3. Methodology

In this study we use Python for data processing and the Scikit learn [23] library for machine learning. For load disaggregation we use NilmTK library v0.2 [11]. All processing and model simulations were carried out using a desktop unit, with two processors with 2.19 GHz, 64 GB RAM.

Currently, not many large flexible appliances are available in households, and therefore the provided flexibility is limited. Household loads, such as wet appliances (tumble dryer, dishwasher, and washing machine), can only be shifted, and refrigerators or freezers offer insufficient flexibility for DR potential alone. The most impactful appliances in typical houses are heating systems and lights, even though lighting system demand has been reduced dramatically over the years due to each dataset. The outcome of the developed model is given in an aggregated and disaggregated manner. The results may be exported to CSV or similar files, allowing a straightforward way to replicate and aggregate it to other houses and flexibilities.

3.1. The Dataset

REDD and Refit are popular datasets and can be found in several studies [24–26]. REDD was released as the first publicly available data set, collected specifically to help researchers conduct NILM and household consumption studies. The data is organized and described to aid energy disaggregation, enabling asset distinction from an aggregated electricity load profile. The dataset was retrieved from six North-American homes by the Massachusetts Institute of Technology (MIT), and includes several weeks of recording. The current monitoring was performed on both phases of power and with a voltage monitor on one phase. The recordings were performed with a high frequency meter in order to detect individual signatures at a high frequency (15 kHz). The houses have different types and number of appliances and capture different time windows in 2011. Each house has a maximum number of recordings of up to 24 individual circuits and each is labelled with a category of appliance, presenting a 0.5 Hz time step of recordings. In addition to this the dataset has up to 20 plug monitors per house, which are recorded at a 1 Hz time step, where various assets are grouped in a single circuit [24]. In this

study, we have used house number 1 to test, which includes 18 appliances and monitors and 2 site meters. The mains sample rate was performed at 1 second and the sub meters (monitors) at 3 seconds.

The Refit [12] dataset was retrieved by the University of Strathclyde. It includes the power consumption records for 20 households in the United Kingdom. The dataset includes aggregated and disaggregated data per appliance, with 8 s time steps. The dataset is meant to support research in the fields of energy conservation, efficiency, advanced energy services, NILM, demand response opportunities, appliance usage analysis, consumption pattern identification, statistical studies, and smart home automation. The dataset was made available on June 16, 2016, and covers the time window of October 2013 to June 2015. Each of the sub files of the 20 houses contain information on occupancy, construction year, appliances owned, type (detached, semi-detached, or mid-terrace, and size regarding the number of bedrooms (3, 4, or 5). From all the appliances identified in the houses, only 9 were monitored. The large majority of devices are described by manufacturer and model. The tested house in the model was house 15, which was built in the 1980's. In total, 48 appliances detected and it is a detached-type house with 5 bedrooms. It is stated that some appliances, such as televisions or computers, may in fact consist of multiple uses, such as lights or others.

Even though flexibility may refer to load increase or decrease, we will focus on the latter due to its higher potential for application, but a similar methodology can be applied to both types of variations.

3.2. Load Categorization and Flexibility Criteria

Different approaches to categorizing appliances can be found in the literature. Some authors [27] categorize household appliances into two groups according to their potential to store energy. According to the authors, only two types of appliances show this property, i.e., heating systems and refrigerator and freezers. These two are, thus, considered to have high potential for flexibility provision. The authors claim that some water appliances, such as washing machines, dryers, or dish washers, which belong to the non-storable energy category, can be also be flexible when controlled automatically. Other authors [28] define the flexibility of household appliances as either shiftable or reducible loads within the limits of complying with users' comfort, and categorize them according to the ability to provide such comfort.

Studies on load disaggregation can extensively be found in the literature. Mocanu et al. [29] developed a load disaggregation model based on a Restricted Boltzmann Machine to extract events of appliance usage from aggregated load data. In order to estimate the flexibility from their results, they considered the entire loads from refrigerators, electric heaters, washer-dryers and dishwashers to be flexible.

Ji and Rajagopal [28] define flexibility as realizable increase or decrease of energy consumption per appliance. The basis of their study is a dataset that consists of appliance-specific consumption patterns collected from 345 household in Austin, Texas, for the year 2016. According to the authors, the variations in consumer patterns of different households for the same appliance group give rise to flexibility potential. For deferrable loads, a user-specific maximum delay time τ is defined. Authors observe that the flexibility of shiftable and curtailable loads, such as electric vehicles (EV) and pool pumps, depend highly on the time of the day.

Li and Pye [30] investigated the effect of flexibility of household appliances on the electricity demand profiles in future scenarios using the UK energy systems optimisation model (UKTM), which is based on the integrated Markal-Efom system (TIMES) model, the international energy agency (IEA) energy system analysis tool [30]. They assume the shifting potential to be 100% for washing machines, tumble dryers, and EV, and a shifting potential of 1 h is suggested for water heaters, space heaters, and refrigerators and freezers.

Regardless of the approach or focus of the studies found in the literature, there is a consensus in categorizing appliances as deferrable and flexible in terms of load curtailment or increase. Therefore, in this study three categories were chosen, where appliances can be: shifted (deferred) but not varied, shifted and varied, and not shifted but varied. The limitations of flexibility were observed to depend on a ratio between power and time, during which a certain power reduction could be applied.

The information reviewed on the time limitations and amplitudes of flexibility is based on a literature review and is summarized in Table 2. Most publications focus on different appliances and it is difficult to find a comprehensive list that can sum them all up. It is, however, advised that for flexibility forecast purposes, the circuits and loads to be monitored are the ones in Table 2, so that the flexibility forecast can reach its full potential.

Table 2. Categorisation and flexibility criteria of appliances.

| Type of Appliance | Appliance | Flexibility from Nominal Power | Delay/Duration in min | Mean ('n') in min |
|-----------------------------|------------------|--------------------------------|-----------------------|-------------------|
| Deferrable but non-flexible | Washing machine | −100% | 180–540 [27,30,31] | 360 |
| | Dryer | −100% | 180–540 [27,30–32] | 360 |
| | Dish washer | −100% | 180–540 [27,31,32] | 360 |
| Deferrable and flexible | Electric Vehicle | −100% | 0–336 [28,30,32] | 120 |
| | | +100% | | 336 |
| | Pool pump | −100% | 300 [28] | 300 |
| | Space heating | −100% | 60 [30,33–35] | 60 |
| | | +100% | 15 [33–35] | 15 |
| | Heat pump | −100% | 60 [31,36] | 60 |
| | Water heating | −100% | 15–180 [27,30,31] | 97.5 |
| | | +42–75% | 120 [32] | 120 |
| | Air conditioning | −100% | 8 [37] | 8 |
| | | +100% | 8 [37] | 8 |
| | Refrigerator | −100% | 15–120 [27,30,31] | 67.5 |
| | | +100% | 200–400 [38,39] | 300 |
| | Freezer | −100% | 15–120 [27,30,31] | 67.5 |
| Non-deferrable but flexible | Lighting | −0–10% | Permanent [40] | 300 |

Enabling flexibility services to small and medium consumers means increasing the number of assets to focus on to achieve the same nominal power as typical industrial units may have. Even though there are a variety of flexible assets in a household, not all of them can be found in every house. Due to their nature, impact, and time of use, the appliances are more likely to be able to participate in explicit demand response programs would be heating or cooling systems, as well as lighting systems. Table 2 shows the variations of up to 100% for the different flexible appliances found in the literature and the duration n of the variation reported in those studies. The flexibility amplitude and duration may both be adjusted with user inputs, updating their level of thermal and visual comfort periodically. In the simulation developed in this study, we consider the mean values for endurance time found in the literature, or an estimated level of 95% of household comfort.

Since the minimum bid time window or response time in the STOR program is 15 minutes, the assets' flexible power can be rearranged according to two inputs. The first has to do with the time when the event begins and the second with its duration. For the sake of the example, let us consider the air conditioning data in Table 2, which estimates that the air conditioning (AC) system can be reduced by 100% during a time n of 8 minutes, until it is noticed or finds a predefined discomfort. If statistical data exists, one could use a statistical distribution or choose a regression type to obtain the power as the time varies. We consider a linear behavior in all appliances, meaning that 50% of AC power reduction could be sustained for 16 minutes, 25% during for 32 minutes, and so on. If an event duration is signaled to last for 2 h, then the amount of flexible power during 8 periods of 15 minutes will be approximately 6.67% of the measured power.

Figure 2 presents the sequence diagram to treat the data, from obtaining the load consumption dataset to delivering a flexibility forecast load diagram. The sequence of the model starts with the access to a dataset of appliances of a given household. The dataset contains load diagrams with power and time, with a 1-sec time step to a 1-minute time step, depending on the application, number of

appliances, and recording time window. Such data is partially used to train a model, and it should do so periodically to account for variations in occupancy or seasonal pattern changes.

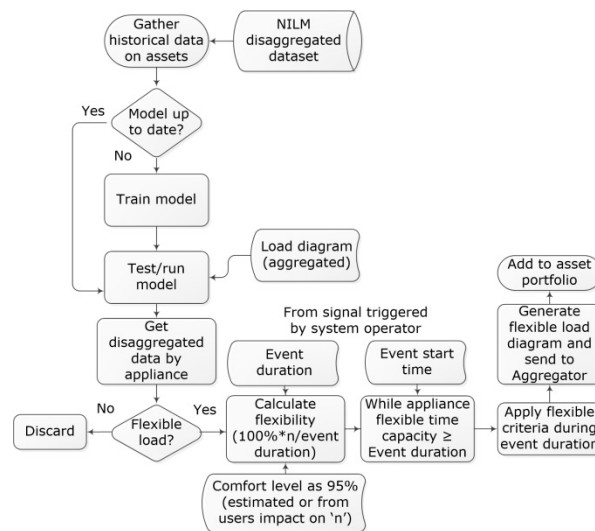


Figure 2. Model implementation diagram for load flexibility estimation.

In order to improve the training time it is advisable to choose a targeted number of appliances as the top ones in terms of power to train the model with. If no changes are observed and the model is trained already, a given aggregated load profile may be tested. This can be retrieved by a load forecast tool. The output of the model is a disaggregated profile by appliance (according to the trained ones) for the same time window of the submitted test. The model is then ready to pursue the flexibility calculations. It starts by choosing those appliances which are flexible from those which are not (discarding them) from the disaggregated results. It then collects the signal from a system operator informing the expected event duration and start time. With such information a linear interpolation of the amplitude of the power is done against the duration of the event, according to Table 2 and considering a comfort level that can be updated in order to personalized flexible power perceptual reduction (flexibility criteria). Knowing the event starting time, the application criteria is applied to the disaggregated load diagram for each appliance and a final output is obtained.

3.3. Simulation

The interactions between the actors can be seen in Figure 3 for a STOR service. The system operator starts by signalling to the aggregator that a demand response will be required in a certain time window. The aggregator, considering an incentive, requests the availability and flexibility forecast of its assets and receives a response. With a response based on the forecast and available assets, the aggregator chooses the bid it makes to the market and waits for the activation order. Once the order is received, a close to real-time order is also given to the client and chosen flexible assets. After the service has been provided the financial settlement is complete.

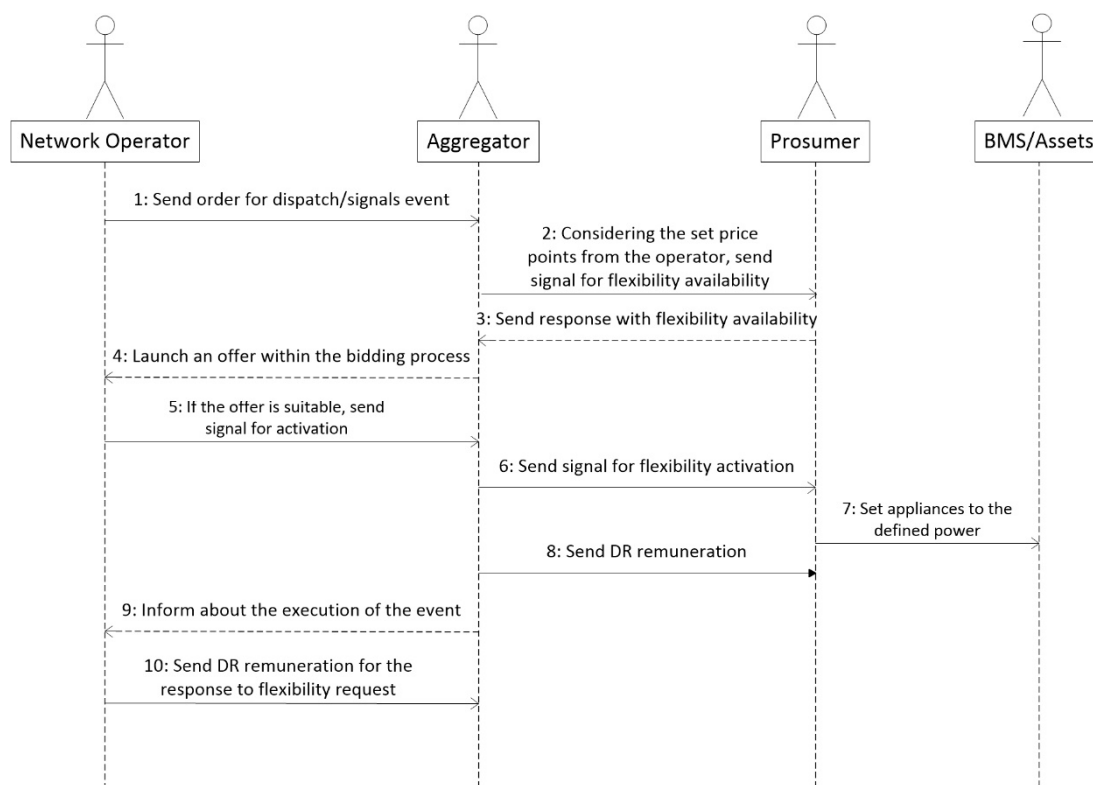


Figure 3. Demand Reduction Bidding Sequence of Interactions diagram.

The Refit dataset portion used for training captures three months of data from October 10, 2014, to December 31, 2014, with over 7.9 million observations in that period, whereas the REDD dataset captures a little over one month, from April 18, 2011, to May 24, 2011, comprising just over 1 million observations. Given the variety of appliances, and that some have relatively low power when used, only the top 8 appliances (in terms of power) were investigated, using the “select_top_k (k = 8)” method provided by the NilmTK library. Given the maximum time of participation in the DR STOR program of 2 h, the corresponding flexibility based on Table 2 is estimated and shown in Table 3. This flexibility criteria or percentages are applied to the corresponding load and time step within the requested 2 h.

Table 3. Flexibility percentage applied to each appliance adjusted to a 2 h period.

| Device | Pool Pump | Space Heating | Heat Pump | Water Heating | AC | Refrigerator | Freezer | Lighting |
|-----------------------------|-----------|---------------|-----------|---------------|-------|--------------|---------|----------|
| Flexibility applied for 2 h | 100% | 50.00% | 50.00% | 81.25% | 6.67% | 56.25% | 56.25% | 10.00% |

The two algorithms were applied, and the metrics were analyzed. Given the results, the best prediction according to the error metrics and F1 score were used to apply the flexibility criteria. As can be seen in Table 4, the Refit dataset with the CO algorithm, even though its performance is close to FHMM, reports a higher average F1 score, and in general reports lower errors. On the other hand, the REDD dataset performed better with the FHMM, however having higher times of training and testing. Figure 4 shows the disaggregated training data used for the top 8 appliances.

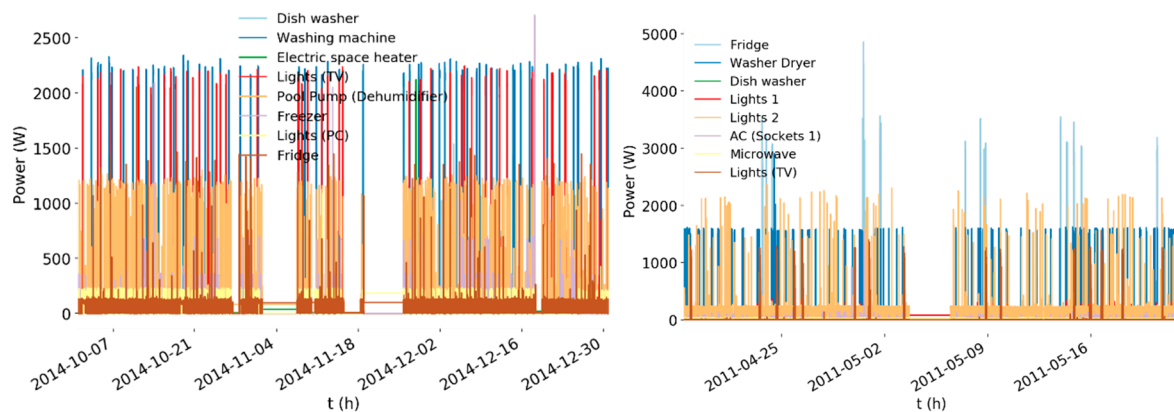


Figure 4. Disaggregated Training data for defined time windows: Refit dataset using combinatorial optimisation (CO) (left), and Residential Energy Disaggregation Dataset (REDD) dataset using Factorial Hidden Markov Model (FHMM) (right).

Figure 5 shows the aggregated data to be tested on the disaggregation and flexibility models. The days are a part of the datasets under study, so that the results could be known and a test could be applied. In practice, this data would be a forecast of the load consumption for a given house, which is typically done across the electricity sector based on weather conditions, constraints, and foreseen interventions, and past data (weekdays, weekends, and also accounting for seasonal changes).

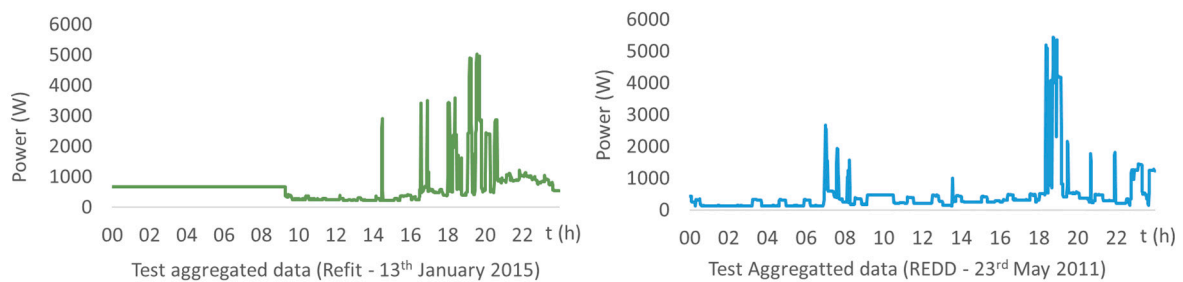


Figure 5. Aggregated load diagram for testing using Refit (left) and REDD (right) datasets.

4. Simulation Results

Both datasets were submitted to the CO and FHMM algorithms, and having defined the training and test windows the following three metrics provided by the NilmTK library were analyzed.

The normalized error in assigned power is given by the mean normalized error power (MNEP) metric, which helps to identify outliers. It is given by the sum of the differences between the assigned power and actual power of an appliance n in each time slice t , normalized by the appliance's total energy consumption, as per Equation (1):

$$\frac{\sum_t |y_t^{(n)} - \hat{y}_t^{(n)}|}{\sum_t y_t^{(n)}} \quad (6)$$

The root mean square error of the power (RMS_error_power) is the standard deviation of the residuals (prediction errors) and is a measure of how spread-out these residuals are. It shows how concentrated the data is around the regression line. This metric is commonly used in climatology,

forecasting, and regression analysis. It is given by the root mean square error between the assigned power and the actual power of appliances n in each time slice t , as explained by Equation (2):

$$\sqrt{\frac{1}{T} \sum_t (y_t^{(n)} - \hat{y}_t^{(n)})^2} \quad (7)$$

The F1 score is the weighted average of each precision and recall observation, taking into account both false positives and false negatives, and its value ranges from 0 (worst) to 1 (best). It is the fraction of time units in which an appliance was predicted to be *On*, but that it was actually *Off* (a false positive or precision), and the fraction of time units in which an appliance was predicted to be *On* that it was actually *On* (true positive or recall). This estimation is also known as the harmonic mean and is given by Equation (3):

$$F1_{score} = \frac{2 \times Precision \cdot Recall}{Precision + Recall} \quad (8)$$

Table 4 shows the results per appliance according to each dataset studied. Those devices that have more than one phase or circuit connected to its designation, such as the case for “Lights”, were aggregated per category in terms of power and treated as one for flexibility calculations.

Table 4. Accuracy metrics for both datasets using F1 score, root mean square (RMS) error power and mean normalized error power (NEP) metrics.

| REFIT Dataset: Train, October 10, 2014, to December 31, 2014; Test, January 1, 2015, to January 30, 2015. | | | | | | |
|---|----------|-------|-------|-------|-----------------|--------|
| Building 15 | F1 Score | | MNEP | | RMS_error_power | |
| Appliances | CO | FHMM | CO | FHMM | CO | FHMM |
| Fridge | 0.55 | 0.49 | 2.59 | 0.97 | 186.12 | 63.63 |
| Freezer | 0.32 | 0.27 | 2.19 | 3.5 | 343.69 | 100.6 |
| Electric space heater | 0.01 | 0.01 | 550 | 503 | 63.95 | 240 |
| Washing machine | 0.06 | 0.04 | 1.92 | 3.7 | 347.1 | 379 |
| Dish washer | 0.06 | 0.044 | 2.84 | 2.25 | 267.66 | 359 |
| Lights (PC) | 0.17 | 0.171 | 34 | 5.34 | 171.47 | 26.5 |
| Lights (TV) | 0.53 | 0.51 | 1.49 | 1.18 | 101.07 | 106 |
| Pool Pump/Dehumidifier | 0.2 | 0.25 | 2.27 | 3.86 | 338.94 | 239.72 |
| REDD Dataset: Train, April 18, 2011, to May 21, 2011; Test, May 21, 2011, to May 24, 2011. | | | | | | |
| Building 1 | F1 Score | | MNEP | | RMS_error_power | |
| Appliances | CO | FHMM | CO | FHMM | CO | FHMM |
| Fridge | 0.45 | 0.47 | 1.11 | 1.06 | 97.99 | 112.72 |
| Dish Washer | 0.23 | 0.15 | 2.41 | 1.75 | 373.23 | 268.62 |
| AC system (Socket 1) | 0.74 | 1 | 1.96 | 0.076 | 46.55 | 2.29 |
| Heating System (Socket 2) | 0.15 | 1 | 2.26 | 0.6 | 205.43 | 27 |
| Light 1 | 0.62 | 0.7 | 1.05 | 2.04 | 57.99 | 102.8 |
| Microwaves | 0.05 | 0.1 | 3.26 | 3.75 | 213.33 | 323 |
| Lights 2 | 0.6 | 0.564 | 0.97 | 0.71 | 34.87 | 37.48 |
| Washer Dryer | 0.01 | 0.111 | 15.29 | 0.996 | 91.38 | 303 |

The datasets capture large time windows, however in order to keep the training time to manageable length, the Refit dataset was reduced. However, it still included approximately 10 million observations, as it was recorded in a 1 second time step referring to four months. The split was performed at 75% and 25% between training and test. The training times for the Refit dataset were 39.20 and 9.22 seconds, corresponding to the FHMM and CO algorithms, respectively. Regarding the dropout ratio, which provides the rate at which packets were lost, building 15 reported a dropout rate of 0.0277. Both algorithms present high errors for the estimations. This may be due to the selection of *top k* training appliances, which when optimized do not match the aggregated data observed by the model. The choice of selecting the top appliances may improve the training speed but may reduce accuracy as a trade-off. To improve the accuracy, a selection of the top appliances for the training data should match the aggregated training data. The only mitigation action taken in this study was to reduce processing time, considering the top 8 appliances of each dataset, however, there were more appliances contributing to the aggregated data, hence contributing to a higher error. In fact, the choice to train

the model could be done on 2 or 3 assets, but for the sake of the exercise and to show how it can be applied to a broader range of appliances, we chose 8. Diversity in appliances can introduce significant variations in signatures, leading to significantly different flexibility estimates.

Regarding the REDD dataset, all observations corresponded to 36 days for building 1 and were all taken into consideration to run the model. This corresponds to 3.1 million observations, since also a 1-second time step was used to record the data. The training times for the REDD dataset were 15.18 and 1.03 seconds, corresponding to the FHMM and CO algorithms, respectively. In this dataset, building 1 reported a dropout rate of 0.1958.

Generally, results for both datasets provided medium to low accuracy and high errors, especially appliances that have more reduced times of usage, such as the microwave. These appliances are, however, not flexible and will not be used in the flexibility forecast. Flexible appliances identified were the freezers and refrigerators, heating systems, lights, and pool pumps. In the Refit dataset, both computer and television usage were considered as lights, as according to the dataset explanation these items could have broader applications. A possible explanation for low accuracy could come from the high dropout rate, specifically in the REDD dataset, which accounts for incorrect data being trained into the model. To improve the accuracy of the models, complete data sets should carefully be retrieved in future projects. This is particularly important for flexibility forecast models, which by themselves add more uncertainty on top of these errors. Figure 6 presents the test and prediction disaggregated data. The time step of the test (dataset) is 3 sec, whereas the prediction for processing reasons was set to 60 sec. For this reason there is a higher level of detail observed in the test plot than in the prediction one. However, one can observe a general consistency in the amplitude and times of use of the appliances.

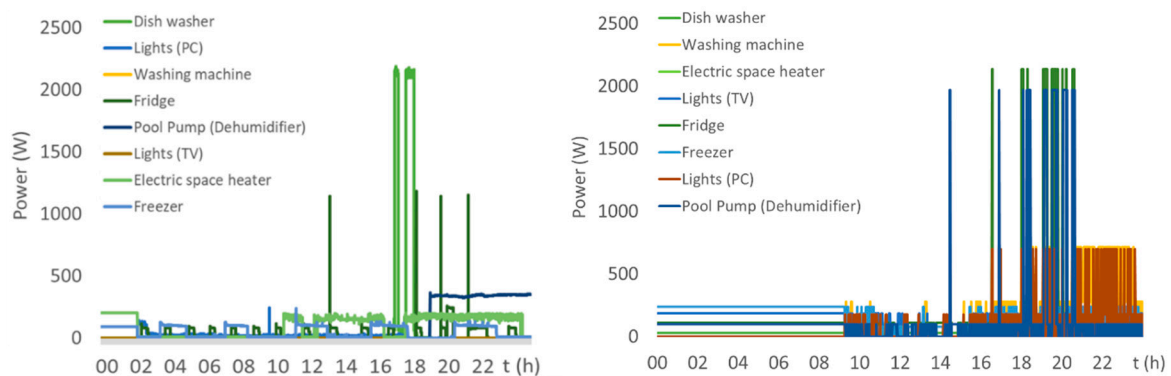


Figure 6. Disaggregated test (left) and prediction (right) for Refit Dataset using CO.

Visualizing the complete test data makes the interpretation somehow imprecise, as it only allows us to see a similar overall trend. Figure 6, however, shows that the total predicted power matches the prediction from the test, and the same can be said for the amplitudes across the appliances. To be able to individually analyze a test and prediction appliance, Figure 7 shows a one-day load diagram using the REDD dataset as an example, with only one appliance (fridge). An overall match can be seen between maximum power reached and a generation correspondence on the duration of the on and off cycles though out the day.

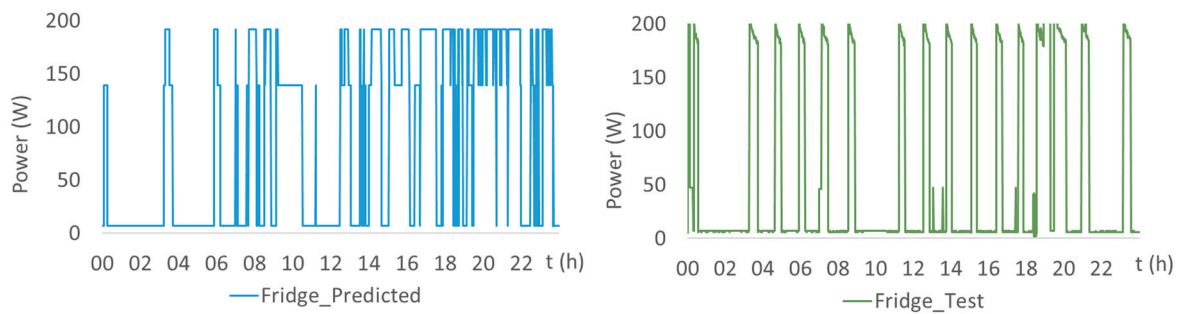


Figure 7. Prediction (**left**) and test (**right**) example for a fridge load on the May 23 (REDD dataset).

There is, however, a mismatch in the number of cycles between the two, which is higher in the predicted values than in the test ones. Such mismatches will be observable in the error analysis, contributing to a decrease in the accuracy of the model.

According to Table 1, the STOR program, which is the one being considered for the building participation, allows a maximum of 2 h participation for each asset. Given the daily load profiles of the test for both datasets to be tested and disaggregated, we considered the morning and afternoon periods, where higher electricity use is observed, which are from 06:00 to 08:00 and from 18:00 to 20:00, as can be seen in Figures 8 and 9 for the REDD and Refit datasets, respectively. It is worth mentioning that different programs may have their # recuperation asset times (e.g., 1 day). This means that if an asset is called to participate in an event, it will not be called throughout the day again to participate in DR, to allow any effect caused by the shift or flexibility to be restored. Results show a flexibility maximum power of 200–245 W and 180–500 W for the REDD and Refit datasets, respectively. The stability of the power is, however, cumbersome, as by itself it does not allow a solid bid to be made, even in a minimum time slot of 15 minutes, except for the morning period in Figure 9 (left).

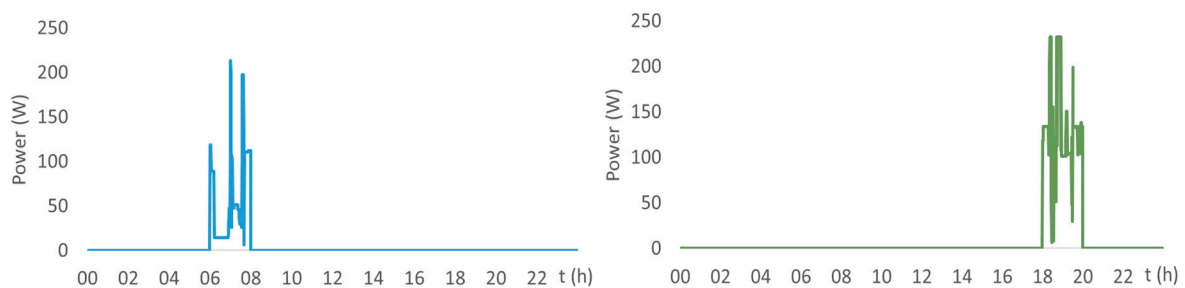


Figure 8. Flexibility Forecast Diagram for REDD dataset for the building 1 morning period from 06:00 to 08:00 (**left**), and the afternoon period from 18:00 to 20:00 (**right**).

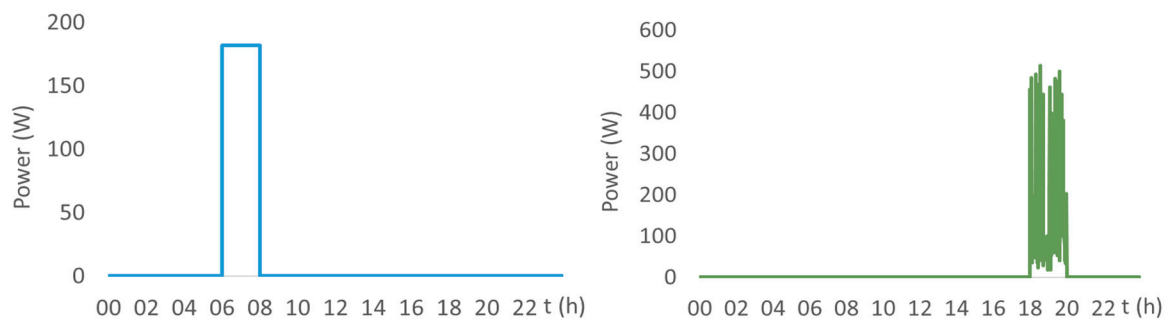


Figure 9. Flexibility Forecast Diagram for REFIT dataset for the building 15 morning period from 06:00 to 08:00 (**left**), and the afternoon period from 18:00 to 20:00 (**right**).

5. Discussion

The accuracies of the NilmTK disaggregation of the test sample for each of the houses tested are generally moderate to low. They are, however, in line with the values found in the literature [25,41]. This uncertainty added to the already existing load forecast uncertainty, which may result in low overall accuracies and lead to failures in delivering committed power to a DR event, meaning penalties. All efforts should be made to reduce the overall error. This may come from higher accuracy measurements from the high frequency meters, targeting specifically flexible devices, or by applying other disaggregation algorithms that could provide higher scores. Furthermore, smaller assets may be included in large power asset contracts to reduce overall uncertainty.

From the flexibility forecasts shown in Figures 8 and 9, it can also be observed that the power is not stable due to the nature of the consumption. This may pose some concerns in securing a stable service when bidding in the market. Another challenge is attaining a sufficient level of power in order to respect the minimum bid of the market, which may reach an order of magnitude of 3 MW, depending on the program. Such challenges may, however, be mitigated with the aggregation of flexibilities, offering in this way a sufficiently stable power obtained from a resulting “base load flexibility” forecast line. Moreover, since the forecast model is based on historical data, the algorithm should include a mechanism to “re-train” the model periodically in order to incorporate changes in consumption patterns (seasonal or occupancy), as foreseen in Figure 2. This would mean that new data sets would have to be retrieved from individual appliances periodically. This cost should be considered in the system when assessing the benefits of such DR services. If a new appliance is added to the system it enters an “unknown” category and is considered non-flexible or “shiftable”. This will happen until the model is trained again to recognize that new appliance.

While implicit DR programs have shown that most customers benefit financially from dynamic tariffs, it is unclear if the benefits from explicit DR services could incentivize participation. Smart contracts or similar distributed ledger technologies of automatized commitments and financial settlements could play a major role in enabling small and medium DR participants. The use of costless, permissionless, and scalable solutions are desirable, so that micro payments can be possible and worth pursuing. Otherwise, the complexity versus potential rewards will make their success doubtful. The ultimate enabler of such DR services will be achieved by standards such as Open ADR [42]. Such standards must promote reliable, scalable, and secure message exchange of events and prices to a vast range of equipment between customers and aggregators. It will also allow the triggering of smart contracts to replicate and generate millions of event requests, truly unleashing the potential of decentralized energy systems and citizen engagement in energy management.

Further Research

Having presented the model, the next steps should be to make it scalable and aligned with the project it was developed for (Delta) [6]. The Delta project pilots are under preparation and this work was useful to understand the limitations, accuracy levels, and mitigation actions. Pilot meters should focus on collecting data only targeting circuits or potential flexible assets and retrieving the corresponding aggregated data, as unnecessary assets will bring noise to the model. In the Delta pilots, both household assets and building level assets will be monitored (centralized chillers, boilers), which means that a mixture of single phase and three phase circuits will be monitored. This reality will have to be incorporated in the model, since currently it only includes single-phase appliances. Moreover, the reaction time of the assets to provide flexibility could be a topic to be investigated, as it could shed some light on which programs could be used besides the STOR program. Finally, the clustering and aggregation of other buildings will be useful next steps of the study. This will allow a better understanding of what the resulting aggregated base flexibility is. This base would be the basis of the offer to the balance market, and could help estimate a range of buildings to be aggregated in order to provide an offer complying with the minimum bid requirements.

6. Conclusion

The study presents the methodology to estimate the load flexibility of a household for DR services. Two datasets are used, namely the REDD and Refit datasets, applied to the Non-Intrusive Load Monitoring (NILM) in order to derive disaggregated data per appliance from smart meter aggregated measurements. Two approaches are used, the Combinatorial Optimization and the Factorial Hidden Markov model for the disaggregation. In order to derive the flexibility available from each device, the devices themselves were categorized as shifted but not varied, shifted and varied, and not shifted but varied. The top 8 devices were taken into account. A thorough literature review was done in order to derive the potential flexibility rate, and the mean time duration that a specific device can be curtailed. To estimate the flexibility in realistic terms and the characteristics of the short-term operating reserve, the STOR program based on the U.K. balance market was considered. Thus, a 2 h interval was used as the time window for the estimated flexibility integration. Simulations were carried out accounting for all of the above parameters, and results showed that flexibility of houses under testing had maximum power values of 200–245 W and 180–500 W for the REDD and Refit datasets, respectively.

Given the individual low levels of flexible power that could be offered to the market, it is unlikely that the financial compensation from explicit DR services could incentivize participation. Other forms of benefits should complement the financial incentives to attract small consumers. Eventual infrastructure costs should be reduced as much as possible in order to allow the already small financial benefits of DR services to exist. The simulations show that the houses tested alone present insufficient power steadiness to serve as a bid into an electricity market. Hence, in future work the application of the methodology to more than one house or a set of buildings is advised, in order to estimate a stable flexible power baseline that could be valid for market purposes. In general, the methodology is considered as a structured way to calculate flexibility that can be extracted by one house, which gives a valuable tool to aggregators as the first step to bring demand response to the residential sector.

Author Contributions: Conceptualisation and methodology, A.L. and L.J.; software, A.L.; validation, A.L.; investigation, A.L. and L.J. writing—original draft preparation, A.L.; writing—review and editing, A.L., N.A., E.K., and M.M.; supervision, E.K. and M.M.

Funding: This research was developed within the DELTA project, which received funding from the European Union's Horizon 2020 research and innovation programme, under grant agreement No. 773960.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bertoldi, P.; Zancanella, P.; Boza-Kiss, B. *Demand Response Status in EU Member States*; JRC Report; EUR 27998 Joint Research Centre: Luxembourg; Publications Office of the European Union: Brussels, Belgium, 2016; p. 153.
2. Kiwi Power. Available online: <https://www.kiwipowered.com/> (accessed on 16 May 2019).
3. Bertoldi, P.; Hirl, B.; Labanca, N. *Energy Efficiency Status Report 2012*; JRC Report; EUR 25405 Joint Research Centre: Luxembourg; Publications Office of the European Union: Ispra, Italy, 2012; p. 143.
4. Dr. BOB. Available online: <https://www.dr-bob.eu/> (accessed on 8 May 2019).
5. Drimpac. Available online: <http://www.drimpac-h2020.eu/> (accessed on 14 May 2019).
6. Delta. Available online: <https://www.delta-h2020.eu/> (accessed on 8 May 2019).
7. Drive. Available online: <https://www.h2020-drive.eu/> (accessed on 8 May 2019).
8. Respond. Available online: <http://project-respond.eu/> (accessed on 8 May 2019).
9. FlexCoop. Available online: <http://www.flexcoop.eu/> (accessed on 8 May 2019).
10. eDream. Available online: <https://edream-h2020.eu/> (accessed on 8 May 2019).
11. NilmTK. Available online: <https://nilmtk.github.io/> (accessed on 16 May 2019).
12. Murray, D. An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. *Sci. Data* **2017**, *122*, 160122. [[CrossRef](#)] [[PubMed](#)]
13. Esa, N.F.; Abdullah, M.P.; Hassan, M.Y. A review disaggregation method in Non-intrusive Appliance Load Monitoring. *Renew. Sustain. Energy Rev.* **2016**, *66*, 163–173. [[CrossRef](#)]

14. Parson, O. Dataport and NILMTK: A Building Data Set Designed for Non-intrusive Load Monitoring. In Proceedings of the 2015 IEEE Global Conference on Signal and Information Processing, Orlando, FL, USA, 14–16 December 2015.
15. Kelly, J. Demo Abstract: NILMTK v0.2: A Non-intrusive Load Monitoring Toolkit for Large Scale Data Sets. In Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, New York, NY, USA, 3–6 November 2014.
16. Naoki, K. *Combinatorial Optimization Algorithms in Resource Allocation Problems Christodou*; Springer: Boston, MA, USA; Kyoto, Japan, 2008.
17. Riesenhuber, M.; Poggio, T. Factorial Hidden Markov Models. Ghahramani, Zoubin Jordan, Michael I. *Adv. Neural. Inf. Process. Syst.* **1996**, AIM-1695, 188–193.
18. Hart George, W. Nonintrusive appliance load monitoring system. *EPRI J.* **1990**, *80*, 45–47.
19. Rossi, J.R. *Mathematical Statistics: An Introduction to Likelihood Based Inference*; International Statistical Review; John Wiley & Sons: Hoboken, NJ, USA, 2018.
20. Zhong, M.; Goddard, N. Latent Bayesian melding for integrating individual and population models. *arXiv* **2005**, arXiv:1510.09130.
21. Kelly, J.; Gate, Q.; Knottenbelt, W. Neural NILM: Deep Neural Networks Applied to Energy Disaggregation. In Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, Seoul, Korea, 4–5 November 2015; p. 10.
22. Zoha, A.; Gluhak, A.; Imran, M.A.; Rajasegarar, S. Non-intrusive Load Monitoring approaches for disaggregated energy sensing. *Surv. Sens.* **2012**, *12*, 16838–16866. [[CrossRef](#)] [[PubMed](#)]
23. Scikit-Learn. Available online: <https://scikit-learn.org/stable/> (accessed on 17 May 2019).
24. Kolter, J.Z.; Johnson, M.J. REDD: A Public Data Set for Energy Disaggregation Research. In *Workshop on Data Mining Applications in Sustainability (SIGKDD)*; Citeseer: San Diego, CA, USA, 2011; pp. 1–6.
25. Fagiani, M.; Bonfigli, R.; Principi, E.; Squartini, S.; Mandolini, L. A Non-Intrusive Load Monitoring Algorithm Based on Non-Uniform Sampling of Power Data and Deep Neural Networks. *Energies* **2019**, *12*, 1371. [[CrossRef](#)]
26. Stankovic, L.; Stankovic, V.; Liao, J. Measuring the energy intensity of domestic activities from smart meter data. *Appl. Energy* **2016**, *183*, 1565–1580. [[CrossRef](#)]
27. Kwon, P.S.; Østergaard, P. Assessment and evaluation of flexible demand in a Danish future energy scenario. *Appl. Energy* **2014**, *134*, 309–320. [[CrossRef](#)]
28. Ji, Y.; Rajagopal, R. Demand and flexibility of residential appliances: An empirical analysis. In Proceedings of the 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, QC, Canada, 14–16 November 2017.
29. Mocanu, E.; Nguyen, P.H.; Gibescu, M. Energy disaggregation for real-time building flexibility detection. In Proceedings of the 2016 IEEE Power and Energy Society General Meeting, Boston, MA, USA, 17–21 July 2016; pp. 1–5.
30. Li, P.H.; Pye, S. Assessing the benefits of demand-side flexibility in residential and transport sectors from an integrated energy systems perspective. *Appl. Energy* **2018**, *228*, 965–979. [[CrossRef](#)]
31. Stötzer, M.; Gronstedt, P.; Task, E.T.G. 21st International Conference on Electricity Distribution demand side management potential a case study for germany. *Energy* **2011**, *958*, 6–9.
32. D’hulst, R.; Deconinck, G.; Claessens, S.; Labeeuw, W.; Vanthournout, K.; Beusen, B. Demandresponse flexibility and flexibility potential of residential smart appliances: Experiences from large pilot test in Belgium. *Appl. Energy* **2015**, *155*, 79–90. [[CrossRef](#)]
33. Reynders, G.; Lopes, R.A.; Marszal-Pomianowska, A.; Aelenei, D.; Martins, J.; Saelens, D. Energy flexible buildings: An evaluation of definitions and quantification methodologies applied to thermal storage. *Energy Build.* **2018**, *166*, 372–390. [[CrossRef](#)]
34. Six, D.; Desmedt, J.; Bael, J.V.A.N.; Vanhoudt, D. Exploring the flexibility potential of residential heat pumps combined with thermal energy storage for smart grids. In Proceedings of the 21st International Conference on Electricity Distribution, Frankfurt, Germany, 6–9 June 2011; pp. 6–9.
35. Nuytten, T.; Claessens, B.; Paredis, K.; van Bael, J.; Six, D. Flexibility of a combined heat and power system with thermal energy storage for district heating. *Appl. Energy* **2013**, *104*, 583–591. [[CrossRef](#)]
36. Hong, J.; Kelly, N.J.; Richardson, I.; Thomson, M. Assessing heat pumps as flexible load. *Proc. Inst. Mech. Eng. Part A J. Power Energy* **2013**, *227*, 30–42. [[CrossRef](#)]

37. Saha, A.; Kuzlu, M.; Pipattanasomporn, M. Demonstration of a home energy management system with smart thermostat control. In Proceedings of the IEEE PES Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 24–27 February 2013.
38. Stadler, I. Demand Response—Nichtelektrische Speicher für Elektrizitätsversorgungssysteme mit Hohem Anteil Erneuerbarer Energien. Master's Thesis, Kassel University, Kassel, Germany, October 2005.
39. Lund, P.D.; Lindgren, J.; Mikkola, J.; Salpakari, J. Review of energy system flexibility measures to enable high levels of variable renewable electricity. *Renew. Sustain. Energy Rev.* **2015**, *45*, 785–807. [[CrossRef](#)]
40. Lindsey, D.; Sharp, F.; Abear, T. The Inconsistency of Dimmers with LED Lamps. In Proceedings of the ACEEE Summer Study Energy Efficiency in Buildings, Pacific Grove, CA, USA, 21–26 August 2016.
41. Ponocko, J.; Milanovic, J.V. Forecasting Demand Flexibility of Aggregated Residential Load Using Smart Meter Data. *IEEE Trans. Power Syst.* **2018**, *33*, 5446–5455. [[CrossRef](#)]
42. Open ADR. Available online: <https://www.openadr.org/> (accessed on 17 May 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).