Demand Response Alert Service Based on Appliance Modeling

Ioanna-M. Chatzigeorgiou [1,*]**, Christos Diou** [2]**, Kyriakos C. Chatzidimitriou** [1] **and Georgios T. Andreou** [1]

[1] School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; kyrcha@ece.auth.gr (K.C.C.); gandreou@ece.auth.gr (G.T.A.)

[2] Department of Informatics and Telematics, Harokopio University of Athens, 17778 Athens, Greece; cdiou@hua.gr

[*] Correspondence: ioannamy@ece.auth.gr

Abstract: Demand response has been widely developed during recent years to increase efficiency and decrease the cost in the electric power sector by shifting energy use, smoothening the load curve, and thus ensuring benefits for all participating parties. This paper introduces a Demand Response Alert Service (DRAS) that can optimize the interaction between the energy industry parties and end users by sending the minimum number of relatable alerts to satisfy the transformation of the load curve. The service creates appliance models for certain deferrable appliances based on past-usage measurements and prioritizes households according to the probability of the use of their appliances. Several variations of the appliance model are examined with respect to the probabilistic association of appliance usage on different days. The service is evaluated for a peak-shaving scenario when either one or more appliances per household are involved. The results demonstrate a significant improvement compared to a random selection of end users, thus promising increased participation and engagement. Indicatively, in terms of the Area Under the Curve (AUC) index, the proposed method achieves, in all the studied scenarios, an improvement ranging between 41.33% and 64.64% compared to the baseline scenario. In terms of the F_1 score index, the respective improvement reaches up to 221.05%.

Keywords: data analytics; artificial intelligence applied to power systems; demand side management; flexibility; demand response; appliance modeling; peak shaving; smart grid

1. Introduction

1.1. Background

A series of measures have been suggested over recent years for the demand side of Power Systems (PSs), aiming to balance the dynamic aspect of Renewable Energy Sources (RESs) introduced in their production side as well as decrease the respective costs and increase the overall system efficiency. These measures comprise what is widely referred to as Demand Side Management (DSM) [1]. One of the main DSM objectives is the transformation of load curves, through the utilization of different strategies, such as peak shaving, strategic conservation or load growth, valley filling, load shifting, or flexible load shape [2]. An important concept within the context of DSM is Demand Response (DR) [3]. DR includes the provision of control signals to the demand side to shift energy usage at important times, e.g., times of extremely high demand, high non-dispatchable production, or when grid reliability is jeopardized. DR can potentially add stability to the system and decrease the need for coal- and gas-fired spinning reserves as well as decrease the operating costs and wholesale prices.

DR programs are either incentive-based or price-based [4] (also referred to as time-based [1]). In incentive-based programs, users are rewarded for changes in energy use compared to a baseline level, while in price-based programs varying prices are used to promote changes in energy use [5]. Incentive-based DR includes Direct Load Control (DLC), Emergency DR (EDR), Interruptible/Curtailable (I/C) service, Capacity Market (CAP),

Demand Bidding/Buyback (DB), and Ancillary Service (AS) market clearing programs. On the other hand, price-based DR includes Time of Use (TOU), dynamic TOU, and Real-Time Pricing (RTP) programs. Combinations or variations of these programs have also been developed.

DR has been mostly applied until now in the industrial sector. However, there has been a rising interest regarding the DR potential in the residential sector. The US energy markets demonstrate an increasing number of both time- and incentive-based residential DR programs during recent years [6]. The implementation of these DR programs in the USA is facilitated by a constantly increasing Advanced Meter (AM is also referred to as a Smart Meter (SM)) penetration rate in the residential sector (57% in 2020) as well as a series of regulatory and legislative changes. In the EU, DR is facing a series of policy and regulatory obstacles. Several EU member states have not yet seriously engaged with DR reforms, and their regulatory structures do not yet actively support the participation of the demand side in the markets [7]. In addition, there is a serious delay in the penetration of SMs in residential installations (expected to be 80% by 2020 but currently reaching only 43%). Furthermore, a lack of interoperability has also been highlighted regarding the different types of AM already installed [8].

Residential DR programs are implemented either through automated control systems or manual shifting by the end users. In Automated Demand Response (ADR) programs, an Energy Consumption Scheduler (ECS) [9] or a Home Energy Management System (HEMS) [10] is necessary as well as a SM installation to optimize the respective appliances' schedules, based on different restrictions, optimization goals, and pricing signals. In some ADR programs, users may override such a control. ADR is developing slowly due to the high costs of the required infrastructure. Another important open issue seems to be the end-users' perceived loss of privacy or control due to automation [11].

In manual DR, end users are expected to shift their consumption following an incentive or a price signal. The end-users' response to a DR event in such programs depends heavily on their demographics [12] as well as their knowledge and perception of their energy use, which is often limited or incorrect [13]. A potential drawback of manual DR is that if end users receive the same incentives or price signals they could shift their appliances in a similar way and create new peaks in their total consumption [14].

EDR comprises a potentially promising manual DR program type for peak shaving. Such an opt-out pilot program in 2015 in California, USA [15] enrolled 74,900 households and used personalized messaging based on neighbor comparison to reduce peak demand during four event days called "Summer Saving Days", with no monetary incentives. The decrease in peak load demand declined from about 2.2% on the first event to about 1.2% on the fourth event. As noted by the authors, "this decline over time is approximately linear and might result from a coincidence of other temporal factors with the performance of the program (e.g., weather) or it might result from the fact that the impact of DR messaging weakens with repeated exposure". A similar approach that included monetary incentives was adopted for a pilot study in China in 2017 with 20,000 enrollments that were invited via a SM on certain days to participate in a 1.5 h peak reduction [16]. The households replying "YES" were considered policy-responding households and received corresponding monetary subsidies based on the electricity savings during the response period. The pilot study reported that the effective participation was 11%, and the average household electricity savings were an average 0.86 kWh during the peak period. "Defeat the Peak" was another peak-shaving program implemented in the USA in 2018, with 16,000 participating households, involving a collective charity donation as the end-users' incentive [17]. Six peak events were set throughout the program and consumers were asked to curtail energy use during specific time periods (3 or 4 h in the afternoon). However, significant savings were reported only when the event coincided with the day of the respective annual peak demand.

1.2. Related Work

In all the above programs, alerts included the straightforward recommendation that users should reduce energy use during a DR event. Pratt and Erickson suggested that more specific recommendations could have increased the impact of their program [17]. Moreover, as mentioned earlier, it is reported in [15] that users' fatigue due to frequent or unrelatable DR recommendations is another factor that may affect participation during DR events. However, the creation of specific, personalized recommendations requires the effective modeling of energy use.

Bottom-up end-use modeling has been widely used during recent years as a potential part of effective DR prediction and semi or full ADR rescheduling. Variations of Markov models [18,19], neural networks [20], classifiers [21], and event-based approaches have been used in the literature [22–25] to acquire appliance models that can successfully estimate their energy use or potential flexibility.

Another approach to this end is to carry out appliance-level power consumption forecasting using different machine learning methods, e.g., deep learning [26,27] or Conditional Hidden Semi-Markov Models (CHSMMs) [28]. The forecasts in these papers, though, are either short term (one time interval ahead, be that a minute or an hour) or make use of external features other than the power consumption (humidity, precipitation, temperature, cloud cover, etc.) [29]. We differentiate in this aspect as we do next-day forecasting of the ON/OFF states of appliances, while, in addition, our resulting model is explainable and thus easier to be examined by an expert for its validity.

An approach that is a closer match to our method is that of [30] that provides behavioral energy consumption pattern mining for appliance–appliance and appliance–time associations derived from smart meter data and uses association rules and clustering to provide insight into consumers' energy consumption decision patterns. The method, though, needs to be amended with the mining of patterns that include both the type of day and the time of day along with a way to rank the households for receiving a personalized recommendation.

Finally, probability distributions of appliance use can be used as preference indicators in appliance scheduling optimization problems. For example, in [31] the hourly energy consumption curve for an appliance type is transformed into a probability distribution by converting the average energy consumption of appliances (per household) to the appliance use distribution using histogram construction algorithms. Such an approach, though, is too simplistic and as we will see in Section 3.4 less performant.

Huber et al. [24] went one step further by demonstrating a first comparison between a histogram, a pattern search, and a Bayesian inference algorithm for the modeling of appliance use. These models, however, are aimed toward the effective modeling of a single end user. In a DR implementation, an aggregator has to identify the part of a consumer group that will most probably respond to a DR request. In this context, the end-use modeling problem becomes rather a problem of probability ranking within this consumer group.

1.3. Contribution

For these reasons, in [25] a targeted messaging mechanism was presented that aims to improve DR communication services by suggesting appliance-based DR actions to the end users during a peak DR event, based on their past-usage appliance information. To this purpose, it is assumed that data regarding the past energy use are available through a SM and appliance data are available either via Non-Intrusive Load Monitoring (NILM) algorithms or via smart plugs. As a first step, appliance profiles are created for each major DR appliance and end user based on past-usage data. Only the users with the highest expected consumption regarding this appliance during the examined peak period are selected to receive the DR alert. This paper extends this work by testing the efficiency of three different appliance models, in a single-appliance and a multi-appliance scenario, based on two different publicly available datasets and a variety of appliance types. In the

multi-appliance scenario, the end users also receive a specific recommendation on which appliance to turn off. The contributions of this work are therefore as follows:

- Different variations for a probabilistic data-driven appliance model that permits us to rank our users based on their past appliance usage are presented;
- In the multi-appliance scenario, appliance-based personalized DR alerts are offered, which can help end users to choose the right actions for reducing energy use during the DR event;
- Overall, an improved data-driven appliance-based DRAS to increase participation in manual DR programs is introduced;
- The DRAS is evaluated on real-world data and on different appliances and households.

The process for the creation of appliance models is presented in Section 2. The evaluation on the real-world data is presented in Section 3, which is followed by the Discussion and Conclusions sections.

2. Method

The approach is based on individual major household appliance measurements, which are used for the development of probabilistic appliance use models. For the purposes of this work, the required measurements are acquired from publicly available residential electrical energy datasets. A training period is set, and appliance models are created based on the appliance usage data during this period. Different single- and multi-appliance scenarios are examined. The part of the dataset that does not belong in the training period is used for the evaluation of the DRAS.

2.1. Appliances

When analyzing the potential shifting of residential load for DR it is important to recognize and categorize residential appliances in terms of their DR potential based on the following aspects: contribution to total load, flexibility over time, and variability over energy use [32]. Electric Vehicles (EVs) can serve as ideal DR appliances since their recharging is energy demanding, interruptible, and can be charged at different charging rates. For this reason, in the US markets, specific rates and rate programs for electric vehicles have been introduced during recent years. Another popular solution in the US market is the introduction of smart thermostats and the respective DR programs for HVAC, pool pumps, and water heaters, which are all common loads among US residencies. Wet or washing appliances, i.e., washing machines, dishwashers, and tumble dryers, have also been widely used in market research as DR appliances [33]. These are deferrable in time but are not variable in terms of energy use. As reported by [34], wet appliances can play an important part in DR, particularly in Europe [35].

2.2. Appliance Modeling

This study builds upon an earlier work on a probabilistic appliance modeling scheme for DR alerts [25] by examining variations of the initial model and reassessing the approach for more single-appliance and multi-appliance scenarios.

Two states are assigned to each appliance (ON/OFF). The ON state is attributed to an appliance if the mean power during a 15 min period (a typical SM data frequency) exceeds a representative threshold. In this work, the threshold of 20 W (mean power during a 15 min period) was used without loss of generality, as it was sufficient in order to distinguish between the ON state of the studied appliances and the cases of measurement noise. One model is created for each appliance, as we make the modeling assumption that the use of each appliance is independent of the use of other appliances. The model aims at estimating the probability, $P(I(t_1, t_2)|d)$, that an appliance is active during the time interval, $I(t_1, t_2)$, between times t_1 and t_2 on a particular calendar day, d. To estimate $P(I(t_1, t_2)|d)$, we make a series of modeling approximations.

First, we model the probability, $P(D|d)$, of event D that the appliance is used on day d. Specifically, let $T(d)$ be the type of day (e.g., "Monday" or "Business Day"). Our model

assumes that there is a regular usage pattern such that the probability of the use of the appliance on d depends only on the type of day, i.e., $P(D|d) \approx P(D|T(d))$ written as $P(D|T)$ for notation simplicity. Types of days are further discussed in Section 2.4.

Assuming the appliance is used on d, the next step is to model the probability that the appliance is active during a specific time interval of the day. To do that, we first model the event $A(t)$, where the appliance is active at time t. This is true if the appliance started consuming power before t and its usage duration was long enough, such that it was active at t. If S is the usage start time and U the usage duration, the event $A(t)$ can be approximated by

$$\bigcup_{k=0}^{k_{max}} \{(S = t - k\Delta t) \cap (U \geq k\Delta t)\} \tag{1}$$

for sufficiently small durations, Δt, where $t_{max} = k_{max}\Delta t$ is the maximum appliance usage duration. Notice that the events $(S = t - k\Delta t) \cap (U \geq k\Delta t)$ are disjoint for different k. Then, assuming the duration of the usage of an appliance is independent of its usage start time, the probability of $A(t)$, where the appliance is active at time t of a usage day, is

$$
\begin{aligned}
(A(t)|D) &= \int_{s=0}^{t_{max}} f_S(t - s|D) P(U \geq s) ds \\
&= \int_{s=0}^{t_{max}} f_S(t - s|D) ds \int_{u=\tau}^{t_{max}} f_U(u) du P
\end{aligned} \tag{2}
$$

where $f_S(s|D)$ is the probability density function (PDF) of S, given that the appliance is used on that day, and $f_U(u)$ is the PDF of the duration, U.

If we work with discrete random variables and probability mass functions, the same calculation can be approximated through

$$P(A(t)|D) = \sum_{i=0}^{t_{max}} P(t - i - 1 < S \leq t - i|D) \; P(U \geq i) \tag{3}$$

Similarly, the event $I(t_1, t_2)$, where the appliance is active sometime during the interval $[t_1, t_2]$, is equivalent to the event

$$D \cap (A(t_1) \cup (t_1 \leq S \leq t_2)) \tag{4}$$

i.e., where the appliance either started during $[t_1, t_2]$ or it started sometime before t_1, but its usage duration was long enough so that the appliance was active at t_1 and afterward. Thus, the probability that the appliance is active during $[t_1, t_2]$ of day d is

$$P(I(t_1, t_2)|d) = P(D|T) \left(P(A(t_1)|D) + \int_{t_1}^{t_2} f_S(s|D) ds \right) \tag{5}$$

We have therefore defined the probability that we can use to rank consumers with a decreasing probability of use of the target appliance during the interval I, in terms of $P(D|T)$, $f_S(s|D)$, and $f_U(u)$. These distributions can be estimated from the available data, as described in the following section.

2.3. Estimating the PDFs

Given a dataset with active power consumption measurements, $P_a(t)$, for appliance a, we identify appliance activation events as intervals for which P_a exceeds a threshold, i.e., $P_a(t) > P_0$. In the experiments of Section 3, we set the threshold at 20W. Then, the probability $P(D|T)$ is the percentage of days of type T that include at least one consumption interval for a.
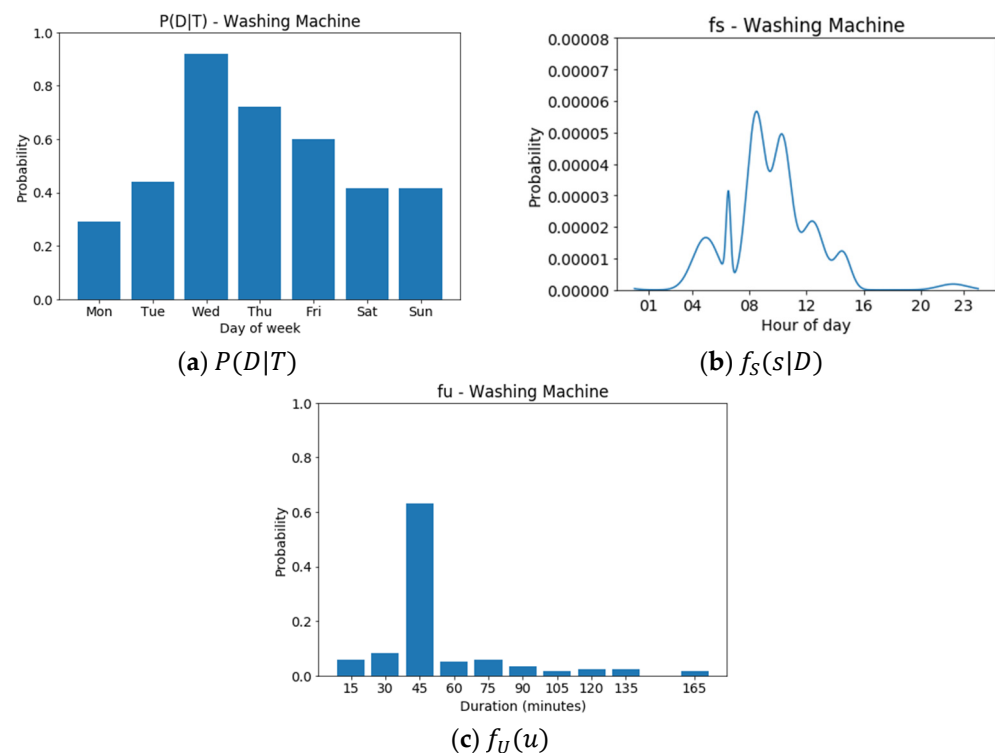
The PDF f_S is modeled as a univariate Gaussian Mixture Model (GMM):

$$f_S(s|D) = \sum_{k=1}^{K} a_k N(s; \mu_k, \sigma_k) \tag{6}$$

where s is the start time, K is the number of mixtures, $\sum_{k=1}^{K} a_k = 1$ are the mixture coefficients, and $N(s; \mu_k, \sigma_k)$ is a univariate Gaussian distribution with mean, μ_k, and standard deviation, σ_k. In our experiments, we set a fixed $K = 10$ for all appliances. The GMM parameters are estimated through the Expectation–Maximization (EM) algorithm on the start times of all the appliance events in the training set.

Finally, we approximate f_U via a probability mass function, which includes the probability of appliance usage duration in 15 min intervals (0–15 min, 15–30 min, etc.). This is also computed using the set of all activation events of the appliance in the training set. Figure 1a–c demonstrates indicative examples of the $P(D|T)$, the $f_S(s|D)$, and the $f_U(u)$ distributions for a washing machine.

(a) $P(D|T)$

(b) $f_S(s|D)$

(c) $f_U(u)$

Figure 1. Indicative examples of (**a**) the $P(D|T)$, (**b**) the $f_S(s|D)$, and (**c**) the $f_U(u)$ distributions for a washing machine.

2.4. Variations of the Model

One can consider many variations of this basic model, depending on the data volume available for training. For example, different options for the type of day, T, include using "workdays" vs. "non-workdays" or differentiating among the individual days of the week (Mon, Tue, Wed, Thu, Fri, Sat, and Sun). The former case assumes there is a regular usage pattern that is similar across business days and non-business days. The latter only makes this assumption for the same day of the week (e.g., Thursdays) and requires more data for training. Similarly, another option is to include the type of day in the calculation of the PDF f_S of the appliance usage start time (in the basic model it is independent of T).

Another option that has been explored in the experiments of this work is to include a variable, Y, denoting whether the appliance under study was used the previous day in the model. The use of Y affects the probability calculation regarding the use of the appliance on a particular day, i.e.,

$$P(D|d) = P(D|T, Y) \tag{7}$$

The distribution $P(D|T, Y)$ is calculated based on the activation events from the data.

2.5. Performance Metrics

As far as performance metrics are concerned, Wenninger et al. [21] successfully highlighted the weakness of accuracy as a metric to evaluate the results of appliance modeling for event-based algorithms. Pereira and Nunes have included in [35] an extensive list of performance metrics for event-based algorithms. Due to the nature of the problem at hand, precision, recall, F_1 score, and AUC were selected as evaluation metrics. In this study, precision and recall represent the following:
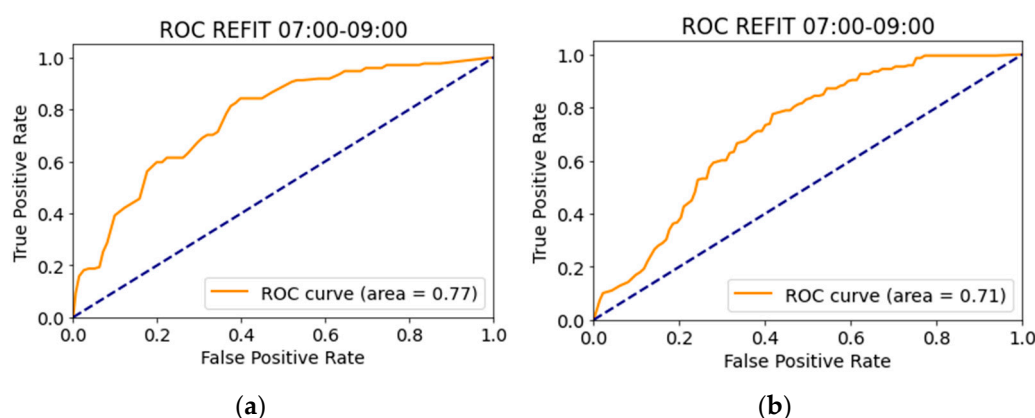
$$Precision = \frac{\text{appliances that were selected by the DRAS and turned ON during the DR event}}{\text{all appliances that were turned ON during the DR event}} \tag{8}$$

$$Recall = \frac{\text{appliances that were selected by the DRAS and turned ON during the DR event}}{\text{all appliances that were selected by the DRAS}} \tag{9}$$

Finally, Area Under the Curve (AUC) is the area under the Receiver Operating Characteristic (ROC) curve, which finds the algorithm configurations that have the best trade-off between recall and the False Positive Rate (FPR). The FPR represents the following:

$$FPR = \frac{\text{appliances that were selected by the DRAS but \textbf{not} turned ON during the DR event}}{\text{all appliances that were \textbf{not} turned ON during the DR event}} \tag{10}$$

Figure 2a,b demonstrates two indicative ROC curves.

Figure 2. ROC curve examples for DR scenarios concerning (**a**) dish washers (**b**) Tumble Dryers.

3. Results

3.1. Dataset

The DRAS was evaluated taking into account the available wet appliances (i.e., washing machines, dishwashers, and tumble dryers) in the REFIT [36] and the HES [37] datasets. The data were down sampled to a 15 min sampling period. The REFIT dataset included 20 washing machines, 15 dishwashers, and 12 tumble dryers. The HES dataset contained 22 washing machines, 16 dishwashers, and 11 tumble dryers. Each dataset was decreased to the minimum period including measurements for all houses (341 days for the REFIT dataset and 168 for the HES dataset). Subsequently, the dataset was divided into a training and an evaluation period. The training and evaluation periods were split 50/50.

3.2. Single-Appliance DR Scenario

In the single-appliance DR scenario, it is assumed that only one type of appliance per household participates in the DR program. The DRAS sends alerts to a pre-specified number of households (2, 4, 6, 8, and 10 out of 20 for the REFIT or 26 houses for the HES dataset) to avoid using this certain appliance.

The selection of the households that will receive an alert is performed based on the probability produced by the appliance use model that the specific appliance will be used in a household within the specific timeframe. This feature is tested in the evaluation periods of the considered datasets. To further test the model's ability to select the appropriate

households, a random selection process is also performed on the dataset evaluation periods (presented as an average over five different runs), and the respective results are compared to the proposed model results. In the following tables, suffix "_s" in the evaluation metrics denotes results obtained through selection by the proposed model, while suffix "_r" denotes results obtained by the random selection process. The different models used are presented in the following subsections.

3.2.1. Workdays versus Non-Workdays

The model assumes in this case that the probability of the use of an appliance on any given day depends only on whether it is a workday or a non-workday (weekends and holidays). This reasoning was evaluated on the wet appliances for both datasets. The results regarding the washing machines for both datasets are presented in Table 1. The results correspond to an evening peak period between 17:00 and 19:00.

Table 1. Workday/non-workday-single appliance (washing machines). Throughout the rest of the tables, the sign # refers to the pre-specified number of households in which the system sends the alerts.

Dataset	#	f1_s	f1_r	prec_s	prec_r	rec_s	rec_r	AUC_s	AUC_r
REFIT	2	0.305	0.095	0.325	0.105	0.288	0.088		0.507
	4	0.379	0.146	0.297	0.117	0.526	0.196		0.500
	6	0.367	0.171	0.252	0.119	0.671	0.302	0.770	0.501
	8	0.338	0.180	0.216	0.117	0.767	0.394		0.487
	10	0.318	0.182	0.195	0.113	0.863	0.475		0.505
HES	2	0.225	0.096	0.229	0.100	0.220	0.093		0.504
	4	0.279	0.131	0.212	0.101	0.407	0.185		0.510
	6	0.291	0.141	0.196	0.097	0.565	0.263	0.722	0.490
	8	0.285	0.158	0.179	0.100	0.689	0.368		0.501
	10	0.257	0.163	0.155	0.099	0.746	0.459		0.505

In all cases presented in Table 1, it is evident that the proposed model results are significantly better than the random selection process results. In terms of AUC, a metric that does not account for the number of selected households, the model results exhibit a 54% and 44% improvement in REFIT and HES, respectively, compared to the random selection process. As expected, as the number of alerts increases, recall increases, while precision decreases, striking a balance between four and six alerts based on the F_1 score. At the same time, the respective values of these metrics are expected to be lower for the HES dataset than for the REFIT dataset, as the former includes more households than the latter, while the magnitudes of the evaluated subsets are the same for both datasets.

3.2.2. Seven Day Types

The model assumes in this case that the probability of the use of an appliance on any given day depends only on the day of the week, i.e., Monday to Sunday. This model was evaluated on all wet appliances for both datasets. As in the previous case study, the following results correspond to an evening peak period between 17:00 and 19:00.

In general, the same observations can be made in Tables 2–4 as in the previous case. Moreover, in the case of the washing machines, the results are comparable (and in some cases slightly worse) than the ones in the previous scenario.

Table 2. Seven day-single appliance (washing machines).

Dataset	#	f1_s	f1_r	prec_s	prec_r	rec_s	rec_r	AUC_s	AUC_r
REFIT	2	0.291	0.099	0.310	0.108	0.275	0.092	0.776	0.501
	4	0.378	0.141	0.295	0.112	0.523	0.189		0.496
	6	0.371	0.169	0.255	0.118	0.679	0.297		0.497
	8	0.335	0.175	0.215	0.113	0.762	0.381		0.502
	10	0.303	0.187	0.186	0.116	0.824	0.486		0.507
HES	2	0.236	0.088	0.241	0.093	0.232	0.084	0.718	0.499
	4	0.279	0.130	0.212	0.100	0.407	0.183		0.522
	6	0.265	0.138	0.178	0.094	0.514	0.259		0.524
	8	0.261	0.156	0.165	0.099	0.633	0.364		0.512
	10	0.261	0.161	0.158	0.098	0.757	0.450		0.492

Table 3. Seven day-single appliance (dishwashers).

Dataset	#	f1_s	f1_r	prec_s	prec_r	rec_s	rec_r	AUC_s	AUC_r
REFIT	2	0.252	0.117	0.234	0.113	0.272	0.122	0.744	0.499
	4	0.324	0.157	0.232	0.115	0.538	0.247		0.496
	6	0.312	0.187	0.201	0.122	0.698	0.397		0.502
	8	0.307	0.190	0.186	0.117	0.866	0.509		0.514
	10	0.281	0.201	0.165	0.119	0.957	0.646		0.502
HES	2	0.218	0.132	0.194	0.121	0.248	0.146	0.645	0.503
	4	0.228	0.143	0.159	0.101	0.406	0.244		0.501
	6	0.221	0.164	0.139	0.105	0.534	0.376		0.507
	8	0.219	0.171	0.131	0.103	0.669	0.492		0.498
	10	0.214	0.170	0.124	0.099	0.789	0.594		0.494

Table 4. Seven day-single appliance (tumble dryers).

Dataset	#	f1_s	f1_r	prec_s	prec_r	rec_s	rec_r	AUC_s	AUC_r
REFIT	2	0.352	0.125	0.285	0.105	0.459	0.154	0.787	0.508
	4	0.358	0.161	0.234	0.108	0.757	0.319		0.504
	6	0.279	0.167	0.168	0.102	0.815	0.455		0.496
	8	0.255	0.182	0.147	0.107	0.950	0.629		0.498
	10	0.210	0.184	0.118	0.105	0.954	0.774		0.489
HES	2	0.299	0.154	0.250	0.135	0.371	0.180	0.685	0.478
	4	0.315	0.172	0.210	0.118	0.624	0.316		0.502
	6	0.291	0.190	0.178	0.119	0.793	0.483		0.508
	8	0.248	0.204	0.145	0.121	0.859	0.654		0.489
	10	0.213	0.210	0.121	0.120	0.897	0.812		0.480

3.2.3. Different Peak Periods

DR programs for peak shaving have been targeting mostly the early evening peak, but the frequency of alerts and the duration of DR events vary between these programs (2 h, 2.5 h, 3 h, and 4 h duration). For the 7-day types and the REFIT dataset, different peak periods were examined for all wet appliances. Apart from the evening peak also used in

the previous case studies, an early morning peak (07.00–09.00) was also chosen as relevant for UK datasets [38,39]. Different evening peak slots were also investigated (17.00–19.00, 17.00–20.00, and 16.00–20.00) to examine the different approaches of DR programs.

Overall, the three appliances studied in the REFIT dataset demonstrate different results, as can be seen in Table 5, which in turn offer information regarding the specific average use of each appliance. For example, the washing machines and tumble dryers exhibit better results (i.e., they are used more consistently) in the afternoon (AUC = 0.788 and 0.792, respectively, for the slot 16:00–20:00), while the dishwashers exhibit better results in the morning (AUC = 0.772 for the slot 07:00–09:00).

Table 5. Different peak periods-single appliance.

REFIT Dataset		Washing Machines		Dishwashers		Tumble Dryers	
#	time	f1_s	AUC_s	f1_s	AUC_s	f1_s	AUC_s
2		0.363		0.315		0.261	
4		0.400		0.270		0.199	
6	07:00–09:00	0.391	0.737	0.247	0.772	0.189	0.735
8		0.357		0.240		0.178	
10		0.347		0.213		0.149	
2		0.304		0.326		0.356	
4		0.402		0.328		0.295	
6	07:00–10:00	0.441	0.716	0.305	0.770	0.254	0.755
8		0.420		0.295		0.244	
10		0.410		0.272		0.213	
2		0.327		0.325		0.350	
4		0.417		0.326		0.299	
6	06:00–10:00	0.435	0.705	0.303	0.759	0.255	0.737
8		0.418		0.293		0.251	
10		0.408		0.278		0.219	
2		0.284		0.250		0.447	
4		0.363		0.351		0.396	
6	16:00–19:00	0.399	0.758	0.355	0.727	0.316	0.780
8		0.382		0.329		0.291	
10		0.349		0.319		0.249	
2		0.350		0.250		0.467	
4		0.436		0.413		0.446	
6	16:00–20:00	0.462	0.788	0.435	0.723	0.354	0.792
8		0.435		0.440		0.325	
10		0.411		0.419		0.277	

3.2.4. Modeling Based on Previous Day Use

The model assumes in this case that the probability of the use of an appliance on any given day depends only on whether it was used the previous day ($Y = 1$ in Equation (7)) or not ($Y = 0$ in Equation (7)). The model was evaluated on the washing machines for the 17.00–19.00 peak period in both datasets. In this case, for the REFIT dataset, the best F_1 score value was 0.316 for six households with an AUC value of 0.725. In the case of the workdays/non-workdays model, the respective values for a selection of six households

were 0.367 for the F_1 score, while the best F_1 score was obtained for a selection of four households, with a value of 0.379. In both the latter cases, AUC reached 0.770.

The results in this case (Table 6) are worse than in the case of the calculation of the probability regarding the appliance use for workdays/non-workdays and for individual days.

Table 6. Previous day use-single appliance (washing machines).

Dataset	#	f1_s	f1_r	prec_s	prec_r	rec_s	rec_r	AUC_s	AUC_r
REFIT	2	0.275	0.108	0.292	0.118	0.259	0.099		0.501
	4	0.310	0.153	0.243	0.122	0.430	0.205		0.490
	6	0.316	0.163	0.217	0.114	0.578	0.287	0.725	0.508
	8	0.295	0.169	0.189	0.110	0.671	0.369		0.498
	10	0.287	0.187	0.176	0.116	0.780	0.487		0.503
HES	2	0.184	0.084	0.188	0.088	0.181	0.081		0.503
	4	0.197	0.120	0.150	0.093	0.288	0.171		0.501
	6	0.215	0.138	0.145	0.094	0.418	0.261	0.651	0.513
	8	0.208	0.146	0.131	0.093	0.503	0.340		0.493
	10	0.214	0.146	0.129	0.089	0.621	0.408		0.493

3.3. Multi-Appliance DR Scenario

The selection of a pre-defined number of houses as receivers of a DR alert comes with a risk. For example, in the case of appliances used infrequently in a household, the selection of a relatively large number of houses as receivers of an alert will result in a number of meaningless alerts (to houses with a very low probability to use the appliance of interest in the specific timeframe). This issue can be avoided through the dynamic selection of houses that will receive a DR alert based on a threshold regarding the probability that the appliance of interest will be used in the desired timeframe. Within this context, in the multi-appliance scenario presented here, alerts are sent based on a probability threshold (0.1, 0.15, 0.2, 0.25, 0.3), resulting in a different number of receivers for each day. In this scenario, it is assumed that more than one appliance per household will participate in a DR implementation. All wet appliances in the REFIT dataset and all wet appliances and water heaters in the HES dataset are considered. The 7-day type model is applied and the 17.00–19.00 peak period is used. The model is tested regarding its ability to predict the households in the following scenarios:

1. Any one of the selected appliances will be used. Each residence receives a general alert asking to lower consumption. A result is considered a True Positive (TP) if any of the studied appliances are ON in the respective household during the DR event;
2. A specific appliance among the selected ones will be used. Each residence receives a personalized alert to turn off the appliance with the highest calculated probability during the peak period. A result is a TP only if this appliance is ON during the DR event.

The results in this case (Table 7) are significantly improved compared to the single-appliance scenario. For the REFIT database, the F_1 score reaches up to 0.513 (for Scenario 1—any of the studied appliances—and an average number of 9.719 alerts to be sent), while in Scenario 2—the corresponding single-appliance case—(7-day type model, 17:00–19:00 peak period, REFIT database) the F_1 score is 0.353 for 10 alerts to be sent.

Table 7. Multi-appliance results.

Dataset	Thres	Scenario 1			Scenario 2		
		msgs	f1_s	AUC_s	msgs	f1_s	AUC_s
REFIT	0.1	9.719	0.513		9.719	0.353	
	0.15	6.298	0.5		6.444	0.315	
	0.2	3.561	0.383	0.768	3.135	0.233	0.809
	0.25	1.567	0.188		1.421	0.127	
	0.3	1.	0.125		1.	0.09	
HES	0.1	7.976	0.488		6.136	0.288	
	0.15	5.259	0.429		4.045	0.27	
	0.2	2.	0.215	0.663	1.538	0.123	0.783
	0.25	0.706	0.086		0.543	0.064	
	0.3	0.141	0.015		0.217	0.024	

3.4. Ablation Study

To check whether a simpler model having only the start time distribution available as the probability for ranking consumers is more accurate than a more complete model as presented in Section 2, we ran the following ablation study:

- For REFIT, for the time interval 17:00–19:00 and focusing on the washing machine, when we removed the probability of an appliance being ON on a given day type, by setting the probability equal to 1 for all day types, the AUC decreased from 0.77 to 0.72;
- Again, for the REFIT dataset, when we set the probability distribution of an appliance being ON on a specific time during the day to a uniform distribution, rendering it indifferent, the AUC decreased from 0.77 to 0.68;
- The same experiment for HES decreased the AUC from 0.72 to 0.7 in the first scenario and from 0.72 to 0.67 in the second scenario.

This is an indication that these two components contribute sufficiently to the forecasting ability of the complete model.

4. Discussion

In terms of the proposed model evaluation, the results are significantly better than the random selection process outcomes (the AUC value ranges, in all cases, between 0.651 and 0.788). More importantly, there are valid reasons to believe that the results can be further improved through the utilization of more data. For example, the REFIT dataset includes as aforementioned 341 usable days. Among them, 170 were used for training. When considering the workdays/non-workdays scenario, this amounts to about 121 workday samples and 49 non-workday samples used to train the proposed model. However, in the case of the differentiation of each individual day, this training set corresponds to an average of only about 24 samples for each individual day.

The scenario regarding the different studied time slots yields similar results in terms of the effectiveness of the proposed model. However, this is partly to be expected as the selected time slots coincide with reported peak periods. Nevertheless, it is interesting to note that the respective results can offer practical information regarding the selection of specific appliances according to the time slot used in a DR implementation. Naturally, this is an area that requires further research to reach more concrete results.

The same holds for the calculation of the appliance use probability based on the operation status of the appliance on the previous day. Compared to the random selection process, the results of the model are significantly better, albeit not as good as the results obtained by the workdays/non-workdays and 7-day type models. Still, the results of this

model show that there are merits in the consideration of the previous use of an appliance in the calculation of its use probability. The results may improve significantly if the previous use consideration is extended to cover more days in the past.

Regarding the multi-appliance scenarios, the results show that the respective models can serve as a basis for practical DR implementations, as they combine a multitude of appliances of interest, a concept for the provision of personalized alerts, and a method for the dynamic selection of households as receivers of meaningful DR alerts

5. Conclusions

This paper presented a sophisticated DRAS, which is part of a peak-shaving DR program. Through the creation of appliance models based on past appliance measurements, the DRAS chooses the households with the highest probability of using their DR appliance. Three alternatives were examined for appliance use modeling along with different types of appliances and peak periods. Specifically, appliance use probability was calculated by differentiating among workdays/non-workdays as well as among each individual weekday. The third modeling alternative related the current and future appliance use probability with its previous use (i.e., the possibility that the appliance was used the previous day). Two main scenarios were examined: a single-appliance scenario and a multi-appliance scenario. The multi-appliance scenario performed significantly well for both datasets. The results of the paper are encouraging toward creating personalized approaches for DR services that could improve attractiveness and increase participation in DR peak-shaving programs. Specifically, in terms of the AUC, the scenarios studied with the proposed models yield values ranging between 0.651 and 0.788, which represent a clear improvement compared to the values in the baseline scenario, which range between 0.48 and 0.52. Moreover, in terms of the F_1 score, the proposed models improve the baseline results up to 221.05%.

The work presented here opens up several pathways for future research. For example, more research is needed in the study of different and more DR appliances to identify the best combination of meaningful DR alerts for each household. Moreover, an integration of end-users' compliance into the model could greatly improve its real-life performance. A future implementation should also consider the concerns of end users in relation to privacy issues [40]. The potential lack of trust could lead to a high dropout rate for DR programs and should be taken into account during real-life implementations.

Author Contributions: Conceptualization, I.-M.C.; Methodology, I.-M.C. and C.D.; Software, I.-M.C. and K.C.C.; Investigation, C.D.; Writing—Original Draft Preparation, I.-M.C.; Writing—Reviewing and Editing, C.D. and G.T.A.; Visualization, K.C.C.; Data Curation, K.C.C.; Validation, K.C.C.; Supervision, G.T.A. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are referenced in [36,37].

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

References

1. Palensky, P.; Dietrich, D. Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads. *IEEE Trans. Ind. Inform.* **2011**, *7*, 381–388. [CrossRef]
2. Gellings, C.W. The Concept of Demand-Side Management for Electric Utilities. *Proc. IEEE* **1985**, *73*, 1468–1470. [CrossRef]
3. Oconnell, N.; Pinson, P.; Madsen, H.; Omalley, M. Benefits and challenges of electrical demand response: A critical review. *Renew. Sustain. Energy Rev.* **2014**, *39*, 686–699. [CrossRef]

4.   Haider, H.T.; See, O.H.; Elmenreich, W. A review of residential demand response of smart grid, Renew. Sustain. *Energy Rev.* **2016**, *59*, 166–178. [CrossRef]

5.   Parrish, B.; Gross, R.; Heptonstall, P. On demand: Can demand response live up to expectations in managing electricity systems? *Energy Res. Soc. Sci.* **2019**, *51*, 107–118. [CrossRef]

6.   Federal Energy Regulatory Commission. *2020 Assessment of Demand Response and Advanced Metering*; Federal Energy Regulatory Commission: Washington, DC, USA, 2020. Available online: https://cms.ferc.gov/sites/default/files/2020-12/2020 AssessmentofDemandResponseandAdvancedMetering_December2020.pdf (accessed on 18 May 2021).

7.   Barbero, M.; Corchero, C.; Canals Casals, L.; Igualada, L.; Heredia, F.-J. Critical evaluation of European balancing markets to enable the participation of Demand Aggregators. *Appl. Energy* **2020**, *264*, 114707. [CrossRef]

8.   Tounquet, F.; Alaton, C. *Benchmarking Smart Metering Deployment in EU*; Publications Office of the European Union: Luxembourg, 2019. [CrossRef]

9.   Mohsenian-Rad, A.H.; Wong, V.W.S.; Jatskevich, J.; Schober, R.; Leon-Garcia, A. Autonomous Demand Side Management Based on Game-Theoretic Energy Consumption Scheduling for the Future Smart Grid. *IEEE Trans. Smart Grid* **2010**, *1*, 320–331. [CrossRef]

10. Zhou, B.; Li, W.; Chan, K.W.; Cao, Y.; Kuang, Y.; Liu, X.; Wang, X. Smart home energy management systems: Concept, configurations, and scheduling strategies. *Renew. Sustain. Energy Rev.* **2016**, *61*, 30–40. [CrossRef]

11. Parrish, B.; Heptonstall, P.; Gross, R.; Sovacool, B.K. A systematic review of motivations, enablers and barriers for consumer engagement with residential demand response. *Energy Policy.* **2020**, *138*, 111221. [CrossRef]

12. Srivastava, A.; Van Passel, S.; Laes, E. Assessing the success of electricity demand response programs: A meta-analysis. *Energy Res. Soc. Sci.* **2018**, *40*, 110–117. [CrossRef]

13. Lesic, V.; Bruine De Bruin, W.; Davis, M.C.; Krishnamurti, T.; Azevedo, I.M.L. Consumers' perceptions of energy use and energy savings: A literature review. *Environ. Res. Lett.* **2018**, *13*, 033004. [CrossRef]

14. Element Energy; Chan, A.; Garcia Moreno, J.; Hughes, M. Electricity Price Signals and Demand Response. 2014. Available online: https://www.gov.uk/government/publications/electricity-price-signals-and-demand-response (accessed on 18 May 2021).

15. Thayer, D.; Brummer, W.; Smith, B.A.; Aslin, R.; Cook, J. Is Behavioral Energy Efficiency and Demand Response Really Better Together? *ACEEE Summer Study Energy Effic. Build.* **2016**, 1–11.

16. Wang, Z.; Li, H.; Deng, N.; Cheng, K.; Lu, B.; Zhang, B.; Wang, B. How to effectively implement an incentive-based residential electricity demand response policy? Experience from large-scale trials and matching questionnaires. *Energy Policy* **2020**, *141*, 111450. [CrossRef]

17. Pratt, B.W.; Erickson, J.D. Defeat the Peak: Behavioral insights for electricity demand response program design. *Energy Res. Soc. Sci.* **2020**, *61*, 01352. [CrossRef]

18. Thomas, E.; Sharma, R.; Nazarathy, Y. Towards demand side management control using household specific Markovian models. *Automatica* **2019**, *101*, 450–457. [CrossRef]

19. Kong, W.; Dong, Z.Y.; Hill, D.J.; Luo, F.; Xu, Y. Short-term residential load forecasting based on resident behaviour learning. *IEEE Trans. Power Syst.* **2018**, *33*, 2017–2018. [CrossRef]

20. Basu, K.; Hawarah, L.; Arghira, N.; Joumaa, H.; Ploix, S. A prediction system for home appliance usage. *Energy Build.* **2013**, *67*, 668–679. [CrossRef]

21. Wenninger, M.; Schmidt, J.; Goeller, T. Appliance usage prediction for the smart home with an application to energy demand side management and why accuracy is not a good performance metric for this problem. In Proceedings of the SMARTGREENS 2017-6th International Conference on Smart Cities and Green ICT Systems, Porto, Portugal, 23–24 April 2017; pp. 143–150. [CrossRef]

22. Kang, Z.; Jin, M.; Spanos, C.J. Modeling of end-use energy profile: An appliance-data-driven stochastic approach. In Proceedings of the IECON 2014-40th Annual Conference of the IEEE Industrial Electronics Society, Dallas, TX, USA, 29 October–1 November 2014; pp. 5382–5388. [CrossRef]

23. Chrysopoulos, A.; Diou, C.; Symeonidis, A.L.; Mitkas, P.A. Bottom-up modeling of small-scale energy consumers for effective demand response applications. *Eng. Appl. Artif. Intel.* **2014**, *35*, 299–315. [CrossRef]

24. Huber, P.; Gerber, M.; Rumsch, A.; Paice, A. Prediction of domestic appliances usage based on electrical consumption. *Energy Inform.* **2018**, *1*, 265–271. [CrossRef]

25. Chatzigeorgiou, I.-M.; Diou, C.; Andreou, G.T. Targeted Messaging for Appliance-based Demand Response. In Proceedings of the 2020 55th International Universities Power Engineering Conference (UPEC), Turin, Italy, 1–4 September 2020; pp. 1–5. [CrossRef]

26. Razghandi, M.; Turgut, D. Residential Appliance-Level Load Forecasting with Deep Learning. In Proceedings of the GLOBECOM 2020-2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; pp. 1–6. [CrossRef]

27. Mehdipour Pirbazari, A.; Farmanbar, M.; Chakravorty, A.; Rong, C. Short-Term Load Forecasting Using Smart Meter Data: A Generalization Analysis. *Processes* **2020**, *8*, 484. [CrossRef]

28. Ji, Y.; Buechler, E.; Rajagopal, R. Data-Driven Load Modeling and Forecasting of Residential Appliances. *IEEE Trans. Smart Grid* **2019**, *11*, 2652–2661. [CrossRef]

29. Din, G.M.U.; Mauthe, A.U.; Marneride, A.K. Appliance-level Short-Term Load Forecasting using Deep Neural Networks. In Proceedings of the 2018 International Conference on Computing, Networking and Communications (ICNC), Maui, HI, USA, 5–8 March 2018; pp. 53–57. [CrossRef]

30. Singh, S.; Yassine, A. Mining Energy Consumption Behavior Patterns for Households in Smart Grid. *IEEE Trans. Emerg. Top. Comput.* **2017**, *7*, 404–419. [CrossRef]
31. Pipattanasomporn, M.; Kuzlu, M.; Rahman, S.; Teklu, Y. Load profiles of selected major household appliances and their demand response opportunities. *IEEE Trans. Smart Grid* **2014**, *5*, 742–750. [CrossRef]
32. Klaassen, E.; Kobus, C.; Frunt, J.; Slootweg, J. Responsiveness of residential electricity demand to dynamic tariffs: Experiences from a large field test in the Netherlands. *Appl. Energy* **2016**, *183*, 1065–1074. [CrossRef]
33. Klaassen, E.; Kobus, C.; Frunt, J.; Slootweg, H. Load shifting potential of the washing machine and tumble dryer. In Proceedings of the 2016 IEEE International Energy Conference (ENERGYCON), Leuven, Belgium, 4–8 April 2016; pp. 1–6. [CrossRef]
34. Yamaguchi, Y.; Chen, C.F.; Shimoda, Y.; Yagita, Y.; Iwafune, Y.; Ishii, H.; Hayashi, Y. An integrated approach of estimating demand response flexibility of domestic laundry appliances based on household heterogeneity and activities. *Energy Policy* **2020**, *142*, 111467. [CrossRef]
35. Pereira, L.; Nunes, N. Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—A review. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1265. [CrossRef]
36. Murray, D.; Stankovic, L.; Stankovic, V. An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. *Sci. Data* **2017**, *4*, 1–12. [CrossRef]
37. Palmer, J.; Terry, N. Powering the Nation 2, (n.d.). Available online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/325741/Powering_the_Nation_2_260614.pdf (accessed on 18 May 2021).
38. Stamminger, R.; Schmitz, A. Load profiles and flexibility in operation of washing machines and dishwashers in Europe. *Int. J. Consum. Stud.* **2017**, *41*, 178–187. [CrossRef]
39. Bobmann, T.; Staffell, I. The shape of future electricity demand: Exploring load curves in 2050s Germany and Britain. *Energy* **2015**, *90*, 1317–1333. [CrossRef]
40. McKenna, E.; Richardson, I.; Thomson, M. Smart meter data: Balancing consumer privacy concerns with legitimate applications. *Energy Policy* **2012**, *41*, 807–814. [CrossRef]