ORIGINAL RESEARCH Open access

Machine Learning Approaches for Understanding Residential Demand Response in Smart Grids

Santiago Morales[1*], Laura Rojas[2], Camila Vega[1], Diego Molina[3]

Abstract

Interest in demand response (DR) has grown significantly in recent years as a strategy to meet the expanding flexibility requirements of contemporary electricity grids. This need for enhanced flexibility is driven by the rising integration of intermittent renewable energy sources into the generation mix and the escalating complexity of consumption patterns caused by the electrification of transportation networks. Currently, under 2% of worldwide demand-side flexibility capacity is being exploited; however, expanding the integration of domestic users as flexibility assets could drastically improve this utilisation rate. Reaching this objective requires a deeper understanding of how residential participants respond during DR events. In this context, modern developments in statistical artificial intelligence and machine learning offer valuable mechanisms to tackle this issue. The present research delivers a thorough evaluation of domestic consumer responses within an incentive-driven DR framework, drawing on household data from a prominent, real-world pilot: the Smart Grid, Smart City (SGSC) project. By employing various machine learning techniques, the correlation between household characteristics and their corresponding DR responses is modelled. Furthermore, the study investigates the possible impacts of these household attributes on residential response patterns, uncovering crucial insights that challenge the documented degrees of consumer participation in DR initiatives and the underlying motivations for varying response magnitudes. Ultimately, the temporal dimensions of these responses are assessed. While no empirical evidence was found to suggest that DR participants in this specific trial adapted or learned over time, the analytical frameworks introduced herein hold promise for extended longitudinal DR investigations. The paper closes with a broader discussion of our findings and prospective trajectories for subsequent research within this developing domain.

Keywords Demand response, Machine learning, Artificial intelligence, Artificial neural networks, Ensemble methods, Residential response behavior

*Correspondence:
Santiago Morales
santiago.morales@gmail.com
[1] Department of Intelligent Renewable Energy Systems, Faculty of Engineering, University of Buenos Aires, Buenos Aires, Argentina
[2] Department of Smart Energy Analytics, Faculty of Engineering, Pontifical Catholic University of Chile, Santiago, Chile
[3] Department of AI-Based Smart Grid Technologies, Faculty of Engineering, National University of La Plata, La Plata, Argentina

Introduction

The rising share of renewable energy sources, combined with the expanding integration of fluctuating load categories (e.g., Electric Vehicles) into the energy portfolio, introduces novel complications for power networks [1]. For power systems to operate efficiently, they must function within strictly defined parameters for both frequency and voltage. Consequently, a continuous equilibrium between electrical supply and demand must be preserved at all times. Historically, grid operators have depended heavily on conventional strategies, including costly infrastructure upgrades, the continuous operation of traditional power facilities, and spinning reserves to ensure adequate backup capacity. Demand

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

response (DR) serves as a viable alternative for managing these complications [2]. DR is recognised as a promising mechanism capable of supplying the flexibility required to offset the fluctuating nature of renewable power production and the reduced predictability of upcoming electrical loads [1]. Implementing DR strategies enables network administrators to cost-effectively maintain the balance between generation and load, thereby circumventing or postponing expensive grid infrastructure expansions, or avoiding capital expenditures on supplementary backup power plants.

Based on a 2020 publication by the IEA [3], the current exploitation of worldwide demand-side flexibility potential remains under 2%. The primary contributors to this demand-side flexibility encompass industrial thermal operations and processes, climate control systems in both domestic and non-residential structures, electric vehicle charging, as well as localized power generation and energy storage [3]. Such demand-side flexibility can be provisioned through Virtual Power Plants (VPPs), specialized demand response providers, and prosumers [2]. Throughout academic literature, demand response is frequently highlighted as a fundamental element for delivering this essential adaptability. Numerous investigations have examined the operational success and efficacy of demand response frameworks, yielding empirical evidence that such initiatives (e.g., dynamic pricing models) can supply the requisite flexibility [4-7]. Nevertheless, the foundational scale of demand response remains notably constrained, necessitating a massive global expansion of flexible resources across the residential, commercial, and industrial domains to reach levels multiple times higher than today [3].

To achieve broad implementation of demand response programmes, it is crucial to engage distributed residential consumers, as they represent a massive reservoir of demand response potential. However, the varied, heterogeneous, and dispersed characteristics of residential DR assets complicate service entities' (i.e., demand-side aggregators) ability to engage effectively in electricity markets using pooled demand-side capacities [8]. Precisely predicting the reaction patterns of DR participants and comprehending the underlying motivations for their actions are vital steps in overcoming these obstacles. Reliable predictions of response behaviour within DR frameworks can reduce ambiguity about available flexibility, which directly influences the dependability of services provided to electricity grids. A deeper insight into how consumers react empowers academic researchers, energy retailers, and system operators to formulate more accurate and targeted pricing or incentive structures for DR initiatives. The structural design of a DR mechanism impacts both the financial viability for involved parties and the overarching success of the DR scheme. Consequently, the utilization of energy big data analytics and related data-driven methodologies serves as an effective instrument for investigating consumer response dynamics [9].

The present study anchors its analysis on data derived from the Smart Grid Smart City (SGSC) trial project [10]. Functioning as one of the most extensive residential DR experiments globally, the SGSC project spanned nearly 2 years. Throughout the trial's lifespan, participating consumers were presented with a variety of product packages [11]. These packages integrated various pricing and incentive models (e.g., dynamic peak pricing, dynamic peak rebate, seasonal time-of-use, etc.) alongside technologies designed to provide usage feedback. Readers seeking a more comprehensive overview of the SGSC pilot are directed to the official project documentation and the research conducted by Motlagh *et al*. [12]. Within the expansive scope of the SGSC experiment, the present study focuses specifically on the dynamic peak rebate (DPR) incentive structure administered by the network trial partner. Under this initiative, participating customers were financially incentivised via rebates to lower their electricity consumption. To solicit their involvement, customers were alerted to impending DPR events through SMS messages dispatched 24 hours before the occurrence [11]. Every participant engaged in this specific DR framework was a residential/domestic consumer.

Contributions and study outline

This research delivers novel and supplementary perspectives designed to build upon previous studies, thereby generating a more holistic understanding of demand response behaviour and its underlying catalysts. Most existing academic literature relies on household load profiles (energy usage data recorded at varying temporal points) to forecast households' response behaviour (the curtailment of consumption following a signal), often neglecting specific household attributes. In contrast, the current investigation incorporates household-specific variables—such as the nature of the

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

heating systems, internet connectivity, and air-conditioning availability—to evaluate participant reactions. The primary contributions of this work are outlined below:

• A comprehensive evaluation of the response behaviour from a substantial cohort of households, utilizing household and appliance data extracted from a major real-world experiment.

• The forecasting of household response actions relying entirely on household-specific variables, achieving precision levels consistent with expectations — considering findings from prior research that predominantly utilized load data for predictive modelling.

• An examination of the temporal architecture of the response patterns and an assessment of the responders' capacity to learn progressively.

• An explicit and comprehensible investigation into how distinct household characteristics and appliance usage impact the resultant energy reductions.

Furthermore, this investigation aims to foster more transparent and reproducible research by utilizing open-source demand-response datasets and publishing the code required to duplicate the exact findings via a public GitHub repository1. Ultimately, this study aspires to refine the modelling and comprehension of DR response behaviour, thereby facilitating enhanced strategic planning and decision-making within the demand response sector.

The manuscript is structured systematically. It first provides an overview of the data-centric methodologies applied in this paper, alongside a literature review of existing data-driven approaches used to evaluate demand response behaviour. This establishes the necessary contextual framework and situates the present methodology within the broader academic discourse. Following this, response dynamics are modelled and assessed by first examining the relevant response datasets. This preliminary phase focuses on "cleaning" the data and identifying inherent traits—as well as potential anomalies—through the application of contemporary exploratory data analysis and statistical tests. Additionally, the potential for DPR participants to learn over time is investigated, and the data are analysed for temporal response patterns. Various statistical and machine learning frameworks are then applied to construct models linking households' DR responses to their specific characteristics. After establishing these correlational models, the analysis seeks to isolate the critical variables and extensively discusses how these attributes influence the households' performance under the DR initiative. Finally, the manuscript concludes with a synthesis of the findings and suggestions for prospective avenues for future studies in this field.

Data-Driven Approaches for Demand Response Behaviour

Fundamentally, simulating how consumers react to DR programs represents an intricate and shifting challenge influenced by both external and internal variables for individual users. Data-driven techniques provide a mechanism for detecting trends, constructing models, and deriving knowledge from accessible datasets [9].

Overview of key data-driven techniques

The wide array of data-driven techniques can be categorized based on their specific uses. Exploratory data analysis tools use descriptive statistics and visual representations to examine the underlying structure, the distribution of variables, and the internal relationships within the datasets [13]. This preliminary data exploration helps guide the subsequent data processing stages.

Data preparation involves the preliminary processing of raw information to remove noise, detect inconsistencies, and locate possible systematic flaws. Within the realm of demand response, this includes identifying inaccurate smart meter measurements, baseline approximations, and possible excessive compensation figures. It is necessary to preprocess the

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

data to identify missing records in smart meter logs, consumer surveys, and the overall dataset, preventing substandard data from feeding into models and generating flawed predictions and conclusions. The issue of absent data can be addressed through complete-case analysis or various imputation strategies (spanning from basic mean imputation to complex multiple imputation utilizing Markov Chain Monte Carlo [14]). When analysing missing data, it is crucial to understand the fundamental mechanism causing the omissions [15], such as whether the absences occur at random. The de-noising process involves identifying and eliminating data inaccuracies and extreme outliers. Data preparation remains a critical phase in the analytical workflow; empirical data is inherently flawed and noisy, and unchecked errors or biases can carry over into later analytical and modelling phases, profoundly impacting the outcomes of data-driven methodologies [16].

Additionally, data mining helps cluster consumer behaviour, facilitating the recommendation of suitable users for DR initiatives, the selection of event participants, and the formulation of DR strategies. Common classification approaches encompass both supervised and unsupervised clustering techniques. Algorithms like k-means and self-organising maps are frequently deployed in demand response settings to analyse reaction patterns. Because a substantial portion of demand response data consists of time series, employing time series analysis tools is critical for uncovering latent temporal frameworks. Such instruments involve evaluating autocorrelation, as well as detecting trends and seasonal cycles. For instance, autocorrelation (also known as serial correlation) quantifies the linear relationship between the lagged elements of a sequence, such as a time series. Thus, a lag autocorrelation represents the correlation between variables separated by k distinct time intervals.

A further significant use case involves data modelling and the approximation of focal variables. Within DR, this translates to forecasting energy consumption to equip power generators with the necessary data for precise production scheduling, evaluating potential load reductions, and assessing the dependability of DR subjects, among other factors.

Prevalent strategies for applied data-driven modelling include ensemble techniques and artificial neural networks [17]. The most frequently utilized ensemble approaches are gradient boosting and random trees. Gradient boosting [18] comprises a collection of weak prediction algorithms (usually decision trees) operating as an additive, forward stage-wise boosting framework. Consequently, the procedure introduces new decision trees consecutively without modifying the weights and parameters of the previously incorporated trees [18]. Furthermore, these models permit the optimization of any differentiable loss function [19].

The random forest algorithm is another decision tree-based ensemble approach, functioning by generating numerous distinct decision trees that are separately trained using a randomised data subset [20]. During the development of each tree, the input variables (or their combinations) at every node are chosen randomly [20]. In contrast to gradient boosting (which constructs weak models in a sequence), random forests operate as an averaging ensemble technique; they independently develop weak decision trees and subsequently average the outputs produced by these individual estimators.

Artificial neural networks represent computational architectures that draw inspiration from, though are not exact replicas of, biological nervous systems. The two primary structures of artificial neural networks are the feed-forward and the recurrent configurations [21]. A dense neural network exemplifies a feed-forward design wherein every layer maintains full connectivity. Specifically, every node within a dense (fully-connected) layer links directly to every node in the following layer [22]. Dense neural networks, along with feed-forward models more broadly, can be conceptualized as universal function approximators designed to accomplish statistical generalisation [22].

In relatively underexplored and unevaluated fields such as demand response, it is crucial to comprehend the mechanisms behind the models' outputs rather than concentrating exclusively on their forecasting accuracy [23]. To achieve this, interpreting data-driven models is essential. Two prominent model-agnostic interpretation strategies are the SHAP (SHapley Additive exPlanations) framework [24] and the permutation feature importance method [20, 25]. These methodologies were adopted because previous studies indicate that when dealing with categorical input variables (as is

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

the case in this research), Gini importance and standard variable importance metrics may exhibit bias favoring features with a greater number of categories [26]. These two systems tackle the concept of feature importance from distinct perspectives.

Fundamentally, the permutation feature importance technique allocates significance to different input variables by evaluating the decline in model efficacy. At the same time, the SHAP method relies on how each feature contributes to the final predictions. Specifically, the model-agnostic permutation algorithm determines a feature's importance by observing the fluctuation in the model's error rates after randomly reordering the values of that particular feature [25]. Because this shuffling severs the link between the target variable and the feature, an increase in error metrics (positive importance) signifies that the model relies heavily on that feature. Conversely, negative importance (a reduction in error metrics post-permutation) implies that the feature holds little significance for the model [20].

The SHAP methodology provides a system for elucidating model forecasts by attributing a specific importance value to each feature for a given prediction [24]. This framework clarifies the output for a specific data instance x by calculating the individual contribution of every feature [24]. A feature's contribution is derived by estimating Shapley values rooted in coalitional game theory [27]. Within this theoretical structure, the feature values of the data instance x act as the coalition players, and the Shapley values indicate the equitable distribution of the payout (which, in this context, is the prediction) across the participating features.

Related work on data-driven methods for DR response behaviour

The examination and modelling of consumer reaction patterns have been extensively investigated throughout DR academic literature (refer to Antonopoulos *et al*. [28] for a comprehensive review). A multitude of studies employ data-driven approaches to simulate and forecast consumer reactions to DR prompts. For instance, Zhou *et al*. [29] calculate decreases in power usage during specific demand response periods by integrating latent (unobservable) variables into statistical prediction frameworks—such as Ordinary Least Squares, k-Nearest Neighbours, Support Vector, and decision tree regression models. These latent attributes are derived from the users' usage records through the sequential application of a Conditional Gaussian Mixture model and a Hidden Markov model. Furthermore, Liu *et al*. [8] utilize simulated consumer response information to train a Long Short-term Memory (LSTM) neural network. This response information was formulated based on a response function that exhibits a quadratic relationship with the incentives provided to the customers.

Paterakis *et al*. [30] address the challenge of forecasting consumer reactions to fluctuating DR pricing signals by implementing a hybrid load prediction architecture. Through wavelet transformation, the time-series consumption data are broken down into distinct frequency signals, which subsequently serve as inputs for dense neural networks. The outcomes generated by these neural networks are then combined to formulate the ultimate response prediction. Similarly, Holtschneider and Erlich [31] utilize a dense neural network to evaluate response dynamics prompted by DR pricing signals. To overcome the challenge of limited training data for response prediction—given that DR events occur infrequently throughout the year—certain studies have utilized non-parametric techniques, such as an ensemble of k-Nearest Neighbours regression frameworks [32], or methodologies grounded in transfer learning [33]. Specifically, Cai *et al*. [33] strive to forecast consumer responses within incentive-driven DR programs by leveraging load profiles from other DR subjects exhibiting comparable energy consumption patterns.

The input parameters for these data-driven models are highly diverse. They encompass engineered variables derived from user consumption profiles, such as maximum and minimum daily load, and mean usage [8, 32], in addition to the raw electricity consumption metrics [29, 30, 33]. Frequently, chronological elements are incorporated into the algorithms, including week, month and hour of the DR occurrence, along with seasonal scores [8, 32], as well as meteorological conditions such as ambient air temperature [29, 33] and humidity [33]. The primary target variable in forecasting DR responses has predominantly been the consumers' actual power usage [8, 29, 30]. Conversely, Cai *et al*. [33] project the amount of load curtailed (the variance between actual power use and the projected baseline), while Kang and Lee [32]

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

predict the ratio of the achieved load decrease relative to the solicited reduction. The datasets used to train these learning models have largely consisted of simulated or semi-synthetic information [8, 29, 30]. However, empirical DR data from real-world applications have also been leveraged in certain instances [32, 33].

Furthermore, categorizing the primary attributes of DR subjects is a standard strategy for evaluating diverse response patterns. Existing literature has predominantly relied on consumer load profiles to conduct this behavioural analysis. Motlagh *et al*. [12] concentrate on extracting insights from usage profiles by contrasting the energy consumption habits of DR participants against those of a control cohort. They employ Principal Component Analysis to identify the core components from consumption metrics recorded at half-hour intervals, subsequently clustering these findings via self-organising maps based on the derived principal components. Alternative clustering methodologies, including k-means [34], expectation maximisation [34], and finite mixture-based clustering [35], have similarly been utilized to comprehend the usage behaviour of DR participants.

Additional prevalent AI methodologies applied to investigate demand response dynamics include mechanism design and cooperative game theory—approaches that are typically structural rather than purely data-driven. These strategies have been applied to the formulation of demand response agreements, as seen in Refs. [36-38], the creation of demand response incentive mechanisms [39, 40], the organization of demand-side and consumer coalitions [41-43], or the coordination of specific devices, such as electric vehicles [44, 45].

It should be highlighted that, although preceding research has yielded significant progress and insights, it generally concentrates on analysing responses through the primary lens of load data. In contrast, the objective of the present research is to investigate the modelling and analysis of response patterns utilizing attributes related to specific households and devices. A particular focus is placed on determining how these specific characteristics influence the overall DR response behaviour.

Individual Datasets Exploration and Analysis

For this investigation, data derived from the Smart Grid, Smart City trial project is utilized. The SGSC initiative represents one of the most extensive commercial-level smart grid technology experiments worldwide [12]. It encompasses a broad array of experimental information, spanning from household and device-specific characteristics to the usage patterns of electric vehicles. The specific SGSC trial datasets incorporated into our research include:

• Customer household dataset

• Peak events dataset

• Peak event response dataset

• Home Area Network Plug Readings (HANPR) dataset

The household dataset comprises information on enrolled residences, including the specific product categories they engaged with, their energy and gas consumption tiers, the architectural style of the residence, and the array of appliances present, among other variables. Furthermore, the Home Area Network Plug Readings dataset records the power consumption of various domestic appliances, providing more detailed insights into the specific electrical devices operated by the customers. The peak events dataset supplies chronological details concerning DPR and DPP occurrences, while the peak event response collection details actual energy use, baseline projections, and associated rebate figures. Before implementing machine learning techniques to simulate DR response patterns, exploratory data analysis tools and data investigation methods are employed to preprocess these diverse datasets.

Household and HANPR data analysis

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

At the outset of preparing the household dataset, we eliminated variables that lacked relevance to our study. This exclusion encompassed data regarding non-tariff products (such as feedback technologies or lifestyle audits), repetitive elements (like internal references), and attributes generally uninformative regarding participant behavior during DR events. While the SGSC trial presented consumers with multiple tariff options, the current analysis specifically targets the incentive-oriented DPR program. Consequently, any households not enrolled in the DPR initiative were removed. Following this filtration procedure, the finalized dataset contained no missing values.

The household dataset comprises both categorical (qualitative) and numerical features, with categorical variables constituting the majority. Analyzing the degree of association across a blend of variable types can be challenging. For this study, we addressed this by dividing the dataset into two distinct groupings based on feature type. We then computed Spearman's correlation for the numerical subset and Cramér's V2 for the categorical subset [46]. **Figures 1 and 2** visually represent the association strengths for the numerical and categorical features, respectively.

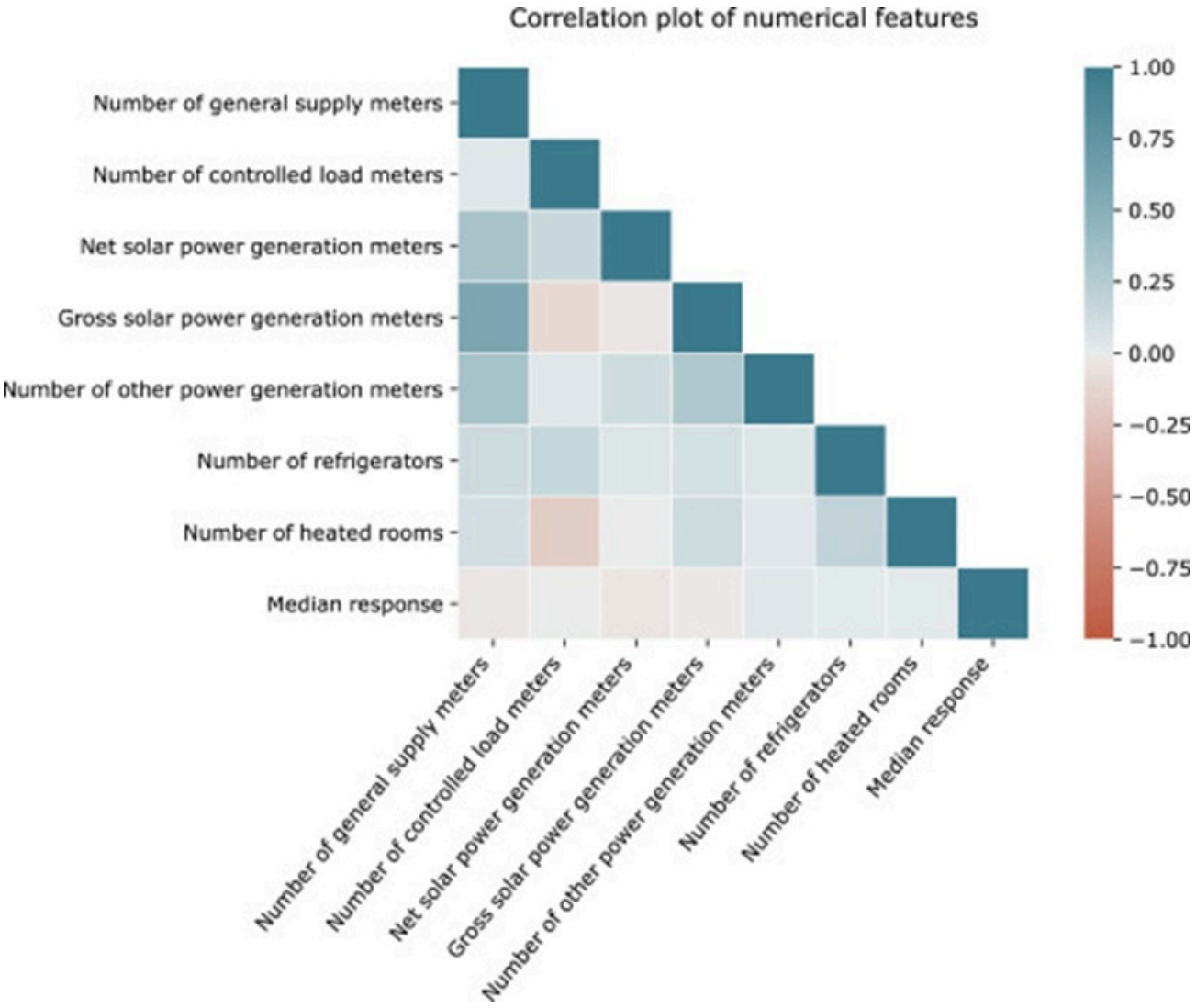

Figure 1. Spearman's correlation among the numerical household characteristics. The colour bar indicates the correlation value ranging from -1 to +1.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

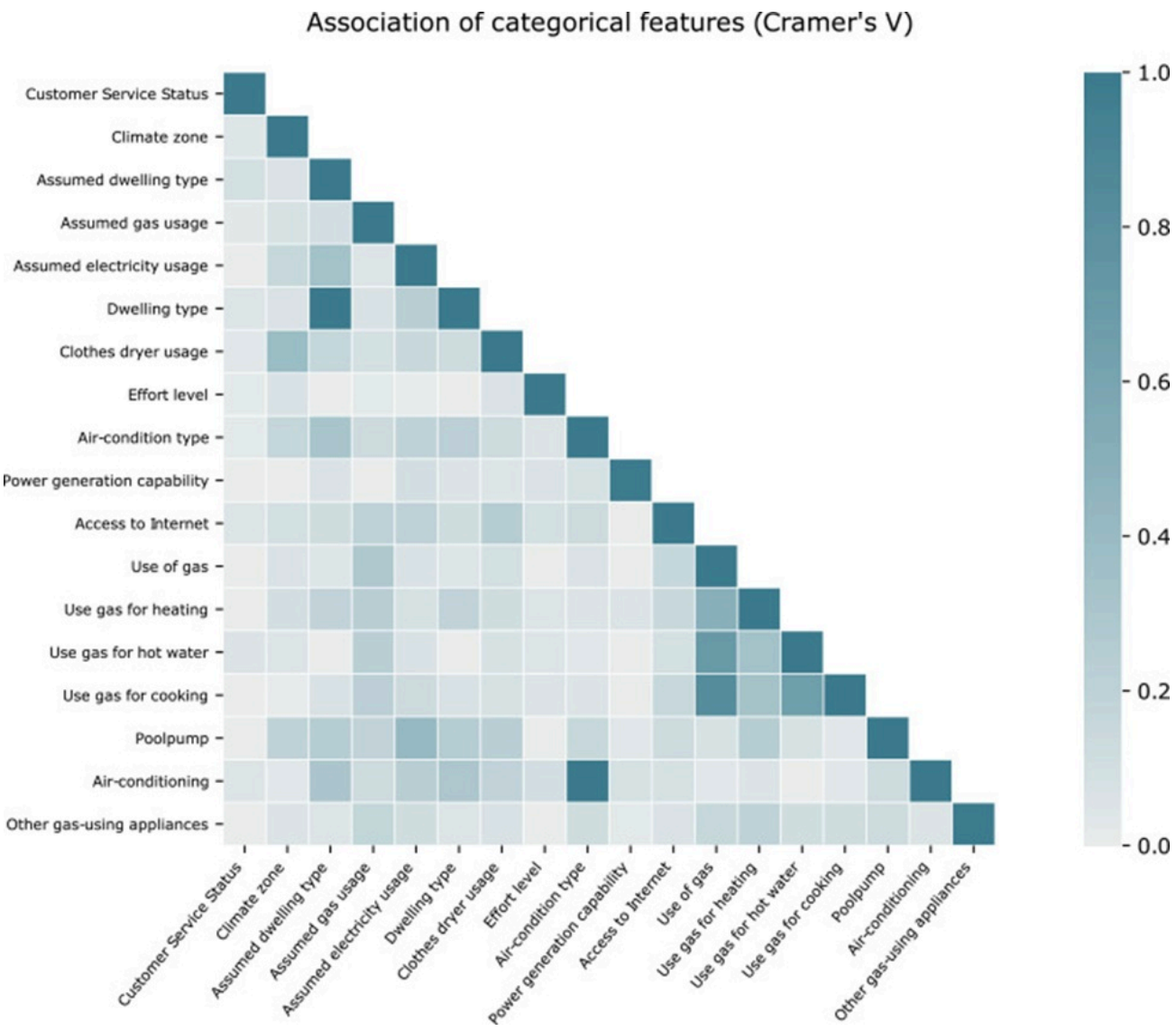

Figure 2. Cramér's V association for the categorical household features. The colour bar indicates the value of this association measure, ranging from 0 to 1.

In our research, where a primary objective of the regression modeling is to assess the relationship between predictive variables and household DR responsiveness, multicollinearity among inputs can skew analyses of feature impacts and lead to inaccurate interpretations of the developed models [24, 47, 48].

Observing **Figure 1**, it is evident that the numerical features generally lack high correlation with one another. The solitary exception demonstrating notable correlation involves features tied to the household's energy generation meters (the number of general supply meters, miscellaneous generation meters, and gross or net solar generation meters), which is an anticipated outcome. From **Figure 2**, we can determine that strong associations are largely absent among the categorical features, aside from the specific instances outlined below:

• Between the assumed dwelling type and the actual dwelling of the participating household.

• Between the air-conditioning type and the response to whether the participating household has air-conditioning.

• Among the gas-related features of the household.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

Regarding the HANPR data, which supplies more detailed insights into the specific devices within each household, we investigated the possibility of engineering new features based on participant appliances. The goal was to potentially utilize these engineered features to forecast individual household responses to DPR events. Ultimately, this dataset contained relevant appliance information for only a marginal fraction of the total participating homes; as a result, we did not incorporate these features further into our modeling.

Peak events and response data analysis

Within the trial project, two specific peak event products were provided to participants: dynamic peak rebate (DPR) and dynamic peak pricing (DPP). The analytical and modeling focus of this paper centers on the incentive-driven DPR framework. Under the DPR structure, engaged households were provided a financial rebate (constant at AUD 4.5/$kW\ h$) as an incentive to reduce their power consumption relative to their calculated baseline.

Throughout the SGSC trial, 18 total DPR events occurred. These DPR events lasted between two and four hours, occurred exclusively on weekdays (predominantly Thursdays and Fridays), and were concentrated mostly during the core winter and summer months. Additionally, they generally took place during the late afternoon, as demonstrated in **Figure 3**. Given that this dataset captures residential consumers in Australia, the occurrence of peak events during summer afternoons aligns logically with the peak operational periods of air conditioning systems.

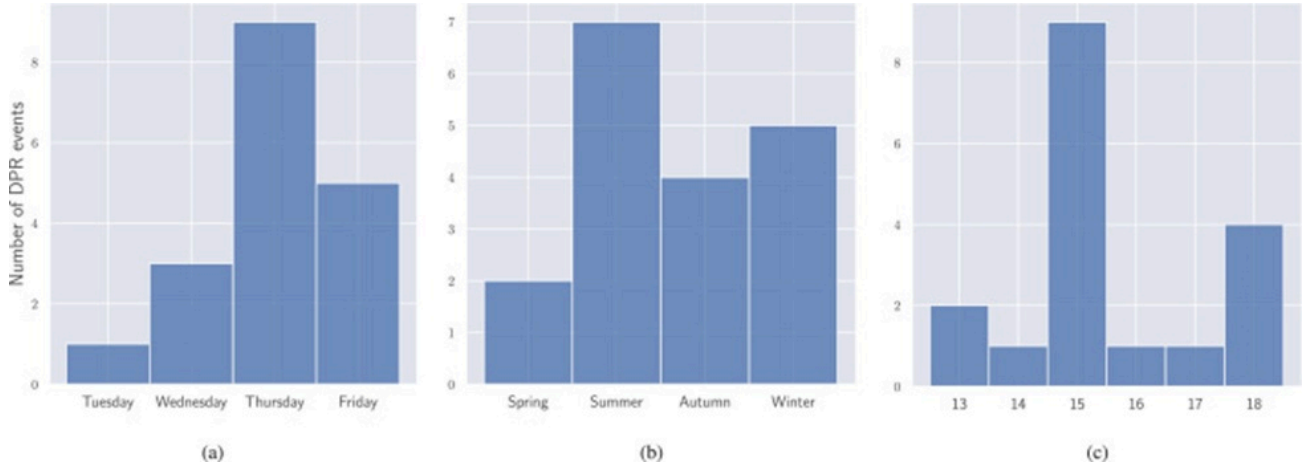

Figure 3. Number of DPR events per day (a), per season (b), and per hour of day (c).

Transitioning to the DPR response records, comprehensive data is available for only 13 of the 18 total DPR events (located within the peak event dataset). Consequently, for these 13 events, we have baseline estimates, actual energy consumption, and rebate amounts for all participants. It is important to clarify that throughout this study, the term “response” regarding a DR event refers strictly to the percentage difference between the actual electricity consumption of the household and its estimated baseline consumption, measured over the duration of that specific DR event — as described in **Eq. 1.**

$$r_{i,j} = \frac{load_{i,j} - base_{i,j}}{base_{i,j}} \times 100 \quad (1)$$

Where $r_{i,j}$ is the response of household i for the DR event j. Respectively, $load_{i,j}$ is the actual consumption, and $load\ i$, is the baseline consumption (estimated by the service provider) of household i for the duration of the DR event j. This evaluation metric was chosen because it allows for equitable comparisons across diverse customers, programmatic contracts, and DR frameworks. However, it also inherently carries limitations; specifically, inaccuracies originating from its

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

constituent components (such as baseline estimation errors) can cascade into the response variable itself. Because baseline estimation challenges are a fundamental hurdle in demand response initiatives [49, 50], we screened for potentially erroneous baseline estimates and removed those data points.

Initially, noting that a standard household utilizes an average of more than 0.45 kWh over 2 hours, we eliminated all instances where the calculated baseline fell below 0.01 kWh. We concluded that readings below this electricity threshold either reflect flawed baseline estimates or indicate that the properties were vacant at the time, thereby offering no analytical value for our response assessment.

Table 1 outlines the descriptive statistics about the customers' actual consumption, baseline consumption, received rebates, and overall responses during the DPR events. These statistical summaries include the mean, standard deviation (std), minimum and maximum extremes and the interquartile ranges (25%, 50% and 75%).

Table 1. Descriptive statistics of the consumers' actual and baseline consumption, rebate amount and response across the DPR events.

Empty cell	Actual consumption (*kWh*)	Baseline consumption (*kWh*)	Rebate amount (*AUD*)	Response (%)
mean	3.331	5.360	14.501	34.247
std	3.620	4.889	17.148	799.552
min	0.000	0.100	0.000	-100.0000
25%	1.031	1.865	2.803	-70.637
50%	2.094	3.936	9.100	-40.411
75%	4.293	7.405	19.660	7.204
max	39.565	50.972	164.800	66218.750

The data reveals that the mean response for the DPR initiative is positive (34.25%), which indicates that rather than reducing their energy draw during a peak (DPR) event, the average household actually escalated their consumption by approximately one-third. Furthermore, the standard deviation (std) concerning customer response during DPR events is notably extreme (≈ 800%). This volatility can likely be traced back to data outliers alongside inherent flaws in baseline calculations. Examining the 3rd quartile (75%) response value confirms it is positive, signifying that in a significant portion of instances (> 25% of the cases), participants failed to curtail their consumption during a DPR event. Both the standard deviation and the substantially high maximum values strongly point to the presence of outliers within the dataset.

We discarded DPR events exhibiting anomalous distribution characteristics. To manage the remaining outliers in the retained DPR events, we utilized unsupervised anomaly/outlier detection methodologies. This outlier filtration was executed on a per-DPR-event basis rather than per household. This approach was adopted primarily because baseline estimations are subject to temporal shifts, and the limited frequency of DPR events per household renders individualized anomaly detection ineffective.

For this objective, we evaluated the Isolation Forest and the Local Outlier Factor (LOF) [51, 52] algorithms. The LOF technique is a density-centric model that uses k-nearest neighbors to gauge the local density of individual samples, isolating data points that exhibit vastly lower densities than their localized clusters. Conversely, the Isolation Forest employs a distinctly different methodology, focusing on directly isolating anomalies rather than establishing a profile for

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

normal behavior [52]. It relies on an ensemble of random splitting trees, flagging data points with short average path lengths within the trees as outliers.

Our research utilized a combined algorithmic strategy akin to the one detailed by Cheng *et al*. [53], attempting to leverage both methods to mitigate their individual shortcomings. We initially deployed the Isolation Forest to extract a preliminary cluster of outliers; subsequently, the LOF algorithm was executed on that isolated set to refine the local outlier detection and finalize the definitive outlier group. The resulting response distribution across households per DPR event, post-outlier removal, is visualized in **Figure 4.**

Observing **Figure 4** reveals that substantial variability in household DR responses persists even after outlier elimination. This finding aligns consistently with established literature [4], which notes that average reductions during DR trials generally span from 10% to 50%. The underlying challenge of high variance within DR contexts has similarly been explored in the research of Aïd *et al*. [36].

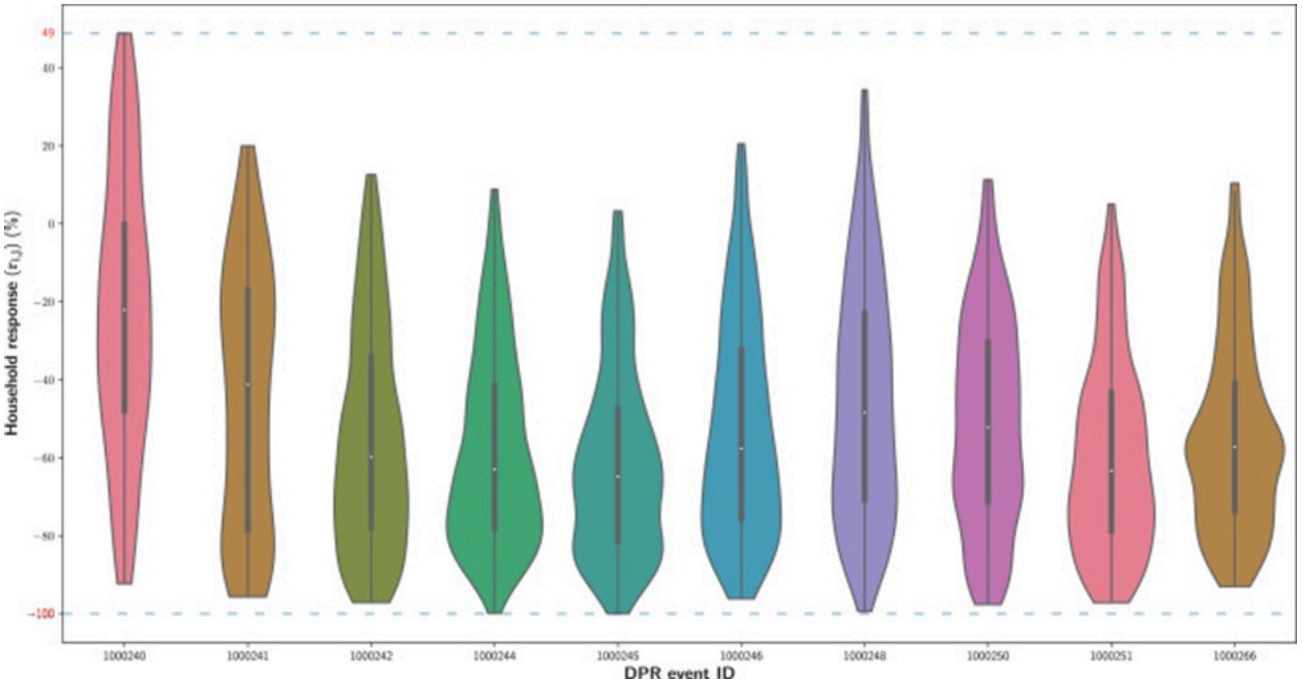

Figure 4. Violin plots showing the probability density of households' response **(Eq. 1)** for ten DPR events. The circle marker shows the median of the data and the box inside each violin plot indicates the interquartile range.

Following the anomaly removal phase, our next analytical step involved evaluating the distribution of the household responses across the array of DPR events. To ensure our evaluation yielded robust insights, we excluded consumers who responded to fewer than five out of the ten available DPR events, establishing a baseline participation threshold. As demonstrated in **Figure 5**, the distribution shape of consumer responses closely mirrors a log-normal distribution.

In this scenario, the response distribution exhibits positive skewness rather than symmetry. This distinct skewed pattern remains consistent across all distinct seasons and is visibly maintained when observing the aggregated data. For data exhibiting a skewed distribution, the median frequently serves as a superior indicator of central tendency. Consequently, we opted to utilize the median response per household (calculated across their respective participated DPR events) as our primary measure of central tendency instead of the mean.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

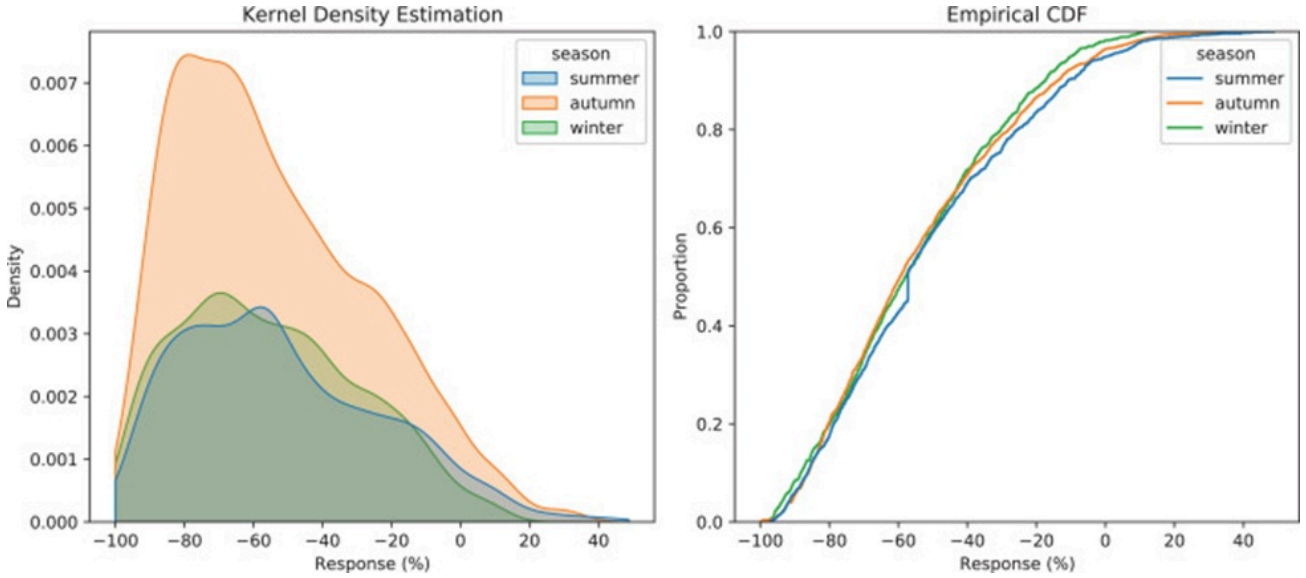

Figure 5. Distribution plots on how households have responded per season.

Temporal structure analysis

A critical proposition to investigate is whether participating households exhibit a learning curve, improving their response rates as time progresses. Validating this hypothesis holds significant implications regarding how consumers might be targeted for future DR events. If it is established that consumers progressively refine their responses over a series of events, demand aggregators should adopt a more dynamic selection strategy. For instance, aggregators might target these adapting consumers more frequently to cultivate enhanced long-term responsiveness. Evaluating this temporal progression can be approached via autocorrelation metrics.

We computed the autocorrelation function (ACF) corresponding to each household's specific DPR responses. The resulting 1-lag and 2-lag autocorrelation distributions are visualized in **Figure 6**. Initially, a noticeable subset of households appeared to present relatively high autocorrelation metrics. However, it is imperative to statistically verify that these observed population autocorrelations are genuinely non-zero, rather than mere artifacts of sampling randomness. We accomplished this verification using the Ljung-Box Q test [54], which functions as a portmanteau statistical evaluation. Rejecting the Ljung-Box Q test's null hypothesis (which posits that the data instances are independently distributed) requires p-values that fall beneath the established alpha significance threshold.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

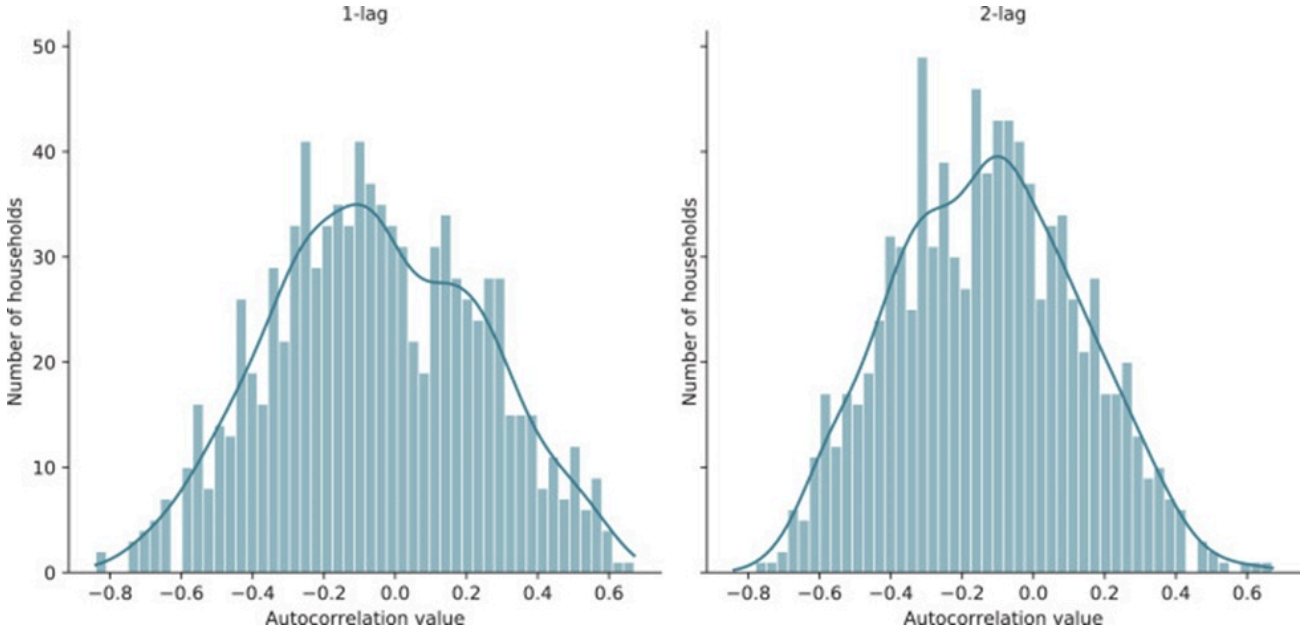

Figure 6. Distribution plots of the 1-lag (left) and 2-lag (right) autocorrelations.

Evaluating the p-values tied to the Ljung-Box Q-statistics at a 5% significance threshold reveals that out of 953 total households, merely 29 demonstrated a statistically significant 1-lag autocorrelation, and only 21 exhibited a statistically significant 2-lag autocorrelation. These findings offer minimal evidence supporting the existence of a temporal learning structure or an adaptive process over time among participants' responses to peak events. This lack of observed learning may be heavily influenced by the fact that most consumers within this dataset engaged in only a restricted number of DR events. This highlights a crucial consideration for future DR trials: executing them over more extended timeframes may provide a clearer lens through which to observe potential temporal structures and participant adaptation.

Moreover, an analysis of the household responses depicted in **Figures 5 and 7** suggests minimal variance in response behavior across different seasons. Similar uniformity is observed across specific days—noting that all DPR events took place during the work week across only three specific days—as well as across distinct months and times of day.

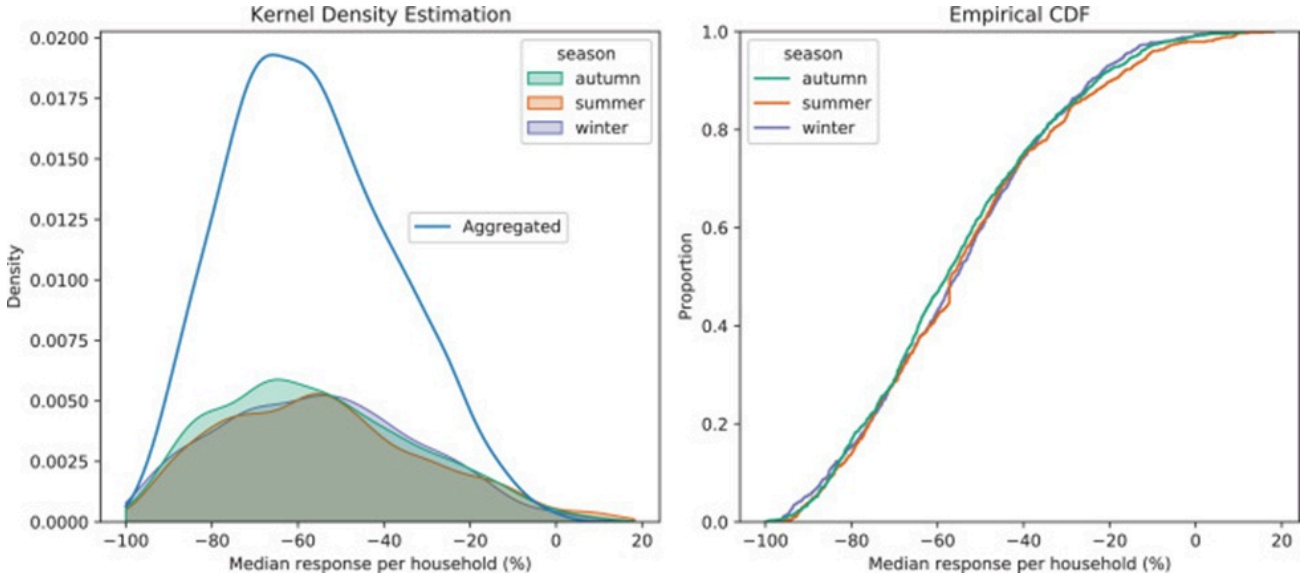

Figure 7. Distribution plots of the median response per household and its breakdown in seasons.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

Modelling of Demand Response Behaviour

In the current section, we investigate the underlying relationships between the household attributes and the calculated median household response to DPR events, ultimately isolating the most influential predictive features. As previously justified **(Figure 5)**, the median response per household was designated as our target variable owing to the skewed nature of the response distribution and the household-level granularity of our input features. The target variable's distribution is displayed in **Figure 7.** To pursue this analysis, we trained several predictive models utilizing the household dataset, specifically employing the following algorithmic approaches:

• Linear model

• Gradient Boosting Regression

• Random Forest regression

• Dense Neural Network

Our initial step involved partitioning the feature data into training, validation, and testing subsets, followed by data preprocessing. Given that our dataset heavily relies on categorical data, the preprocessing phase primarily focused on transforming these categorical variables into numerical formats utilizing embedding techniques (such as ordered target encoding or one-hot encoding). Because the numerical features naturally exhibited comparable distributions and scales, they largely bypassed further preprocessing. Additionally, since the dataset lacked missing values, no data imputation procedures were required. The subsequent paragraphs supply a more comprehensive breakdown of the deployed models alongside the specific encoding frameworks applied to each.

For the linear model, parameter estimation was executed via the widely adopted Ordinary Least Squares (OLS) technique. This specific variation of linear modeling provides excellent interpretability regarding potential linkages between the target variable and its predictors [55, 56], establishing a solid benchmark against which the more complex learning algorithms can be evaluated. In this research, we applied the statsmodels library [57] to implement the OLS linear regression, coupling it with one-hot encoding for the necessary categorical variables.

Our gradient boosting implementation utilized decision trees powered by the CatBoost framework [58]. This specialized library features a permutation-based approach that diverges from classical gradient boosting and incorporates a native embedding mechanism for categorical attributes driven by target statistics [59]. The final model's tree count was optimized by identifying the iteration that generated the peak evaluation metric across the validation set. For the Random Forest algorithm, we utilized the default parameter configuration available within the scikit-learn library [60], mirroring the gradient boosting preprocessing methodology.

Finally, we engineered a dense neural network featuring two hidden layers utilizing the Keras API operating atop the TensorFlow framework [61]. The network's core hyperparameters—specifically the Adam optimizer's learning rate [62] and the unit count per layer—were meticulously tuned utilizing the random search methodology [63] provided by the Keras Tuner library [64]. While we explored "deeper" network configurations featuring more extensive layering, they failed to yield predictive performance enhancements and were consequently excluded from the final model selection. Referencing the comprehensive survey conducted by Hancock and Khoshgoftaar [65], One-Hot encoding serves as the prevalent categorical encoding standard within neural network applications. Nevertheless, to preserve consistency and ensure comparability across our evaluated models, we opted to maintain the CatBoost encoder utilized in the preceding non-linear models.

To rigorously assess their generalized performance and provide robust test error metrics, the diverse machine learning architectures were trained and subsequently validated across 10 distinct dataset splits (comprising training, validation, and test segments). Evaluation on the test set utilized the following metrics: mean absolute percentage error (MAPE),

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

mean absolute error (MAE), and root mean squared error (RMSE). These aggregated results are detailed in **Table 2**. Readers seeking granular insights into the specific codebase implementations or the precise optimum hyperparameter configurations are encouraged to review the provided Jupyter notebooks hosted on the GitHub repository.

Table 2. Average values of evaluation metrics (in the test set) for various learning models.

Predictive model	Root mean square error (RMSE)	Mean absolute error (MAE)	Mean absolute percentage error (MAPE)	Coefficient of determination (R^2)
Linear Regression Model	18.769 ± 0.794	15.284 ± 0.625	0.465 ± 0.099	0.009 ± 0.027
Gradient Boosting Regressor	18.711 ± 0.714	15.246 ± 0.631	0.473 ± 0.101	0.015 ± 0.018
Random Forest Regressor	19.721 ± 1.611	16.107 ± 1.389	0.484 ± 0.101	−0.108 ± 0.240
Deep Neural Network (Dense Architecture)	18.906 ± 0.641	15.469 ± 0.663	0.484 ± 0.101	−0.008 ± 0.063

Analyzing the aggregate performance across the varied error metrics reveals that the Gradient Boosting algorithm delivered the most optimal predictive results on the test set, with the linear model functioning as the immediate runner-up. Overall, performance variances across these four algorithms (encompassing both linear and non-linear frameworks) proved minimal and aligned closely with established academic literature. As a benchmark, the MAPE metric mapping individual customer DR responsiveness typically hovers in the 41%–45% range [32, 33]. Consistently across all models, the MAPE remained notably low, leading to the deduction that these specific household features lack strong predictive authority concerning the median response. However, even when input variables explain only a marginal fraction of a model's variance, they should not automatically be discarded as analytically useless. The core objective of this study is to explore feature relationships and relative feature importance rather than optimize absolute predictive accuracy. When independent variables prove statistically significant, they still facilitate critical insights regarding the interrelationships governing consumer behavior.

Reviewing the summary outputs associated with the linear model highlights an in-sample ≈10%, alongside a markedly low F-statistic p-value. In the context of multiple regression, this specific condition confirms the presence of at least one statistically significant feature, effectively establishing a definitive relationship between the evaluated household attributes (or a select subset) and the median DR response. Diving deeper into the t-test p-values associated with the individual variables reveals the following specific features as statistically significant at the 95% confidence threshold:

- The level of a household's dryer usage
- If a household has internet access
- If the household is using a gas oven
- If the household is using gas for hot water
- If the household has air conditioning
- The number of a household's controlled loads

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

• The number of refrigerators in a household

• Whether the household stayed on the trial for the duration

The indicators listed above offer concrete evidence validating a statistically significant—though admittedly subtle—correlation linking these specific input variables to a household's median DR response. Armed with this validation, we push our analysis further by isolating these pivotal features and unpacking their specific contributions to shifting consumer behaviors within the framework of an incentive-based DR program.

Identification of important features and discussion of their effects on response behaviour

Establishing a connection between household attributes and demand response (DR) behavior naturally leads to investigating which specific characteristics most heavily influence consumer participation. Pinpointing these critical features not only enhances the accuracy of response predictions but also deepens our understanding of the underlying data and the predictive models utilized.

Within this research, we employed two prevalent model-agnostic techniques to assess feature importance: the permutation feature importance method and the SHAP framework. To minimize bias and generate a more decorrelated feature set, we implemented these methods across distinct models. Specifically, the permutation algorithm was applied to the linear regression and random forest models. On this importance graph, the x-axis quantifies feature significance (the variance in the model's performance metric with and without the permuted feature). At the same time, the y-axis displays the features ordered by their magnitude of importance. Conversely, the SHAP technique was applied to the dense neural network and gradient boosting models. Here, a feature's importance is calculated as the mean absolute Shapley value derived from the test data.

Utilizing these two interpretation frameworks across four disparate models generated four unique feature importance rankings. To reconcile these variations and formulate a unified hierarchy of critical features, we applied an ensemble feature ranking technique based on a Markov Chain rank aggregation method [66].

Grasping the underlying mechanisms driving consumer behavior in DR programs profoundly impacts the deployment of demand-side flexibility services. A comprehensive understanding of residential response patterns empowers service providers to optimize household targeting, refine selection methodologies for specific DR events, and architect more equitable, effective DR initiatives that foster high participant engagement.

Our research investigates the influence of household characteristics on residential DR response behavior through a dual-pronged approach. First, we analyze the distribution of residential responses mapped against critical household attributes. Second, we utilize the adaptable SHAP interpretability framework to evaluate how these features shape the outputs of the Gradient Boosting model. By scaling the SHAP method's localized explanations for the test set samples, we derive global model insights. We prioritized the Gradient Boosting regression model because it demonstrated superior predictive accuracy among all evaluated algorithms in this study. Applying two distinct methodologies allows us to cross-validate our findings, ultimately yielding potentially less biased insights regarding residential behavior during DR events.

Specifically, we isolate the highest-ranked features identified by the ensemble algorithm and analyze how fluctuating levels of categorical input variables impact household responsiveness. The influence of various numerical attributes on consumer behavior during DR events is visualized in **Figure 8**. For instance, homes with more refrigerators generally ignore DR requests. In contrast, a lower refrigerator count implies that the household will decrease consumption more aggressively during DR events. The justification for utilizing CDF plots is that comparing them across categories reveals

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

whether different classifications distinctly alter participant response behavior. Categories exhibiting “steeper” CDF curves suggest that those specific households typically demonstrate higher responsiveness, reducing their actual electricity usage (relative to their baseline) more significantly than households in competing categories—and vice versa.

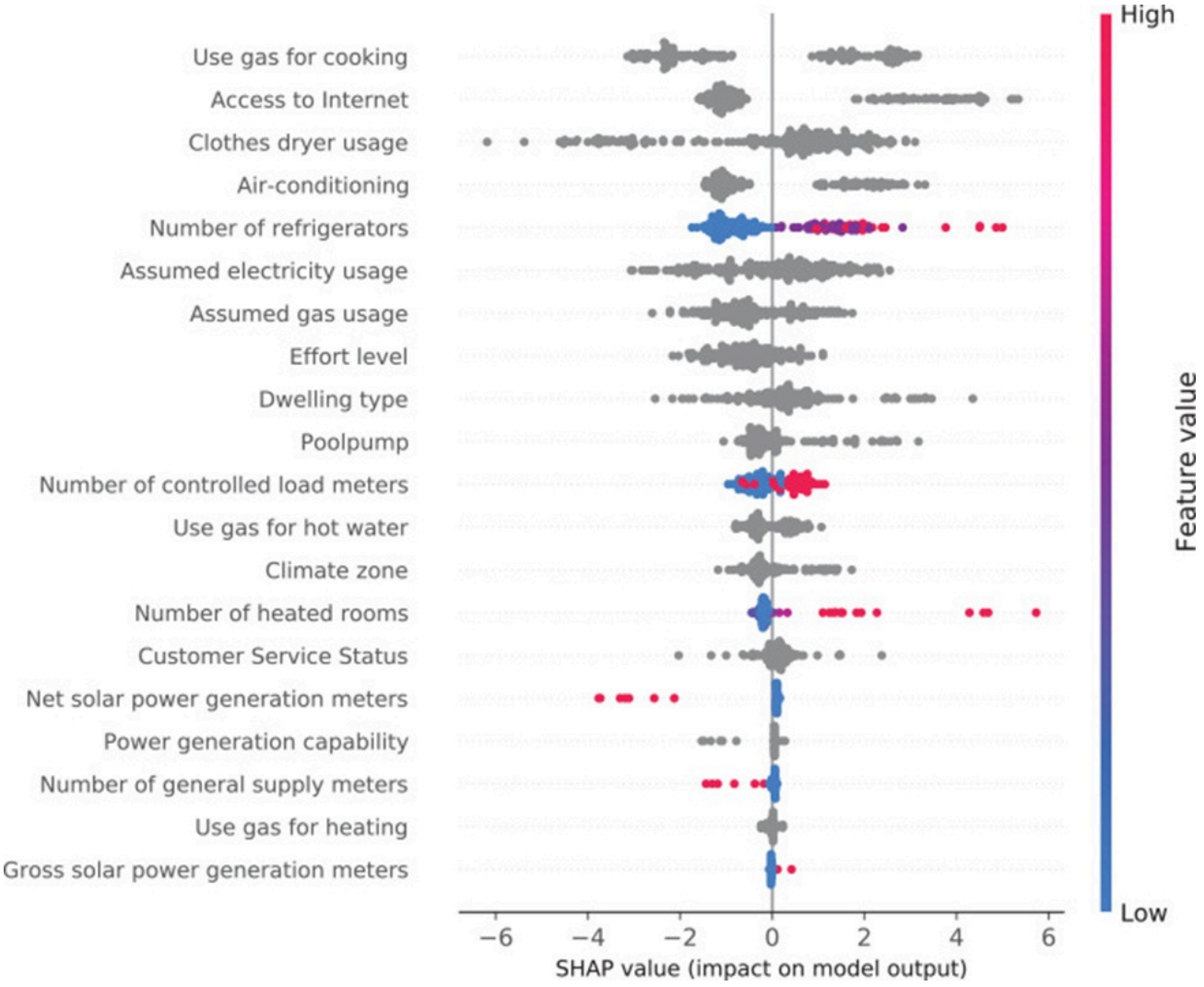

Figure 8. SHAP values of each household characteristic (feature) for every sample of the data.

Figure 8 visually articulates each feature’s contribution to the Gradient Boosting regression model. Every data point within these graphs represents a Shapley value associated with a specific feature for a single model prediction. **Figure 8** merges feature importance with feature effects. The variables are ordered by importance magnitude, with color gradients reflecting the feature’s value transitioning from low to high. Categorical variables are denoted by grey, as they lack an inherent high-low numerical scale. While this visualization effectively communicates how numerical household attributes influence model predictions, it offers limited insight into categorical variables. To discern the precise relationships governing categorical data, this chart shows the Shapley value of that feature alongside its categories for all instances in the test dataset. The y-axis dictates the feature’s SHAP value, indicating the degree to which knowing that feature’s value alters the model’s prediction for that specific sample.

Furthermore, because household response is mathematically defined as the percentage variance between actual and baseline consumption, negative SHAP values signify a reduction in electricity usage, translating to a superior response. The data points within this visualization are also color-coded according to the household’s electricity usage level. If an

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

interaction effect exists between the plotted feature and the overall electricity usage tier, it emerges visually as a distinct coloring pattern.

A rigorous review of the preceding visualizations yields several critical insights regarding how household traits influence DR behavior. Both analytical approaches confirm that households equipped with internet access consistently exhibit superior response rates. We deduce that internet connectivity does not directly affect DR performance; rather, it is most likely associated with specific resident characteristics. Additionally, homes featuring air-conditioning systems perform better than those without, and residences lacking a gas-fired oven typically out-respond homes that utilize one.

Another compelling observation is that medium-usage electricity consumers generally respond more effectively than their low- or high-usage counterparts. This contradicts certain prior studies in the literature suggesting that high-consumption households deliver significantly larger load reductions compared to low-use properties [7]. This discrepancy might stem from the specific structural design of this DPR peak program, which offered a constant rebate amount per unit of energy saved. Furthermore, participants residing in warmer climate zones demonstrate enhanced response behaviors compared to those in mild climate conditions. Similarly, residential units reporting medium to high levels of clothes dryer utilization exhibit stronger responses than properties that rarely or never use a dryer.

The data also indicates that consumers maintaining moderate gas consumption tend to be more responsive during peak DR events than low-gas consumers. Conversely, households exhibiting high consumption levels for both gas and electricity generally perform worse regarding DR response. The specific type of dwelling appears to exert minimal influence on a customer's response behavior. However, we can tentatively infer that consumers residing in apartments (units) characterized by medium electricity usage demonstrate superior responsiveness.

Conclusion

Maximizing the broad potential of demand-side flexibility fundamentally relies on the extensive integration of residential DR resources. Nevertheless, residential response behavior presents distinct hurdles. Developing precise models of participant behavior and cultivating a deeper comprehension of the underlying drivers can mitigate these challenges by enabling superior consumer targeting, refined selection strategies, and the design of meticulously customized DR schemes.

In this study, the authors model and evaluate DR response behavior not by relying on primary load data (which is primarily used in such studies), but by leveraging household-characteristic datasets from an incentive-based DR program. To achieve this, the research utilizes data derived from one of the most extensive open-source residential (household) DR trials available. We thoroughly explored the datasets and preprocessed them accordingly (e.g., executing multicollinearity checks, handling missing values, and performing anomaly/outlier detection and filtering). We investigated the hypothesis that participating households might learn to optimize their responses over time; however, we found no corroborating evidence to support this theory (at least concerning this specific DR scheme and the available dataset).

We successfully demonstrated a statistically significant correlation between a household's response and several specific characteristics, deploying various machine learning techniques to pursue enhanced predictive accuracy. The predictive accuracy achieved by these models aligns with expectations based on existing literature that focused exclusively on load data predictions. Future research could combine both data types (load data paired with household-related features) to potentially improve predictive precision regarding residential response behavior.

By applying two distinct model-agnostic methods across various machine learning architectures, we further explored and identified the critical features required for modeling household response behavior. Subsequently, we analyzed the direct effects of these essential features utilizing statistical distribution plots and interpretable AI methodologies grounded in coalitional game theory. Our findings indicate that households characterized by internet access, air-conditioning systems, power-intensive appliances (such as a clothes dryer), and lower gas usage generally deliver better-than-average

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

responses. Furthermore, our analysis uncovered insights that raise questions about the accuracy of self-reported consumer engagement metrics in DR schemes and the individual rationales driving customers' responses to DR signals.

Finally, taking a long-term perspective, we assert that data-driven analyses like ours offer vital guidance for better designing future residential DR trials, ensuring the gathered data successfully fills existing knowledge gaps for practitioners and researchers. For instance, in this study, we were unable to detect a temporal trend indicating a progressive improvement in household responsiveness as they participated in more DR events over time. This limitation is likely due to the restricted volume of DR events within the SGSC trial dataset (post-filtration, we found only 10 events yielded reliable, high-quality, usable data). Another critical question that this data cannot resolve is the elasticity of the expected percentage of demand reduction (response) with respect to the magnitude of the financial reward offered by the network operator or DR aggregator. Moving forward, we expect answering this question will be paramount in designing, optimizing, and budgeting future residential DR schemes.

Overall, with the ubiquitous roll-out of smart meters and the large-scale availability of consumer-level response data, we argue that data-driven methodologies will play an increasingly pivotal role in formulating fair and successful DR frameworks that maintain high participant engagement.

Acknowledgements

The authors would like to acknowledge the support of the Energy Technology Partnership Scotland (ETP) through their Industry Doctorates scheme and our industrial sponsor Upside Energy. The work was also supported by the UK Engineering and Physical Sciences Council (EPSRC) through the UK National Centre for Energy Systems Integration (CESI) [EP/P001173/1] Community Energy Demand Reduction in India (CEDRI) [EP/R008655/1] and by Innovate UK through the Responsive Flexibility (ReFlex) project [ref:104780].

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 11 Oct 2024 Revised: 17 Dec 2024 Accepted: 03 Jan 2025
Published online: 15 June 2026

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Eid C, Codani P, Perez Y, Reneses J, Hakvoort R. Managing electric flexibility from Distributed Energy Resources: a review of incentives for market design. Renew Sustain Energy Rev. 2016;64:237-47.

2. Luo X, Wang J, Dooner M, Clarke J. Overview of current development in electrical energy storage technologies and the application potential in power system operation. Appl Energy. 2015;137:511-36.

3. IEA. Demand response (More efforts needed). Paris: IEA; 2020. Report No.: Tech. Rep.

4. Faruqui A, Sergici S. Household response to dynamic pricing of electricity: a survey of 15 experiments. J Regul Econ. 2010;38(2):193-225.

5. Jessoe K, Rapson D. Knowledge is (Less) power: experimental evidence from residential energy use. Am Econ Rev. 2014;104(4):1417-38.

6. Newsham GR, Bowker BG. The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: a review. Energy Policy. 2010;38(7):3289-96.

7. Herter K. Residential implementation of critical-peak pricing of electricity. Energy Policy. 2007;35(4):2121-30.

8. Liu D, Sun Y, Qu Y, Li B, Xu Y. Analysis and accurate prediction of user's response behavior in incentive-based demand response. IEEE Access. 2019;7:3170-80.

9. Ma Z, Xie J, Li H, Sun Q, Si Z, Zhang J, et al. The role of data analysis in the development of intelligent energy networks. IEEE Netw. 2017;31(5):88-95.

10. Department of the Environment and Energy. Smart-grid smart-city customer trial data. 2015. Available from: https://data.gov.au/dataset/ds-dga-4e21dea3-9b87-4610-94c7-15a8a77907ef/details.

11. Langham E, Downes J, Brennan T, Fyfe J, Mohr S, Rickwood P, et al. Smart grid, smart city customer research report. Sydney: Institute for Sustainable Futures as part of the AEFI consortium for Ausgrid and EnergyAustralia; 2014. Report No.: Tech. Rep.

12. Motlagh O, Foliente G, Grozev G. Knowledge-mining the Australian smart grid smart city data: a statistical-neural approach to demand-response analysis. In: Springer International Publishing; 2015. p. 189-207.

13. Kotu V, Deshpande B. Chapter 3 – data exploration. In: Kotu V, Deshpande B, editors. Predictive analytics and data mining. Boston: Morgan Kaufmann; 2015. p. 37-61.

14. Schafer JL. Analysis of incomplete multivariate data. 1st ed. New York: CRC press; 1997.

15. White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. Stat Med. 2010;29(28):2920-31.

16. Zhang S, Zhang C, Yang Q. Data preparation for data mining. Appl Artif Intell. 2003;17(5–6):375-81.

17. Kotu V, Deshpande B. Chapter 2 – data mining process. In: Kotu V, Deshpande B, editors. Predictive analytics and data mining. Boston: Morgan Kaufmann; 2015. p. 17-36.

18. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29(5):1189-232.

19. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer-Verlag; 2009.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

20. Breiman L. Random forests. Machine Learn. 2001;45(1):5-32.

21. Russel S, Norvig P. Artificial intelligence a modern approach. 3rd ed. 2012.

22. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016.

23. Kim B, Khanna R, Koyejo OO. Examples are not enough, learn to criticize! Criticism for Interpretability. In: Lee D, Sugiyama M, Luxburg U, Guyon I, Garnett R, editors. Advances in neural information processing systems. Vol. 29. Curran Associates, Inc.; 2016. p. 2280-8.

24. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in neural information processing systems. Vol. 30. Curran Associates, Inc.; 2017. p. 4765-74.

25. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics. 2010;26(10):1340-7.

26. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinform. 2007;8(1):25.

27. Chalkiadakis G, Elkind E, Wooldridge M. Computational aspects of cooperative game theory. Synth Lect Artif Intell Mach Learn. 2011;5(6):1-168.

28. Antonopoulos I, Robu V, Couraud B, Kirli D, Norbu S, Kiprakis A, et al. Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review. Renew Sustain Energy Rev. 2020;130(109899):35.

29. Zhou D, Balandat M, Tomlin C. A Bayesian perspective on Residential Demand Response using smart meter data. In: 2016 54th annual allerton conference on communication, control, and computing (Allerton). IEEE; 2016. p. 1212-9.

30. Paterakis NG, Tascikaraoglu A, Erdinc O, Bakirtzis AG, Catalao JPS. Assessment of demand-response-driven load pattern elasticity using a combined approach for smart households. IEEE Trans Ind Inform. 2016;12(4):1529-39.

31. Holtschneider T, Erlich I. Optimization of electricity pricing considering neural network based model of consumers' demand response. In: 2013 IEEE Computational Intelligence Applications in Smart Grid (CIASG). IEEE; 2013. p. 154-60.

32. Kang J, Lee S. Data-driven prediction of load curtailment in incentive-based demand response system. Energies. 2018;11(2905):11.

33. Cai L, Wen H, Gu J, Ma J, Jin Z. Forecasting customers' response to incentives during peak periods: a transfer learning approach. Int Trans Electr Energy Syst. 2020;30(7):e12251.

34. Varghese AC, P V, Kumar G, Khaparde SA. Smart grid consumer behavioral model using machine learning. In: 2018 IEEE innovative smart grid technologies – Asia (ISGT Asia). IEEE; 2018. p. 734-9.

35. Haben S, Singleton C, Grindrod P. Analysis and clustering of residential customers energy behavioral demand using smart meter data. IEEE Trans Smart Grid. 2016;7(1):136-44.

36. Aïd R, Possamaï D, Touzi N. Optimal electricity demand response contracting with responsiveness incentives. 2019. arXiv:1810.09063.

37. Élie R, Hubert E, Mastrolia T, Possamaï D. Mean field moral hazard for optimal energy demand response management. Math Financ. 2021;31(1):399-473.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

38. Meir R, Ma H, Robu V. Contract design for energy demand response. In: Proceedings of the twenty-sixth international joint conference on artificial intelligence, IJCAI-17. 2017. p. 1202-8.

39. Ma H, Robu V, Li N, Parkes DC. Incentivizing reliability in demand-side response. In: IJCAI 2016: 25th international joint conference on artificial intelligence. 2016. p. 352-8.

40. Ma H, Parkes DC, Robu V. Generalizing demand response through reward bidding. In: Proceedings of the 16th international conference on autonomous agents and multiagent systems. ACM; 2017. p. 60-8.

41. Kota R, Chalkiadakis G, Robu V, Rogers A, Jennings NR. Cooperatives for demand side management. In: Proceedings of the 20th European conference on artificial intelligence. 2012. p. 969-74.

42. Norbu S, Couraud B, Robu V, Andoni M, Flynn D. Modelling the redistribution of benefits from joint investments in community energy projects. Appl Energy. 2021;287:116575.

43. Robu V, Vinyals M, Rogers A, Jennings NR. Efficient buyer groups with prediction-of-use electricity tariffs. IEEE Trans Smart Grid. 2018;9(5):4468-79.

44. Robu V, Gerding EH, Stein S, Parkes DC, Rogers A, Jennings NR. An online mechanism for multi-unit demand and its application to plug-in hybrid electric vehicle charging. J Artif Intell Res. 2013;48:175-230.

45. Stein S, Gerding E, Robu V, Jennings NR. A model-based online mechanism with pre-commitment and its application to electric vehicle charging. In: Proc of the 11th int conference on autonomous agents and multiagent systems - Volume 2. AAMAS '12. 2012. p. 669-76.

46. Cramér H. The two-dimensional case. Princeton University Press; 1946.

47. Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in regression analyses conducted in epidemiologic studies. Epidemiology. 2016;6(2):227.

48. Basu I, Maji S. Multicollinearity correction and combined feature effect in shapley values. 2020. arXiv:2011.01661.

49. Nolan S, O'Malley M. Challenges and barriers to demand response deployment and evaluation. Appl Energy. 2015;152:1-10.

50. Weng Y, Rajagopal R. Probabilistic baseline estimation via Gaussian process. In: 2015 IEEE Power & Energy Society General Meeting. IEEE; 2015. p. 1-5.

51. Breunig MM, Kriegel HP, Ng RT, Sander J. LOF: identifying density-based local outliers. SIGMOD Rec. 2000;29(2):93-104.

52. Liu FT, Ting KM, Zhou ZH. Isolation-based anomaly detection. ACM Trans Knowl Discov Data. 2012;6(1).

53. Cheng Z, Zou C, Dong J. Outlier detection using isolation forest and local outlier factor. In: Proceedings of the conference on research in adaptive and convergent systems. New York: Association for Computing Machinery; 2019. p. 161-8.

54. Ljung GM, Box GEP. On a measure of lack of fit in time series models. Biometrika. 1978;65(2):297-303.

55. Escanciano JC, Lobato IN. An automatic Portmanteau test for serial correlation. J Econ. 2009;151(2):140-9.

56. Murphy KP. Machine learning: a probabilistic perspective. MIT Press; 2012.

57. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. In: 9th Python in science conference. 2010. p. 92-6.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

58. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. 2019. arXiv:1706.09516.

59. Micci-Barreca D. A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. SIGKDD Explor Newsl. 2001;3(1):27-32.

60. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-30.

61. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: large-scale machine learning on heterogeneous systems. 2015. Available from: https://www.tensorflow.org/.

62. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Bengio Y, LeCun Y, editors. 3rd International Conference on Learning Representations (ICLR 2015); 2015 May 7-9; San Diego, CA, USA. 2015. https://doi.org/10.48550/arXiv.1412.6980.

63. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res. 2012;13(null):281-305.

64. O'Malley T, Bursztein E, Long J, Chollet F, Jin H, Invernizzi L, et al. Keras Tuner. 2019

65. Hancock JT, Khoshgoftaar TM. Survey on categorical data for neural networks. J Big Data. 2020;7(1):28.

66. Dwork C, Kumar R, Naor M, Sivakumar D. Rank aggregation methods for the web. In: Proceedings of the 10th international conference on world wide web. New York: Association for Computing Machinery; 2001. p. 613-22.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press