Utilities Policy 79 (2022) 101446

Contents lists available at ScienceDirect

Utilities Policy

journal homepage: www.elsevier.com/locate/jup

Full-length article

Predicting residential electricity consumption patterns based on smart meter and household data: A case study from the Republic of Ireland

Zhifeng Guo [a], Jesse R. O'Hanley [a,b,*], Stuart Gibson [c]

[a] Kent Business School, University of Kent, Canterbury, UK
[b] Centre for Logistics and Heuristic Optimisation, University of Kent, Canterbury, UK
[c] School of Physical Sciences, University of Kent, Canterbury, UK

ARTICLE INFO

Keywords:
Residential electricity consumption
Household load profiles
Machine learning

ABSTRACT

We use machine learning algorithms to investigate various aspects of residential electricity consumption for households in the Republic of Ireland. Temperature, day of week, and month of year have an apparent causal effect on consumption. The prevalence of six distinct intra-day load profiles, identified by clustering, changes dramatically between weekdays and weekends as well as seasonally. Key socio-demographic and dwelling characteristics associated with annual load profiles include household makeup and size and occupation of the primary income earner. We further discuss policy and management implications of our findings and propose avenues for future research.

1. Introduction

Reducing nonessential energy consumption, including residential electricity, is widely considered one of the most effective and low-cost means of mitigating global climate change and enhancing sustainable development, particularly in developed economies. The residential sector accounts for a significant portion of overall electricity demand – 26.9% worldwide in 2018 (IEA 2020) – with consumption patterns largely reflective of household characteristics (Ürge-Vorsatz et al., 2015; Wang et al., 2018). With increased population growth and rising living standards, residential electricity consumption has continued to grow over the decades. Gaining a better understanding of household electricity consumption patterns, behaviors, and household characteristics that influence consumption should prove enormously useful to policy-makers and electricity generation firms in developing effective strategies to reduce residential electricity consumption.

Recent advances in big data and machine learning algorithms come at an ideal time to help address this challenge. It is now possible to digitally store massive amounts of electricity usage data. Advanced machine learning methods, in turn, provide ideal tools for analyzing these large datasets, opening up new avenues to explore electricity consumption patterns in greater depth. For example, variations in electricity consumption over time can be analyzed at almost any level of granularity, from yearly to monthly, daily, hourly, or even minute-by-minute (Crone and Kourentzes 2009). As such, it becomes possible to identify how peaks and troughs in demand evolve and begin to delineate typical household energy consumption patterns or load profiles. Here, advanced machine learning techniques are ideally suited for categorizing and predicting household load profiles by identifying key variables driving energy consumption (i.e., feature selection), which can subsequently form the basis for more accurate electricity demand forecasting and support demand side management (Koprinska et al., 2015).

Uncovering behavior patterns hidden in typical load profiles is challenging due to the lack of a full mapping between behavior and electricity consumption data. In earlier studies, researchers often investigated behavior through questionnaires and then applied traditional statistical models to try to find any relationships. In the internet era, it is possible to gain fresh insights into different aspects of residential electricity consumption behaviors. In particular, it is now feasible to combine social media network data with energy usage data and then analyze the data using flexible neural network structures to reveal relationships between behavior and residential power demand.

Although there is considerable research on household electricity consumption, there are some apparent deficiencies. For one, most studies fail to comprehensively explore relationships between residential electricity patterns and interactions among key drivers, including time-related factors (e.g., season, month, day, hour) (McLoughlin et al., 2015), air quality-related factors (e.g., particulate matter, ground-level

* Corresponding author. Kent Business School, University of Kent, Canterbury, CT2 7FS, UK.
E-mail address: j.ohanley@kent.ac.uk (J.R. O'Hanley).

https://doi.org/10.1016/j.jup.2022.101446
Received 30 June 2022; Received in revised form 24 October 2022; Accepted 24 October 2022
Available online 3 November 2022

ozone) (He et al., 2020), weather-related factors (e.g., rainfall, temperature, humidity) (Li et al., 2019), and economic factors (e.g., GDP, unemployment, inflation) (Okoligwe and Ihugba 2014). Another limitation is that researchers often investigate energy consumption patterns and behaviors from a static point of view. Energy consumption patterns may vary dynamically through different time intervals. For example, in temperate regions, daytime during winter is much shorter than it is in summer, which directly impacts heating needs and people's daily activities. A third critique of the existing literature is a general failure to associate electricity usage with household data and historical usage. Both household characteristics and historical electricity usage are likely to influence future usage strongly. Most studies focus solely on characterizing consumption patterns according to dwelling and socio-economic data, while ignoring recent historical usage (e.g., over the past few months).

This study attempts to fill these identified gaps. In particular, using smart meter data from the Republic of Ireland, we examine the effect of temperature, day of week, and month on electricity usage with the help of a generalized additive model. Second, using k-means clustering to extract different intra-day load profiles, we examine how household electricity consumption patterns vary annually and weekly. Third, after employing k-medoids clustering to generate annual load profiles, we apply an elastic net model to predict annual household consumption patterns based on household-level data and recent past electricity consumption to generate useful user-profile information.

The remainder of this paper is structured as follows. In the next section, we describe the data used in this study and our methodology for analyzing residential electricity consumption patterns. We then present main results and findings. We conclude with a critical discussion of the implications of our findings and propose avenues for future research.

2. Literature review

Existing studies on this general topic of residential electricity demand profiling can be grouped into several different research themes. The first focuses on obtaining typical load profiles using different levels of granularity for electricity consumption data (i.e., daily, hourly, and minute) and different types of clustering algorithms, such as k-means (Hartigan and Wong 1979), self-organizing maps (SOM) (Kohonen 1990), and ensemble clustering (Yu et al., 2012). For example, Räsänen et al. (2010) propose using SOM combined with k-means and hierarchical clustering for handling large time-series datasets. Their methodology is applied to year-long hourly electricity consumption data from North Savo, Finland. Experimental results indicate their proposed methodology produced more accurate household load profiles than existing ones used by an electric power company to estimate load.

Meanwhile, Khan et al. (2016) present an incremental density-based ensemble clustering method for segmenting factories according to electricity consumption data. Using data from manufacturing factories in Guangdong Province, China, the authors find that their algorithm outperforms several state-of-the-art data stream clustering algorithms. Other relevant studies include those by Benítez et al. (2016) and Chévez et al. (2017). Benítez et al. (2016) present a Hausdorff distance-based dynamic clustering algorithm for identifying and visualizing temporal load profiles. Compared to traditional clustering methods like k-means and Fuzzy c-means, the proposed method produces more well-defined and balanced clusters. Chévez et al. (2017) use k-means to detect homogeneous areas of residential electricity consumption and associated socio-demographic characteristics.

A second line of research examines the importance of feature selection prior to carrying out data clustering. Feature selection is critical in the big data era as high-dimensional electricity consumption data become more widely available with the introduction of smart meters. With high-dimensional data, some traditional clustering algorithms are no longer helpful. Hence, feature selection becomes key to tackling challenges associated with analysis of high-dimensional data. An illustrative example is a study by Räsänen and Kolehmainen (2009), who present a feature selection approach for clustering time series based on extracting statistical features within time series. Advantages of their approach include dimensionality reduction of the original time series, increased robustness to missing observations, and the ability to handle time series of different lengths. Motlagh et al. (2019) proposed a baseline feature-based clustering algorithm to alleviate the limitations of extreme dimensionality of load time series by converting load time series into map models that can be readily clustered. Choksi et al. (2020) proposed a feature-based clustering algorithm aimed at dimensionality reduction, load profile characterization, and probabilistic load variation assessment by combining classical k-means with empirical feature selection.

A third area of research on household load profiles concerns the relationship between household-level data and electricity consumption patterns. For example, McLoughlin et al. (2012) use a multiple linear regression model to estimate total electricity consumption, maximum demand, load factor, and time of use of maximum electricity demand based on different dwelling types and occupant socio-economic variables. Beckel et al. (2014) examine the feasibility of inferring household characteristics, like employment status and number of occupants, from smart meter data. They were able to achieve 70% or more accuracy for each household characteristic using supervised machine learning techniques. More relevant to our work, Singh et al. (2019) apply k-means clusters to segment customers based on electricity consumption metrics, socio-demographic/economic and dwelling characteristics, and georeferenced weather data to improve peak and off-peak load predictions and support targeted demand management programs.

3. Methodology

3.1. Data

The data used in this study come from 4232 Irish households randomly selected across the country between July 14, 2009 and December 31, 2010. The dataset consists of smart meter electricity consumption data logged at 30 min intervals, along with household surveys carried out in December 2009. The data were collected as part of an electricity customer behavior trial carried out by the Irish Commission for Energy Regulation (CER 2012) and were obtained from the Irish Social Science Data Archive (https://www.ucd.ie/issda). The types of household data collected, along with summary information about dwelling, space heating, water heating, and cooker types, are provided in a series of tables in an appendix (see Table A.1-A.6). Key things to point out are that most surveyed homes (72%) are heated by oil or gas (only 6% are heated by electricity), electricity represents the single largest category for water heating (35%) and cooking (70%), and properties tend to be rather large (less than 2% are apartments and only 14% are terraced houses). After data cleaning, a sample of 3639 households remained to analyze time and climate effects on electricity consumption (see §3.2) and pattern shifting over time (see §3.3). A subset of 2874 households could be matched with household survey data to carry out a household segmentation analysis (see §3.4).

Daily temperature data from 47 weather stations across Ireland (see Figure A1) for the same time interval as the electricity monitoring data were sourced from the Irish Meteorological Service (https://www.met.ie). Due to a lack of knowledge about the location of households, hourly temperatures were averaged across all stations and used as a common temperature time series for all households. Though not ideal, average heating degree days at each station showed low variability (see Figure A1), suggesting that most households in Ireland experience similar temperatures regardless of location.

3.2. Time and climate effects on electricity usage

It is hypothesized that Irish residential electricity consumption

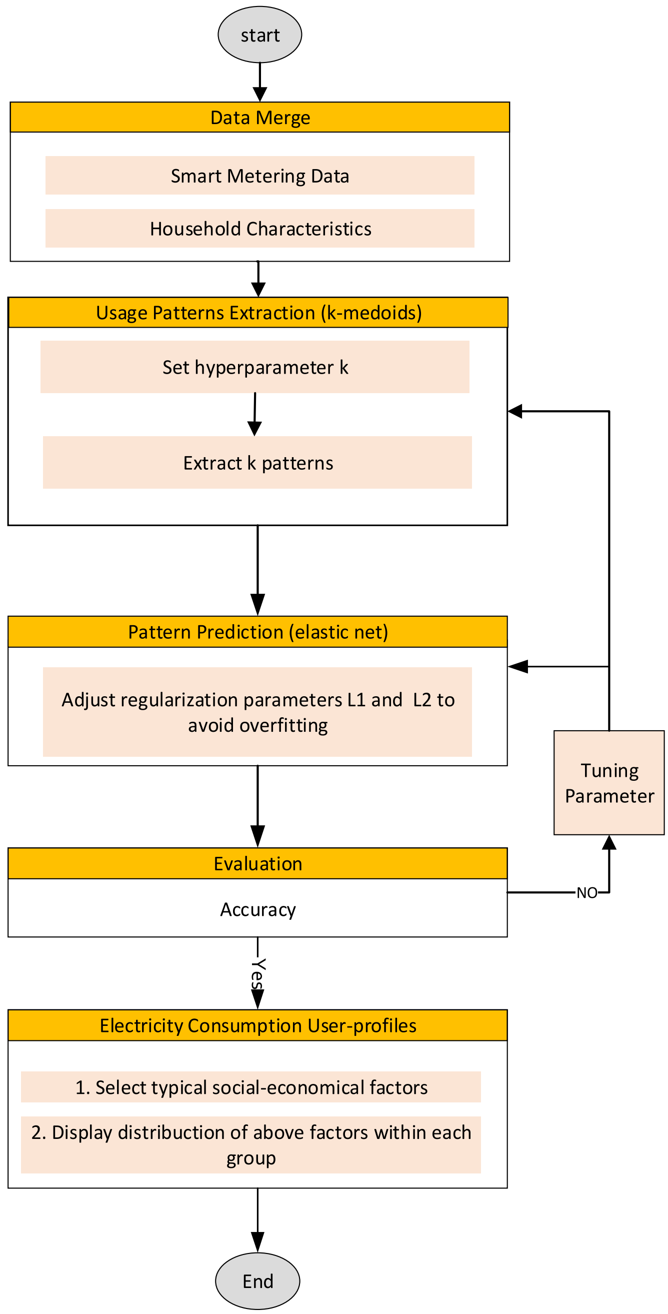

Fig. 1. Overview of how to predict annual household electricity consumption patterns.

behavior is mainly affected by seasonally varying factors like temperature and month (a proxy for daylight). In addition, electricity consumption often has clear day-of-week patterns. The effect of such influencing factors on electricity usage may be non-linear; hence we applied a generalized additive model (GAM) (Hastie and Tibshirani 1990) to study the relationship between average electricity consumption and influencing factors.

GAM is a non-parametric extension of a generalized linear model (GLM) in which non-linear smooth functions express relationships between the response and predictor variables to capture non-linearities within the data. For our analysis, we considered three predictor variables for household electricity usage (temperature, day of week, and month), which resulted in the following GAM:

$$Elect_i = \beta_0 + g_1(Temp_i) + g_2(Day_i) + g_3(Month_i) + \varepsilon_i \quad i = 1, 2, \dots, T \qquad (1)$$

where $Elect_i$ is daily usage on the i^{th} day, $Temp_i$ is average daily temperature on day i, Day_i is the day of week on day i (values 1 to 7), $Month_i$ is the month of year on day i (values 1 to 12), each $g_s(\bullet)$ is a smooth function of the corresponding covariate with thin plate regression splines as a smoothing basis, and ε_i is an error item with normal distribution. We used the penalized iteratively re-weighted least squares (PIRLS) algorithm to solve model (1).

We also consider the following linear regression model as a benchmark for model (2).

$$Elect_i = \beta_0 + \beta_1 * Temp_i + \beta_2 * Day_i + \beta_3 * Month_i + \varepsilon_i \quad i = 1, 2, \dots, T \qquad (2)$$

where $Elect_i$, $Temp_i$, Day_i, $Month_i$ and ε_i are the same as defined previously and β_0, β_1, β_2 and β_3 are linear regression coefficients. Model (2) was solved via least squares (Fahrmeir et al., 2007).

3.3. Electricity consumption pattern shifting

Residential daily electricity consumption patterns mainly depend on people's electricity consumption behavior. Behaviors are in part governed by household size and socio-economic status (e.g., number of adults and children, level of education, income, and employment status), but also seasonal and climatic factors (e.g., daylight, temperature, and precipitation). Accordingly, daily electricity consumption patterns do not remain fixed over time. To better understand this, we combine data re-aggregation with the k-means algorithm (Hartigan and Wong, 1979) to analyze changes in household intra-day electricity consumption, referred to as electricity consumption pattern shifting.

We first extracted total intra-day electricity consumption patterns for all households by taking intra-day usage time series data, re-aggregating it to a daily basis, and then applying the k-means algorithm on the re-aggregated data. The result is an assignment of each household ℓ on day i to pattern p. The number of each pattern on a particular day can then be tallied to determine the distribution of patterns and analyze how the distribution varies gradually (or not) through time to help identify underlying drivers. The proposed method is explained in more detail below.

Data re-aggregation: Before applying the k-means algorithm to extract intra-day electricity consumption patterns, it was necessary to reformat the original dataset from wide to long format, as shown in Table 1. Here, each row indicates a whole day's record of electricity usage for household ℓ ($\ell = 1, 2, \dots, n$). The first column is household index ℓ or ID, the second column denotes the day i ($i = 1, 2, \dots, T$) electricity usage was recorded, and the remaining columns indicate electricity usage c_{ℓ,i,t_j} for household ℓ on day i during a 30-min time interval t_j ($j = 1, 2, \dots, 48$).

Pattern extraction and assignment: In the pattern extraction phase, k-means is applied on columns t_1 to t_{48} in Table 1. The outputs consist of intra-daily patterns, namely the mean value of each cluster and the household-date-pattern assignment matrix (Table 2). From this, we derive the pattern-date-proportion matrix (Table 3). In Table 3, value $f_{v,d}$

Table 1
Aggregated intra-day electricity usage.

ID	Day	t_1	t_2	…	t_{48}
1	1	$c_{1,1,t_1}$	$c_{1,1,t_2}$	…	$c_{1,1,t_{48}}$
1	2	$c_{1,2,t_1}$	$c_{1,2,t_2}$	…	$c_{1,2,t_{48}}$
⋮	⋮	⋮	⋮	⋱	⋮
1	T	c_{1,T,t_1}	c_{1,T,t_2}	…	$c_{1,T,t_{48}}$
2	1	$c_{2,1,t_1}$	$c_{2,1,t_2}$	…	$c_{2,1,t_{48}}$
2	2	$c_{2,2,t_1}$	$c_{2,2,t_2}$	…	$c_{2,2,t_{48}}$
⋮	⋮	⋮	⋮	⋱	⋮
2	T	c_{2,T,t_1}	c_{2,T,t_2}	…	$c_{2,T,t_{48}}$
⋮	⋮	⋮	⋮	⋱	⋮
n	1	$c_{n,1,t_1}$	$c_{n,1,t_2}$	…	$c_{n,1,t_{48}}$
n	2	$c_{n,2,t_1}$	$c_{n,2,t_2}$	…	$c_{n,2,t_{48}}$
⋮	⋮	⋮	⋮	⋱	⋮
n	T	c_{n,T,t_1}	c_{n,T,t_2}	…	$c_{n,T,t_{48}}$

Table 2
Example household-date-pattern assignment matrix.

Day ID	1	2	3	…	T
1	p_1	p_1	p_4	…	p_5
2	p_3	p_1	p_7	…	p_2
⋮	⋮	⋮		⋱	⋮
n	p_5	p_k	p_k	…	p_6

Table 3
Pattern-date-proportion matrix.

Day Pattern	1	2	…	T
p_1	$f_{1,1}$	$f_{1,2}$	…	$f_{1,T}$
p_2	$f_{2,1}$	$f_{2,2}$	…	$f_{2,T}$
⋮	⋮	⋮	⋱	⋮
p_k	$f_{k,1}$	$f_{k,2}$	…	$f_{k,T}$

represents the proportion of each pattern p_v on day i as calculated from Table 2.

Table 3 lets us easily track how each pattern changes over time. To more easily visualize this, we further define a pattern shifting matrix (Table 4) and utilize it to analyze electricity consumption pattern shifting between any two days d and d'. In Table 4, value $S_{vw} \in [0,1]$ represents the shifting proportion, defined as:

$$S_{vw} = \begin{cases} \dfrac{|P_v \cap P'_w|}{|P_v|}, |P_v| > 0 \\ 0, |P_v| = 0 \end{cases} \tag{3}$$

where P_v represents the set of pattern v households on day d and P'_w represents the set of pattern w households on day d'.

Our proposed method has two main advantages over a more traditional approach of directly clustering on wide formatted household electricity usage time series data. First, it avoids clustering on high-dimensional data. Second, it is more effective at revealing changes in daily electricity consumption patterns over time. Without data re-aggregation, it is more difficult to detect meaningful intra-day patterns in long time series possessing a distinct daily cycle. Typically, only long-term cycles within the data can be extracted without re-aggregation. An alternative is to cluster on a day-by-day basis, but this presents issues regarding the large number of times clustering needs to be performed. Moreover, the patterns obtained by clustering on each day separately are unlikely to be universal. Our proposed way of clustering avoids both of these problems.

3.4. Household segmentation

We further investigated how household characteristics and past usage can be used to forecast annual electricity usage patterns on a daily interval. An important application of this analysis is the segmentation of potential new customers, enabling electricity providers and energy comparison websites to make appropriate price plan recommendations.

As mentioned previously, daily electricity consumption patterns often vary over time. On the contrary, household socio-economic descriptors tend to be relatively stable. This creates an inherent challenge whereby stable (socio-economic) features are used to predict an unstable (electricity consumption pattern) feature. To overcome this, we applied k-mediods (Reynolds et al., 2006) on household characteristics and past electricity consumption information to first cluster households around different typical yearly load profiles. Unlike k-means, k-medoids chooses actual data points as centers, which has the advantage of providing greater interpretability of the links between underlying household characteristics and consumption patterns. Having extracted a set of typical load profile patterns, we subsequently applied an elastic net model (Zou and Hastie 2005) to predict the likelihood of a household having any given load profile. Finally, we generated user-profile information for each load profile. Fig. 1 shows an overview of our methodology. The methodology consists of four distinct steps: data merging, load profile pattern extraction, load profile prediction, and generation of user-profile information. Each of these steps is discussed in more detail below.

Table 4
Pattern shifting matrix.

Pattern shifting matrix.				
d' / d	P'_1	P'_2	…	P'_k
P_1	S_{11}	S_{12}	…	S_{1k}
P_2	S_{21}	S_{22}	…	S_2
⋮	⋮	⋮	⋱	⋮
P_k	S_{k1}	S_{k2}	…	S_{kk}

3.4.1. Data merge

Household characteristics are detailed in two tables provided in an appendix (see Tables A.9-A.10). This includes type of residence, number of occupants, employment status, level of education, and social class. As part data processing, we first removed any households with missing data. Next, we split the smart meter data into time intervals: 2009-07-14 to 2009-12-31 and 2010-01-01 and 2010-12-31. Data from 2009 were treated as “historical” usage; from this, average daily usage is computed. The three categories of data – household information, 2009 historical usage, and 2010 contemporary usage – were then used in the next step for load profile pattern extraction. The first two data types were later used for load profile prediction.

3.4.2. Load profile pattern extraction

In the second step, household characteristics and historical and contemporary electricity usage data were used to segment consumers using the Partitioning Around Medoids (PAM) algorithm. Importantly, PAM can accept mixed-type distance similarity. Indeed, several household characteristics are categorical, like residence type and education level. For hyperparameter k (i.e., number of clusters), values between 3 and 6 were considered. The silhouette coefficient can be used to assess the quality of clusters based on the degree of similarity and dissimilarity among them. However, the cluster validity index tends to result in choosing a smaller number of clusters, which is not ideal for customer segmentation. The output from this step was a set of load profiles with similar household information, historical electricity usage, and contemporary electricity consumption patterns.

3.4.3. Load profile prediction

The load profile prediction step aims to classify households according to load profile patterns based on household and historical electricity usage data. For this, we used an elastic net model (Zou and Hastie 2005), which combines lasso (Tibshirani, 1996) and ridge regression (Hoerl and Kennard, 1970). Elastic net overcomes some of the limitations of the lasso model. For example, in the “large p, small n” case (i.e., high-dimensional data with few samples), lasso selects at most n variables before it saturates. If there is a group of highly correlated variables, then the lasso model tends to select one variable from the group and ignore the others. In order to overcome this, the elastic net adds a quadratic part to the penalty, also known as Tikhonov regularization, which comes from ridge regression.

We used coordinate descent to solve for the regression parameters and parameter grid search to optimize the model’s hyperparameters, namely, the number of clusters, elastic net penalty, and regularization parameter. With parameter grid search, parameter values can be determined to optimize the model’s accuracy. We also implemented a deep neural network (LeCun et al., 2012), random forest (Breiman

2001), gradient boosting machine (Friedman 2001), and support vector machine (Cortes and Vapnik 1995) as benchmarking models.

3.4.4. User-profiles

In the final step of our analysis, socio-demographic and dwelling information were analyzed for each load profile pattern. Household characteristics, including number of occupants, employment status, education level, and social class, were anticipated to be strongly linked with different load profiles. A better understanding of the linkages between household characteristics and consumption behaviors will likely help in devising more effective demand-side management programs in the electricity sector.

4. Results

Model implementation and computational experiments were carried out using Python and R on a PC with an Intel Core i5-8250U CPU running at 1.60 GHz with 8.0 GB RAM.

4.1. Influence of time and climate on electricity usage

We implemented the GAM model (2) in the R package mgcv (Wood 2017), where each smooth function $g_s(\bullet)$ is estimated by a penalized regression spline. Results of the GAM and benchmark linear regression models are displayed in Tables 5 and 6. Estimated temperature, day of week, and month curves are plotted in Fig. 2, with the shaded area representing twice the standard error bands.

For both models, we observe that all predictor variables are statistically significant at the 0.1% level. As one would expect, there is an inverse relationship between electricity demand and temperature, as indicated by the negative sign for the corresponding regression coefficient estimate. For the GAM model, the near zero *p* values for s(Temperature), s(Day), and s(Month) suggests that these smoothing terms each have a powerful influence on electricity usage.

The performance of each model is reported in Table 7. We observe that GAM produces a superior fit compared to a multiple linear regression, which is evident by the higher adjusted R^2 value (90% of variation explained by the three predictors) and lower Akaike information criterion (AIC), root mean square error (RMSE), mean absolute percentage error (MAPE) values.

Inspection of Fig. 2 reveals a clear monthly pattern to electricity usage, with above average usage in the late autumn and winter (November to February) and below average usage in the spring, summer, and early autumn (March to October). Intuitively, this would appear to be driven by behavioral responses to seasonal variation, namely reduced time at home and use of appliances in warmer months with more daylight (March to October) and greater use of appliances and increased secondary electric heating and water heating in colder months with less daylight (November to February). There is also a clear weekly pattern for residential electricity usage. Demand is noticeably higher at the weekend (Saturday and Sunday) when people are typically at home than during weekdays when they are typically at work. One interesting observation is that demand steadily declines from Monday to Friday. Why this should be the case is not clear.

Compared to a basic regression model, the GAM model is able to capture non-linearities in the temperature response function. Moreover, while month of year will partly capture temperature-driven electricity usage patterns, the use of temperature as a separate variable better accounts for daily variation in the need for water heating, secondary electric heating, and other activities. Indeed, household electricity consumption shows a near linear increase as temperature drops from 20 to 5 °C and then rapidly increases as temperatures drop below 5 °C (see Fig. 2a).

Table 5
Linear regression model results.

Parameter	Est.	Std.	t value	p-value
Intercept[a]	26.938	0.265	101.466	$<2 \times 10^{-16}$
Temperature[a]	−0.610	0.014	−43.828	$<2 \times 10^{-16}$
Day[a]	0.221	0.036	6.016	3.32×10^{-9}
Month[a]	0.263	0.022	11.916	$<2 \times 10^{-16}$

[a] Significant at the 0.1% level.

4.2. Shifting electricity consumption patterns

We identified six typical intra-day load profiles (see Fig. 3 and Table 8). Patterns 1–3 are all similar in that they show a bump in electricity usage in the morning starting around 06:00, increased electricity usage beginning at 16:00 that eventually peaks in the evening at around 19:00, presumably when dinner is being prepared, followed by a noticeable decline starting at 22:00 when people go to bed. Pattern 1 stands out from these and the rest in terms of the very low amount of electricity used in a 24-h period and low temporal variability. Pattern 6 is similar to patterns 1–3, but peaks much earlier in the late afternoon/very early evening around 16:30, presumably corresponding to an early dinner time. Patterns 4 and 5 differ considerably from the others. Pattern 4 has a prominent peak in the middle of the day at noon that then drops off but remains high in the evening before declining at 22:00. Pattern 5, meanwhile, has the highest electricity usage overall, which begins steadily rising at 05:00 in the morning until 17:30 in the early evening and then rapidly decreases.

It is worth noting that mean annual electricity consumption for the different load patterns is relatively high (8730.3 kWh), nearly double what one might expect (~4500 kWh). We suspect that this is mainly because few small dwellings (apartments and terraced houses) are included in the household survey dataset, which is probably not fully representative of Irish households more generally.

A graphical representation of the pattern-date-proportion matrix showing the number of households exhibiting each daily load profile pattern for one and a half years is displayed in Fig. 4. As can be seen, pattern 1 and pattern 3 represent the dominant load profiles among Irish households, accounting on average for 34% and 33%, respectively, over the study period. Pattern 1 shows apparent seasonal variation characterized by a summer peak in July and a winter nadir in December. In contrast, patterns 2, 5, and 6 show opposite seasonal variation, with highs in December and lows in July. More specially, we observe that the proportion of pattern 1 households reduces from 42 to 44% in July to as low as 20–21% in December, while patterns 2, 5, and 6 households collectively account for 15–17% in July to as much as 47–49% in December (see Table A7). This kind of pattern oscillation demonstrates the extent that households modify their behavior and electricity usage in response to long-term changes in temperature, amount of daylight, and other seasonal factors. Patterns 3 and 4, meanwhile, form a more constant proportion of households, albeit with high weekly and yearly volatility.

Importantly, there is a noticeable weekday versus weekend relationship to the proportion of pattern 3 and pattern 4 households, with pattern 3 being more frequent on weekdays and pattern 4 more frequent on weekends. Other patterns also have a strong weekly cycle, with patterns 2 and 6 being more prevalent on weekdays and pattern 5 on weekends. We also note a very distinctive load profile distribution on Christmas Day, with patterns 4 and 5 usually representing a small proportion of households (12% over the study period) but forming a majority or clear plurality (49–51%) on Christmas Day.

An investigation of seasonal transitions (Fig. 5a and Table A8) reveals several perhaps unexpected insights. Moving from summer (July 14, 2009) to winter (December 24, 2009), one might expect, as a result of pattern 1 being counter-cyclical with patterns 2, 5, and 6, that the decline of pattern 1 and increase of patterns 2, 5, and 6 would be simply down to pattern 1 households transitioning to those other patterns. The story is more complex than that. While a significant proportion (26%) of

Table 6
GAM model results.

Parameter	Est.	Std.	t value	Eff. DF	F value	p-value
Intercept[a]	23.935	0.051	469.3			$<2 \times 10^{-16}$
s(Temperature)[a]				4.667	62.78	$<2 \times 10^{-16}$
s(Day)[a]				4.483	23.76	$<2 \times 10^{-16}$
s(Month)[a]				9.299	69.34	$<2 \times 10^{-16}$

[a] Significant at the 0.1% level.

pattern 1 households transition to patterns 2, 5, and 6, more (27%) transition to the other dominant pattern 3. Meanwhile, the bulk of pattern 3 households (54%) shift to patterns 2, 5, and 6 between summer and winter. Thus, the significant growth in these less common patterns during the summer is driven more by pattern 3 than pattern 1 shifting. Patterns 2, 4, and 6, like patterns 1 and 3, also show a significant shift to other patterns. Only pattern 5 seems relatively stable, with 73% of households maintaining this pattern between summer and winter. The result is that winter shows a much more uniform distribution among patterns 1–6 in winter compared to summer.

Looking at weekday-to-weekend pattern shifting (see Fig. 5b and Table A9), we find that pattern 1 is by far the most stable, with only around 25% of households having a pattern 1 on a weekday transitioning to some other pattern on the weekend. In contrast, patterns 2, 5, and 6 show the most flux, with 75–85% of households having one of these patterns on a weekday and moving to a different pattern on the weekend. Patterns 3 are 4, meanwhile, are somewhat more stable between weekdays and weekends, with 60–63% of households transitioning to a different pattern on the weekend.

In terms of overall makeup (see Fig. 5b and Table A10), pattern 1 accounts for a roughly equal share of households on weekdays (43%) and weekends (45%). Pattern 4, meanwhile, increases substantially on weekends, going from 9% of households on a weekday to 18% on the weekend. Conversely, pattern 3 shows a noticeable decrease, going from 32% of households on weekdays to 25% on weekends; patterns 2, 5, and 6 show more modest decreases or increases on weekends than weekdays.

4.3. Household segmentation

4.3.1. Load profile patterns

We identified a total of 5 distinct annual electricity consumption patterns (see Fig. 6 and Table 9). All five have a similar usage profile, typified by higher (lower) usage in the winter (summer), albeit with very different amounts of average daily usage. Pattern 2 households consume the most, with average daily usage of 39.92 kWh, while pattern 5 households consume the least, with average daily usage of just 9.98 kWh. The most common usage profile is pattern 1, which accounts for nearly a third (29%) of all Irish households. Pattern 1 households consume an intermediate amount of electricity, with daily usage at 26.02 kWh.

We further investigated the proportions of different intra-day electricity consumption patterns associated with each annual load profile (Tables A.11-A.12). We observe that the five annual patterns are linked to distinct combinations of intra-day patterns that vary between summer and winter. Specifically, annual load profile 1 is majority (0.56) daily pattern 3 in summer, but formed mostly of patterns 4 and 5 (0.63) in winter. Load profile 2, on the other hand, has roughly equal proportions of patterns 2, 4, and 6 (0.18–0.21 each) and a slightly higher proportion of pattern 3 (0.28) in summer, but is majority pattern 5 (0.62) in winter. Annual load profiles 3 and 5 have in summer very high proportions of pattern 1 (0.87–0.91), a small proportion of pattern 3 (0.07–0.10), and little of the other patterns (0.02–0.03). In winter, daily pattern 1 (0.47–0.67) remains the dominant pattern, but with higher proportions of patterns 3 and 4 (0.27–0.41) and increases in the other patterns (0.06–0.13). Finally, annual load profile 4 is composed mainly of daily patterns 1 and 3 (0.90) in summer, but has similar shares of patterns 1, 3, and 5 (0.16–0.21 each) and a plurality of pattern 4 (0.33) in winter.

4.3.2. Load profile prediction

Fig. 7 displays the training and test accuracy of the five machine learning models (deep neural network (DNN), random forest (RF), gradient boosting machine (GBM), elastic net, and support vector machine (SVM)) used to classify households into one of the five annual load profiles based on household and historical electricity usage information. Parameter settings for the five machine learning models are provided in Table A.13. The elastic net model performs the best overall, with a test accuracy of nearly 85%, thus demonstrating the effectiveness of combining the L_1 and L_2 regularizations, closely followed by the random forest (RF) model with an 83% test accuracy. Despite having the best training accuracy (96%), support vector machine had the lowest test accuracy (70%), indicating it overfitted the training data substantially.

4.3.3. User-profiles

As part of our analysis, we try to elucidate the connection between a household load profile and various socio-demographic and dwelling characteristics, namely household type (live alone vs only adults vs adults and children), occupation (aka social grade) and employment status of the chief income earner, number of household appliances, number of residents, and number of bedrooms (see Table A.1-A.2). Results are summarized in Table 10 and further discussed below.

User Profile 1: This user profile represents the largest subgroup of Irish households (29%) with moderate electricity usage (26 kWh daily average). User profile 1 can generally be classified as middle-income, working adult families and more affluent retirees with moderate energy usage. Household type is predominately made up of adults only (62%), some with children (33%), and very few living alone (5%). In terms of social grade, a majority are either lower middle class (29% C1) or nonworking (34% DE), with most being employees (53%) or retirees (29%). Profile 1 users have an intermediate number of bedrooms (3–4) and intermediate number of household appliances (15–18).

User Profile 2: This subgroup of households has the highest overall electricity usage of any user profile (40 kWh daily average), including the highest Christmas Day usage peak (see Fig. 6). This user profile can be generally classed as affluent families with children and high energy usage. A majority of households have children under 15 years old (50%) and are mainly employed in higher/intermediate (26% AB) or junior (35% C1) administrative and professional roles. Unsurprisingly, this user profile is also characterized by the highest percentage of self-employed (19%), the lowest percentage of retirees (13%), the most number of appliances (15–19), and the most number of bedrooms (3–5).

User Profile 3: This subgroup of households can be mainly described as lower-income adult families and less affluent retirees with low energy usage (14 kWh daily average). Most households are composed of adults (58%) only or single adults (33%). A majority are retired (51%). Among the employed and self-employed, most households are lower middle class (20% C1) in junior administrative and professional roles. This user subgroup has the fewest household appliances (13–16) and lives in smaller-sized houses (2–4 bedrooms).

User Profile 4: Households in this subgroup are similar to user profile 3, but with higher electricity usage (20 kWh daily average). In comparison to user profile 3, households are slightly less affluent, with more lower-middle class (22% C1) and working class (18% C2) main income

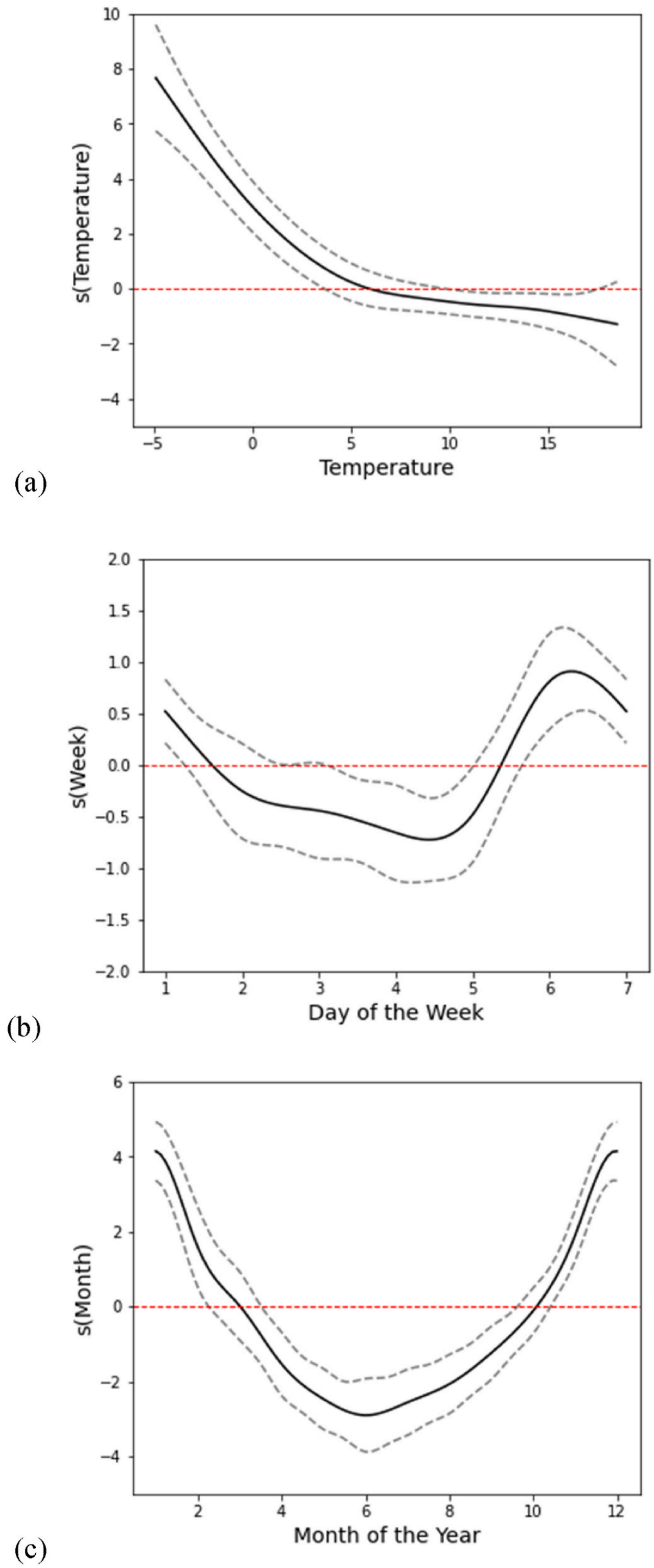

Fig. 2. Effect of temperature (a), day of week (b), and month of year (c) on residential electricity demand.

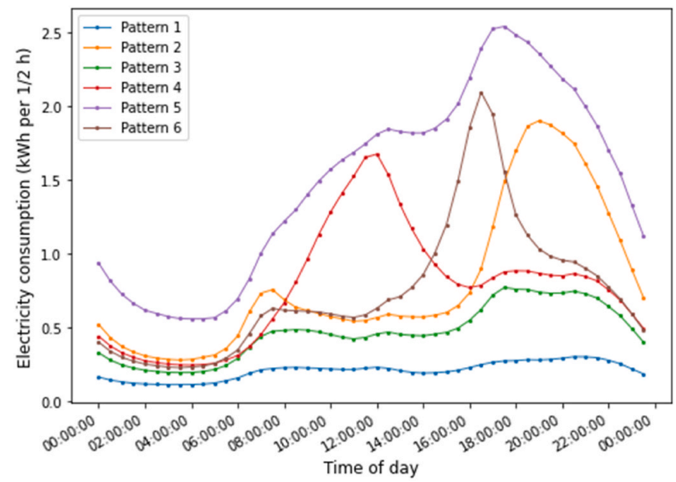

Fig. 3. Typical intra-day consumption patterns (kWh per ½h).

Table 7
Comparison of GAM and linear regression model accuracy.

Model	Adj. R^2	AIC	RMSE	MAPE
Regression	0.79	2095.48	1.69	5.14%
GAM	0.90	1720.30	1.16	3.04%

Table 8
Summary statistics for patterns 1–6 (n = 3639).

Pattern	1	2	3	4	5	6
Proportion	34.4%	9.9%	32.6%	9.8%	3.4%	9.9%
Avg. daily usage (kWh)	9.88	38.19	22.21	36.49	71.32	35.33
Peak avg. usage (kWh/ 30 min)	0.42	2.35	1.12	2.25	5.36	1.87
Max avg. ramp rate (kWh/min)	0.08	0.59	0.24	0.56	1.80	0.33

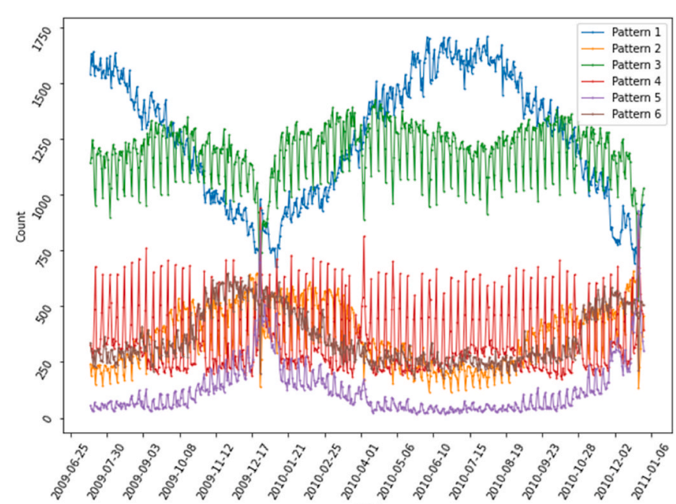

Fig. 4. Frequency of each intra-day consumption pattern over time.

earners, have fewer adults living alone (20%), more families with children (15%), and more appliances (15–18), and live in larger houses (3–4 bedrooms).

User Profile 5: The final and smallest subgroup of Irish households (13%) has the lowest energy usage (10 kWh daily average) of any user profile. This user profile is best described as middle-income single adults and more affluent retirees with low energy usage. A clear majority live alone (58%), the highest of any user profile; very few have children (12%). Nearly half are employed in higher/intermediate (13% AB) or junior (36% C1) administrative and professional roles. Households with

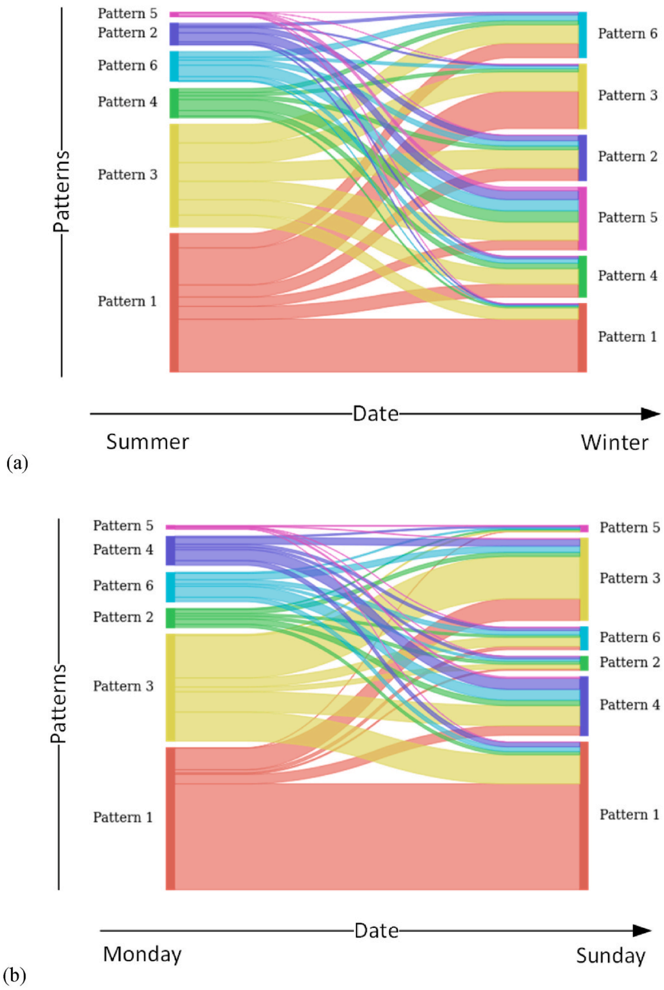

Fig. 5. Sankey diagrams of summer (July 14, 2009) to winter (December 24, 2009) to winter pattern shifting (a) and weekday (July 27, 2009) to weekend (August 2, 2009) pattern shifting (b).

Table 9
Annual electricity consumption pattern summary statistics (n = 2874).

Pattern	1	2	3	4	5
Proportion (%)	29.37	24.36	13.99	18.86	13.43
Avg. daily usage (kWh)	26.02	39.92	13.73	19.75	9.98

this user profile also have relatively few appliances (14–17) and live in smaller-sized houses (2–4 bedrooms).

In summary, household type, occupation, and employment status are all highly correlated with household electricity usage of Irish households. Each of these factors is likely to strongly influence lifestyle, including work and leisure patterns, which, in turn, result in different electricity usage profiles. Our procedure for extracting typical annual load profiles and linking this to socio-demographic and dwelling data to generate user profiles offers a simple, transparent, and effective approach to a challenging, cross-domain matching problem that combines massive smart meter data with household data to extract meaningful market segmentation information.

5. Conclusion and discussion

This study considers various aspects of electricity consumption based on an analysis of smart meter electricity data and household level characteristics, namely temporal and climate influences on electricity consumption, daily consumption pattern shifting over time, and prediction of annual load profiles. For the first of these, we established an apparent causal effect of temperature, day of week, and month on household electricity consumption. We found that the GAM model that uses smoothing functions for independent variables produced a noticeably better fit than standard linear regression (+0.11 adjusted R^2 value). More importantly, we observed that 5 °C represents a critical threshold for electricity consumption – for temperatures below 5 °C, electricity demand rapidly increases, whereas at 5 °C and above, there is a small negative effect of increasing temperature on electricity demand. This monotonically decreasing temperature sensitivity curve is notably different from the “U” shaped curve of other countries, like the US and China. Season also has a powerful influence on electricity consumption. In spring to early autumn (March and October), electricity demand is less than in late autumn and winter (November to February). Day of week also affects electricity usage, with Mondays and weekends seeing noticeably higher demand than other weekdays. It is likely that this is driven by increased leisure time at home (weekends) and increased cleaning and food preparation on the first day of the week (Monday).

As part of our electricity consumption pattern shifting analysis, we identified six distinct intra-day patterns between July 2009 and December 2010. Moving from summer to winter, household intra-day load profiles changed dramatically in shape and volume. For example, the most common pattern reduced from 42 to 44% of households in July to 21–25% in December, indicating the extent to which daily electricity consumption behaviors change over a year (i.e., higher energy use associated with secondary electric heating, water heating, and a more in-door lifestyle during colder months with less daylight versus more outdoor time and less energy use in warmer months with more daylight). In addition, we observed a clear weekday versus weekend cycle and a very noticeable spike in demand on Christmas day. We note that the data collection period corresponds to just after the 2007-08 Financial Crisis when the Irish economy was in a state of flux. Indeed, GDP in Ireland fell 5.1% in 2009 compared to 2008 (European Commission 2013). This context may have influenced observed household electricity consumption behaviors.

Finally, we assessed the performance of various machine learning models to predict annual electricity consumption patterns based on household characteristics and historical usage. The best fitting model, the elastic net model, achieved nearly 85% test accuracy for a 5-class classification problem. Further analysis of socio-demographic and dwelling characteristics associated with each load profile revealed some interesting findings. Besides household makeup and size, occupation of the main income earner (social class indicator) had a strong influence on electricity usage. It was found that those employed in higher/intermediate and junior administrative and professional roles had distinctly higher energy demand than semi-skilled and unskilled workers and the unemployed, even when children lived in the household.

These results have meaningful implications for electricity suppliers/operators and policymakers. Firstly, given the strong influence of temperature, day of week, and season on electricity usage, power suppliers and power grid operators can predict electricity demand based on climate forecasts and proactively take measures to balance supply and demand. Secondly, given the extent of daily electricity consumption pattern shifting seasonally, over a week, and on national holidays (Christmas), the use of time-of-use price schemes may go a long way to smoothing out electricity demand over a day. Finally, our market

Table 10
Primary socio-demographic and dwelling characteristics associated with each annual load profile.

User profile (proportion)	Avg. daily consumption (kWh)	Household type	Occupation (proportion)	Employment status (proportion)	No. Household appliances	No. People	No. Bedrooms
1 (29.4%)	26.0 kWh	Live alone (5%) Only adults (62%) Adults and children (33%)	AB (18%) C1 (29%) C2 (19%) DE (34%)	Employee (53%) Self-employed (12%) Unemployed (6%) Retired (29%)	15–18	2–4	3–4
2 (24.4%)	39.9 kWh	Live alone (3%) Only adults (47%) Adults and children (50%)	AB (26%) C1 (35%) C2 (18%) DE (20%)	Employee (61%) Self-employed (19%) Unemployed (6%) Retired (14%)	15–19	2–6	3–5
3 (14.0%)	13.7 kWh	Live alone (33%) Only adults (58%) Adults and children (9%)	AB (11%) C1 (21%) C2 (12%) DE (56%)	Employee (35%) Self-employed (7%) Unemployed (7%) Retired (51%)	13–16	2–3	2–4
4 (18.9%)	19.8 kWh	Live alone (20%) Only adults (65%) Adults and children (15%)	AB (8%) C1 (22%) C2 (18%) DE (52%)	Employee (35%) Self-employed (7%) Unemployed (7%) Retired (51%)	15–18	2–3	3–4
5 (13.4%)	10.0 kWh	Live alone (58%) Only adults (30%) Adults and children (12%)	AB (13%) C1 (36%) C2 (17%) DE (34%)	Employee (57%) Self-employed (8%) Unemployed (8%) Retired (27%)	14–17	2	2–4

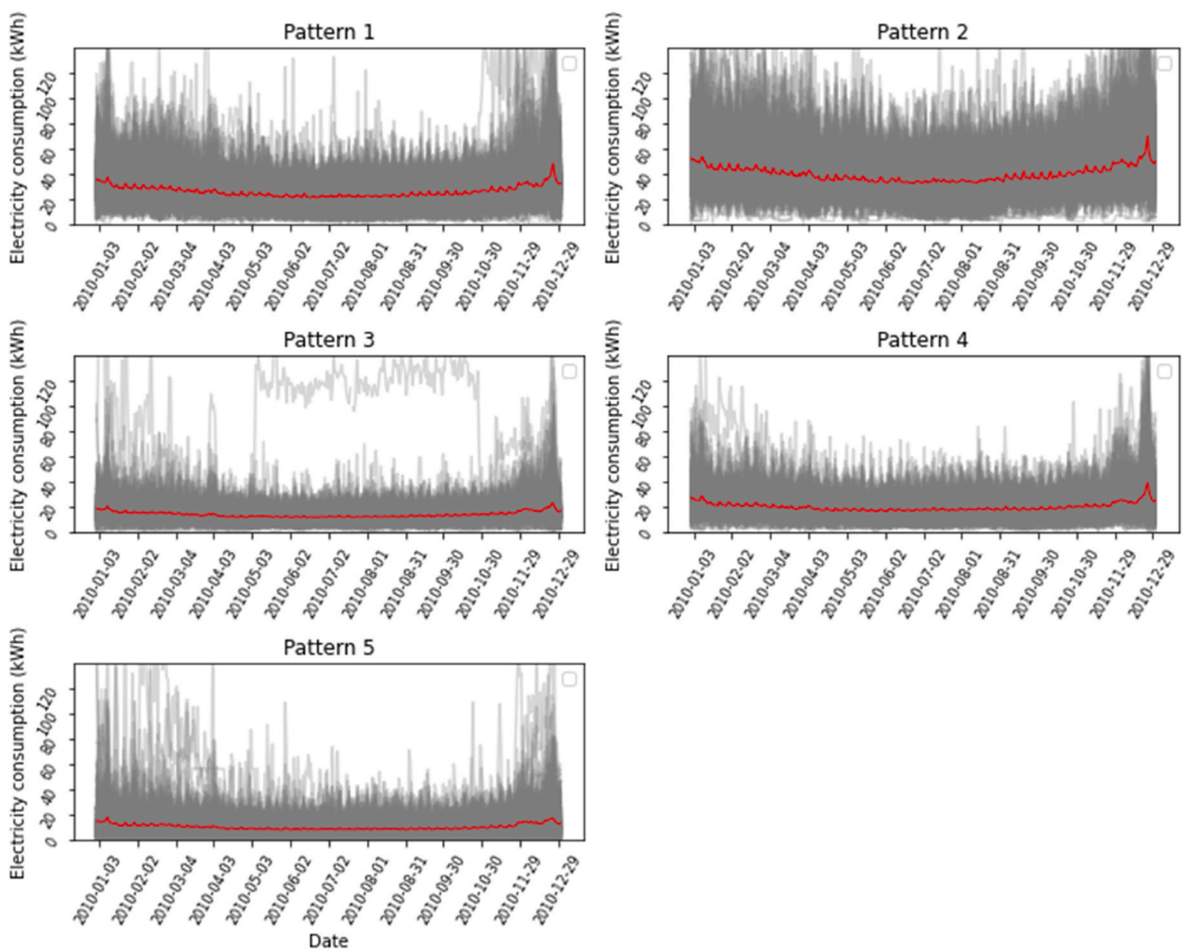

Fig. 6. Annual electricity consumption patterns on a daily basis.

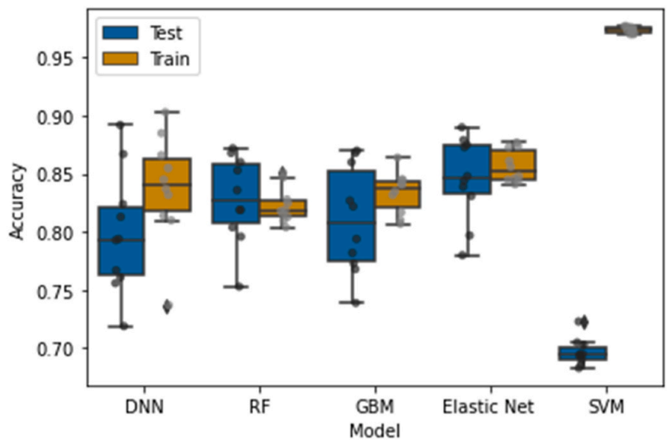

Fig. 7. Boxplot of training accuracy and test accuracy for the five machine learning models. Accuracy based on the fraction of samples (annual load profile patterns) correctly classified.

segmentation of user profiles could support more tailored energy plans for new customers based on simple household information.

In terms of potential future lines of research, extensions of our methodological approach might focus on formulating and implementing a real-time demand response system. Our current work did not examine the extent to which supply-side mechanisms household electricity consumption, which future work might address. In addition, a more robust analysis could factor in other weather variables besides just temperature, like precipitation, cloudiness, and humidity, which may affect electricity demand as well as the potential influence of government policy. Finally, our general approach could be extended to other utility sectors, like water and natural gas usage, and the commercial user sector. In short, there is ample opportunity for follow-up work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

Z. Guo was supported by a University of Kent Vice Chancellor's Research Scholarship. This support is gratefully acknowledged. The authors also extend their gratitude to two anonymous referees and the editor-in-chief for providing constructive comments on earlier drafts of this article.

APPENDIX

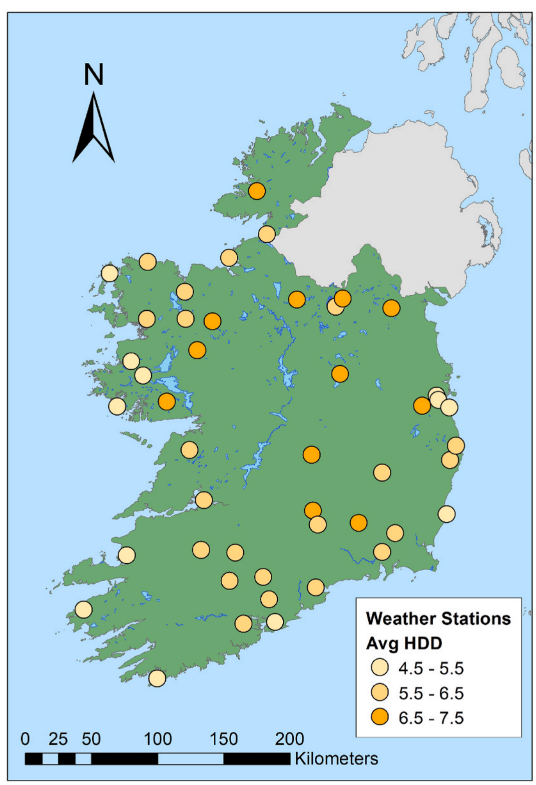

Fig. A.1. Locations and associated average heating degree days (HDD) of weather stations in the Republic of Ireland.

Table A.1
Partial list of household characteristics in the CER dataset.

Characteristic	Description	Categories
Education level	Level of education of chief income	Primary level
		Secondary level
		Third level
Employment status	Employment status of chief income earner	Employee
		Self-employed
		Unemployed
		Retired
Occupation	National Readership Survey (NRS) social grade of chief income earner	AB – upper and middle class
		C1 – lower middle class
		C2 – skilled working class
		DE – working class and nonworking
Internet access	Internet access availability	Yes
		No
Household type	Number/ages of people living with	Live alone
		All over 15 years old
		Adults and children under 15 years old
House type	Type of property	Apartment
		Semi-detached
		Detached
		Terraced
		Bungalow
		Refused
Homeownership	Own or rent property	Rent (from a private landlord)
		Rent (from a local authority)
		Own outright (not mortgaged)
		Own with mortgage
		Other
No. Bedrooms	Number of bedrooms	Range 1 to 6
No. People	Number of residents	Range 1 to 6
No. Household appliances	Number of household appliances	Range 11 to 29

Table A.2
NRS Social grade classification scheme.

Social Grade	Social Class	Description
A	Upper middle class	Higher managerial role (administrative or professional)
B	Middle class	Intermediate managerial role (administrative or professional)
C1	Lower middle class	Supervisory or clerical and junior managerial role (administrative or professional)
C2	Skilled working class	Skilled manual worker
D	Working class	Semi-skilled and unskilled manual worker
DE	Nonworking	Pensioners, casual and lowest grade workers, unemployed with state benefits only

Table A.3
Distribution of house type in the CER dataset.

House Type	Percentage
Apartment	1.66
Semi-detached house	30.05
Detached house	27.06
Terraced house	14.25
Bungalow	26.78
Refused	0.20

Table A.4
Distribution of spacing heating type in the CER dataset.

Heating Type	Percentage
Electricity (electric central heating or storage heating)	3.35
Electricity (plug-in heaters)	2.76
Gas	24.59
Oil	47.18
Solid fuel	21.17
Renewable (e.g., solar)	0.43
Other	0.52

Table A.5
Distribution of water heating type in the CER dataset.

Heating for Water	Percentage
Central heating system	7.95
Electric (immersion)	33.84
Electric (instantaneous heater)	0.95
Gas	14.33
Oil	24.26
Solid fuel boiler	9.69
Renewable (e.g., solar)	0.97
Other	8.01

Table A.6
Distribution of cooker type in the CER dataset.

Cook Type	Percentage
Electric cooker	69.70
Gas cooker	25.66
Oil fired cooker	2.38
Solid fuel cooker (stove aga)	2.26

Table A.7
Proportion of each pattern on selected days of the year.

Pattern Date	P_1	P_2	P_3	P_4	P_5	P_6
2009-07-14	0.423	0.066	0.314	0.090	0.015	0.093
2009-12-24	0.208	0.139	0.196	0.125	0.191	0.140
2009-12-25	0.269	0.038	0.148	0.259	0.230	0.055
2009-12-26	0.256	0.072	0.204	0.177	0.145	0.145
2010-07-14	0.440	0.064	0.326	0.082	0.012	0.074
2010-12-24	0.199	0.138	0.187	0.123	0.223	0.130
2010-12-25	0.254	0.037	0.145	0.251	0.255	0.058
2010-12-26	0.239	0.062	0.210	0.185	0.178	0.126

Table A.8
Typical summer (July 14, 2009) to winter (December 24, 2009) pattern shifting matrix.

2009-12-24 2009-07-14	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0.379	0.088	0.268	0.094	0.068	0.104
P_2	0.050	0.254	0.063	0.100	0.413	0.121
P_3	0.115	0.182	0.191	0.151	0.179	0.182
P_4	0.058	0.129	0.107	0.187	0.377	0.141
P_5	0.036	0.055	0.036	0.091	0.727	0.055
P_6	0.030	0.169	0.092	0.145	0.374	0.190

Table A.9
Typical weekday (July 27, 2009) to weekend (August 2, 2009) pattern shifting matrix.

Sunday Monday	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0.746	0.010	0.153	0.067	0.002	0.022
P_2	0.166	0.156	0.213	0.280	0.062	0.123
P_3	0.273	0.044	0.397	0.189	0.012	0.085
P_4	0.161	0.084	0.230	0.373	0.053	0.099
P_5	0.064	0.043	0.170	0.362	0.255	0.106
P_6	0.153	0.077	0.212	0.362	0.049	0.147

Table A.10
Proportion of each intra-day pattern over a selected week.

Pattern Date	P_1	P_2	P_3	P_4	P_5	P_6
2009-07-27 (Monday)	0.428	0.058	0.323	0.088	0.013	0.090
2009-07-28 (Tuesday)	0.419	0.065	0.308	0.091	0.017	0.100
2009-07-29 (Wednesday)	0.419	0.064	0.325	0.092	0.014	0.086
2009-07-30 (Thursday)	0.443	0.059	0.327	0.091	0.012	0.068
2009-07-31 (Friday)	0.423	0.061	0.317	0.099	0.019	0.081
2009-08-01 (Saturday)	0.432	0.056	0.288	0.136	0.019	0.069
2009-08-02 (Sunday)	0.446	0.042	0.247	0.176	0.021	0.068

Table A.11
Proportion of the six intra-day patterns associated with the five annual load profiles on July 14, 2010 (summer weekday)

Intra-day pattern Annual profile	P_1	P_2	P_3	P_4	P_5	P_6
1	0.210	0.057	0.560	0.073	0.000	0.100
2	0.070	0.214	0.283	0.201	0.048	0.184
3	0.873	0.003	0.098	0.017	0.000	0.009
4	0.481	0.017	0.416	0.043	0.002	0.041
5	0.912	0.003	0.070	0.006	0.000	0.009

Table A.12
Proportion of the six intra-day patterns associated with the five annual load profiles on December 25, 2010 (Christmas day).

Intra-day pattern Annual profile	P_1	P_2	P_3	P_4	P_5	P_6
1	0.108	0.043	0.123	0.370	0.263	0.093
2	0.019	0.049	0.048	0.231	0.620	0.033
3	0.466	0.023	0.270	0.138	0.057	0.046
4	0.181	0.051	0.205	0.329	0.164	0.070
5	0.670	0.012	0.132	0.134	0.028	0.024

Table A.13
Machine learning model parameters.

Model	Parameter	Value
DNN	Number of net layers	4
	Number of hidden layers	2
	Nodes per layer	[22, 200, 200, 5]
	L1	0.005
	L2	0.003
	Maximum number of iterations	500
	Activation function	Rectifier
	Learning rate	0.1
RF	Number of trees	2000
	Maximum tree depth	20
	Early stopping based on stopping_metric convergence	2
	Relative tolerance of metric-based stopping criterion	0.001
GBM	Number of trees	2000
	Maximum tree depth	5
	Early stopping based on stopping_metric convergence	2
	Relative tolerance of the metric-based stopping criterion	0.01
	Learning rate	0.001
Elastic Net	Link function	Multinomial
	Solver	L_BFGS
	Regularization factor between L1 and L2	0.9
SVM	Cost	7
	Gamma	0.1
	Kernel	Radial basis

References

Beckel, C., Sadamori, L., Staake, T., Santini, S., 2014. Revealing household characteristics from smart meter data. Energy 78, 397–410.

Benítez, I., Díez, J.L., Quijano, A., Delgado, I., 2016. Dynamic clustering of residential electricity consumption time series data based on Hausdorff distance. Elec. Power Syst. Res. 140, 517–526.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

CER, 2012. Commission for Energy Regulation (CER) Smart Metering Project - Electricity Customer Behaviour Trial, 2009-2010, first ed. Irish Social Science Data Archive. SN: 0012-00. Available at: https://www.ucd.ie/issda/data/commissionforenergyregulationcer/. (Accessed 6 May 2021). Last accessed:

Chévez, P., Barbero, D., Martini, I., Discoli, C., 2017. Application of the k-means clustering method for the detection and analysis of areas of homogeneous residential electricity consumption at the Great La Plata region, Buenos Aires, Argentina. Sustain. Cities Soc. 32, 115–129.

Choksi, K.A., Jain, S., Pindoriya, N.M., 2020. Feature based clustering technique for investigation of domestic load profiles and probabilistic variation assessment: smart meter dataset. Sustainable Energy, Grids and Networks 22, 100346.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.

Crone, S.F., Kourentzes, N., 2009. Input-variable specification for neural networks-an analysis of forecasting low and high time series frequency. In: International Joint Conference on Neural Networks, pp. 619–626, 2009.

European Commission, 2013. European Economic Forecast: Spring 2013. Publications Office, Directorate-General for Economic and Financial Affairs. Available at: https://data.europa.eu/doi/10.2765/46506. (Accessed 28 January 2022). Last accessed.

Fahrmeir, L., Kneib, T., Lang, S., Marx, B., 2007. Regression. Springer-Verlag, Berlin/Heidelberg.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: a k-means clustering algorithm. J. Roy. Stat. Soc.: Series C (Applied Statistics) 28 (1), 100–108.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Monographs on Statistics and Applied Probability, vol. 43. Chapman and Hall/CRC, New York/Boca Raton.

He, P., Liang, J., Qiu, Y., Li, Q., Xing, B., 2020. Increase in domestic electricity consumption from particulate air pollution. Nat. Energy 5, 985–995.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: applications to nonorthogonal problems. Technometrics 12 (1), 69–82.

IEA, 2020. Key World Energy Statistics 2020. International Energy Agency (IEA). Available at: https://www.iea.org/reports/key-world-energy-statistics-2020/final-consumption. (Accessed 7 September 2022). Last accessed.

Khan, I., Huang, J.Z., Ivanov, K., 2016. Incremental density-based ensemble clustering over evolving data streams. Neurocomputing 191, 34–43.

Kohonen, T., 1990. The self-organizing map. Proc. IEEE 78 (9), 1464–1480.

Koprinska, I., Rana, M., Agelidis, V.G., 2015. Correlation and instance based feature selection for electricity load forecasting. Knowl. Base Syst. 82, 29–40.

LeCun, Y.A., Bottou, L., Orr, G.B., Müller, K.R., 2012. Efficient backprop. In: Neural Networks: Tricks of the Trade. Springer, Berlin/Heidelberg, pp. 9–48.

Li, Y., Pizer, W.A., Wu, L., 2019. Climate change and residential electricity consumption in the Yangtze River Delta, China. Proceedings of the National Academy of Sciences USA 116 (2), 472–477.

McLoughlin, F., Duffy, A., Conlon, M., 2012. Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: an Irish case study. Energy Build. 48, 240–248.

McLoughlin, F., Duffy, A., Conlon, M., 2015. A clustering approach to domestic electricity load profile characterisation using smart metering data. Appl. Energy 141, 190–199.

Motlagh, O., Berry, A., O'Neil, L., 2019. Clustering of residential electricity customers using load time series. Appl. Energy 237, 11–24.

Okoligwe, N.E., Ihugba, O.A., 2014. Relationship between electricity consumption and economic growth: evidence from Nigeria (1971-2012). Academic Journal of Interdisciplinary Studies 3 (5), 137.

Räsänen, T., Kolehmainen, M., 2009. Feature-based clustering for electricity use time series data. International Conference on Adaptive and Natural Computing Algorithms 5495, 401–412.

Räsänen, T., Voukantsis, D., Niska, H., Karatzas, K., Kolehmainen, M., 2010. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. Appl. Energy 87 (11), 3538–3545.

Reynolds, A.P., Richards, G., de la Iglesia, B., Rayward-Smith, V.J., 2006. Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. J. Math. Model. Algorithm. 5 (4), 475–504.

Singh, S., Yassine, A., Benlamri, R., 2019. Consumer segmentation: improving energy demand management through households socio-analytics. In: IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), pp. 1038–1045. IEEE.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc. B 58 (1), 267–288.

Ürge-Vorsatz, D., Cabeza, L.F., Serrano, S., Barreneche, C., Petrichenko, K., 2015. Heating and cooling energy trends and drivers in buildings. Renew. Sustain. Energy Rev. 41, 85–98.

Wang, F., Li, K., Duić, N., Mi, Z., Hodge, B.M., Shafie-khah, M., Catalão, J.P., 2018. Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns. Energy Convers. Manag. 171, 839–854.

Wood, S.N., 2017. Generalized Additive Models: an Introduction with R, second ed. Chapman and Hall/CRC, New York/Boca Raton.

Yu, Z., You, J., Wong, H.-S., Han, G., 2012. From cluster ensemble to structure ensemble. Inf. Sci. 198, 81–99.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. Roy. Stat. Soc. B 67 (2), 301–320.