UNIVERSITY OF CAMBRIDGE | Energy Policy Research Group

Machine Learning on residential electricity consumption: Which households are more responsive to weather?

EPRG Working Paper 2113

Cambridge Working Paper in Economics 2142

Jieyi Kang and David M Reiner

Abstract The introduction of smart meters has created opportunities for both utilities and policymakers to understand residential electricity consumption in greater depth. Machine learning techniques have distinct advantages over traditional approaches in dealing with extremely large volumes of high-resolution usage data. We introduce a novel clustering method to detect household behaviour using different types of weather data as proxies. Based on this approach, we combine Irish smart meter and weather data to identify and characterize clear differences in the daily patterns between workdays and weekends in both summer and winter and investigate how households respond to changing weather patterns. We also examine the relationships between response groups and household demographic features using different statistical tests. We find the magnitude of the effect of occupancy-related variables in the clustering of weather sensitivity to be larger than income-related factors. This proposed new approach could be the basis of a classification model to identify households that are more responsive to different types of weather. Tariff design could benefit from such a model and enable specific schemes to be developed that would target weather-sensitive households and result in improved load management.

Keywords Weather sensitivity; smart metering data; unsupervised learning; clusters; residential electricity; consumption patterns; Ireland

JEL Classification C55, D12; R22 ; Q41

Contact jieyi.kang@hotmail.com
Publication May 2021

Machine Learning and residential electricity consumption: Which households are more responsive to weather?*

Jieyi Kang[1,2] and David Reiner[1,3]

[1]Energy Policy Research Group, University of Cambridge
[2]Department of Land Economy, University of Cambridge
[3]Judge Business School, University of Cambridge

Abstract

The introduction of smart meters has created opportunities for both utilities and policy-makers to understand residential electricity consumption in greater depth. Machine learning techniques have distinct advantages over traditional approaches in dealing with extremely large volumes of high-resolution usage data. We introduce a novel clustering method to detect household behaviour using different types of weather data as proxies. Based on this approach, we combine Irish smart meter and weather data to identify and characterize clear differences in the daily patterns between workdays and weekends in both summer and winter and investigate how households respond to changing weather patterns. We also examine the relationships between response groups and household demographic features using different statistical tests. We find the magnitude of the effect of occupancy-related variables in the clustering of weather sensitivity to be larger than income-related factors. This proposed new approach could be the basis of a classification model to identify households that are more responsive to different types of weather. Tariff design could benefit from such

*To whom correspondence should be addressed: E-mail: jieyi.kang@hotmail.com

a model and enable specific schemes to be developed that would target weather-sensitive households and result in improved load management.

1 Introduction

Past quantitative studies of residential energy consumption have mainly focused on energy tariff pricing, explanations for differences in energy consumption, and models to predict consumption. However, many of these studies are based on aggregate levels of consumption, particularly those using econometrics (Bianco, Manca and Nardini, 2009; Karanfil, 2009; Sanquist et al., 2012). The objectives of these studies are various, including the research on relationships between daily consumption and household social-economic background (Hackett and Lutzenhiser, 1991; Druckman and Jackson, 2008; Jones, Fuertes and Lomas, 2015); studies focused on effects of weather variables on the total regional electricity consumption (Valor, Meneu and Caselles, 2001; Pardo, Meneu and Valor, 2002; Hor, Watson and Majithia, 2005). Due to the limitations in the resolution of their data however, these researchers were unable to conduct more detailed studies. As installation of smart metering in households has increased in recent years, analysis of high-resolution electricity consumption data becomes possible. The nature of high-frequency data brings opportunities to understand energy consumption with a granularity that would have been unimaginable even a few years ago. As a result, there are now new areas of research available arising from this high-resolution data in the energy sector.

One main area of focus is load management, especially the prediction of electricity consumption, which is of interest to both utilities and policymakers. Previous prediction models have generally been based on aggregated grid consumption data. Now with the technological advancements in metering, the high-frequency load data has the potential to help related parties to understand consumer behaviour better to achieve higher efficiencies. Prediction models can massively benefit from such data and significantly improve their accuracy (Beccali et al., 2008;

Ghofrani et al., 2011). Another research question which has been constantly discussed is dynamic pricing of electricity (Faruqui and Malko, 1983; Sanghvi, 1989; Herter, McAuliffe and Rosenfeld, 2007; Faruqui and Sergici, 2010; Alberini and Filippini, 2011). The deployment of smart meters enables utilities to set dynamic pricing structures than flat prices. There have been a great number of trials for various types of pricing schemes (Newsham and Bowker, 2010; Haider, See and Elmenreich, 2016), e.g. time-of-use tariffs, critical peak prices and etc. By introducing fluctuating prices, it could be helpful to reduce energy consumption and save the environment to some extent. Another motivation for dynamic pricing is that utilities intend to encourage customers to shift away from the peak times to reduce the power load during critical periods (Herter, 2007; Faruqui and Sergici, 2010). To analyse the efficiency of the tariff design, the effects of the pricing structures can be investigated thoroughly using high-resolution data.

Furthermore, understanding the correlations between customers' social-economic profiles and electricity consumption is also one of the classic applications of smart metering data (McLoughlin, Duffy and Conlon, 2012; Beckel et al., 2015). Accurate segmentation of electricity customers can assist in higher energy efficiencies and lower operation losses. Previously, using only aggregated daily or monthly household consumption data, it was difficult to look into the details of how and why households' consumption behaviours differ during specific time periods (Cramer et al., 1984; Silk and Joutz, 1997; Kaza, 2010). Previous studies therefore could only focus on longer periods of household consumption to explore differences in social-economic backgrounds or of property characteristics of houses. With smart metering data, utilities and researchers can finally create and understand the customer demand curves and the habits and behavioural patterns underlying the profiles. Due to the huge volumes of data involved and the complexity of the data processing, machine learning techniques, such as clustering and classification, have been increasingly adopted rather than more traditional econometric tools.

The majority of research featuring data mining techniques is focused on how to cluster

customers based on their load curves (Räsänen et al., 2010; McLoughlin, Duffy and Conlon, 2012; Razavi et al., 2019). The objectives of those studies are to identify the connections between the demand curves and the characteristics of households. However, there have been few studies using these tools on smart metering data for behavioural studies, and in particular there has been limited work on the effects of weather on household behavioural patterns. Load curves can partially reflect households' consumption patterns, for instance, identifying peak times or the amount consumed during a specific period. Nevertheless, an aggregated curve cannot reveal household preferences in any detail. Greater understanding of household behavioural patterns could benefit both policymakers and utilities.

To fill the gap, our main objectives are to understand how household daily life patterns are reflected in the demand response to weather sensitivities. This study brings together the smart metering and the survey data from the Irish Electricity Smart Metering Customer Behaviour Trials in 2012 with the weather data during the trial. Weather sensitivities of electricity consumption of each household during different periods of the day are the core of this study. They are used as proxies to discover household behaviour patterns, for example, when members of a household normally need to go out and at what time of day people are more likely to have spare time (or at least when they are most flexible in terms of their behaviour). The work proposes a novel method using machine learning techniques to identify the patterns using a two-step process: 1) Define the demand change indexes under different weather conditions; and 2) Employ clustering techniques on the indexes defined in Step 1 to generate representative sensitivity curves. With this method, the study offers a new perspective on the differences in household responses to weather changes drawing on time of use preferences derived from smart metering data. Combining these results with load curves can provide a better understanding of daily residential electricity consumption patterns.

This remainder of the paper is organised as follows: Section 2 presents a literature review

of past studies of smart metering data, especially works on electricity consumer segmentation and the correlations between weather and electricity consumption. Both the scope and methods are discussed. Section 3 includes the data and the methodology used here, consisting of data prepossessing, definition of demand change indexes, and the algorithms and a brief description of the clustering process. In Section 4, the results of the weather sensitivity curves are presented and explained in detail. A summary of the work and the conclusions are drawn in Section 5.

2 Literature review

Among all the grid operation improvements, the deployment of smart meters particularly benefit short-term load management. While short-term (hourly to daily) load forecast plays a critical role in load management, it has been rather difficult to model the demands by low-frequent data. Under this circumstance, forecasting for residential electricity demand particularly benefits from smart meter data. With household level load data, the models for both long-term and short-term demand forecasting have been well-established. Quilumba et al. (2015) propose improving the accuracy of short-term load forecasts by considering customer behaviour. Using clustering techniques on smart meter data, they create models for load forecasts from 30 minutes up to one day-ahead predictions. Taieb et al. (2016) prove that for disaggregated demand, an additive quantile regression model outperforms the traditional model with an normality assumption, based on the smart metering data from a trial in Ireland over a period of 1.5 years. Ghofrani et al. (2011) combines traditional Gauss-Markov process modelling with automatic meter readings to achieve a higher prediction accuracy, although it increases the computational cost. Some other studies using smart meter data focus on identifying the efficiencies of domestic appliances. Firth et al. (2008) attempt to reveal the trends in the use of appliances from a high-resolution dataset. They found that a 10.2% rise of "standby" appliances (such as consumer electronics) consumption accounts for the largest share of the overall demand increase.

Weiss et al. (2012) prove that disaggregation of individual appliances is possible by using a set of algorithms on smart metering data.

Apart from the modelling-related research, another stream of studies using smart meter focus on the effects of household characteristics on residential consumption. Gouveia and Seixas (2016) combine the meter readings with a door-to-door survey of 110 questions administered to 265 households in Portugal to unravel residential consumption profiles. They carried out clustering analysis using daily consumption and formed three profiles. The main variables used for profile analysis included: dwelling location and type, age, gender and educational attainment of household members. A U-shape pattern with higher consumptions at the beginning and end of a year and lower demand in the middle was found to be the most common type accounting for 77% of the households. Beckel et al. (2014) attempted to reveal household characteristics purely from consumption data with a supervised method. In their work the Irish CER trial data was used and the household's socio-economic status, appliance stock, properties of the dwelling, and the consumption behaviour of the occupants were considered as class labels in the research. The electricity consumption profiles were firstly formed through different indexes of consumption behaviour, for instance, the ratio of peak to off-peak. Then by a supervised-learning process, with input of electricity consumption data, the model would be able to identify household characteristics only depending on the consumption patterns. The experimental results show that among all other household characteristics, the occupancy state of the house, the number of persons in the house and the appliance stock can be identified directly from the consumption load profiles very well with an accuracy of more than 70%.

2.1 Data mining methods in smart metering data

The rich information brought by high resolution real-time smart meter data can improve the efficiency of grid operations. However, such massive data flows pose a major challenge for the

utilities to store and extract knowledge from the data (Viegas et al., 2015). Traditional tools such as database software are inadequate when dealing with huge amounts of data. In response, computational techniques, particularly, machine learning, have become increasingly appealing.

The applications can vary from operations, such as load forecasting, simulating Demand Side Management (DSM), and detecting bad data, to marketing – tariff design and potential customer identification. The core of the implementation is the segmentation of electricity consumers and load clustering. Wijaya et. al. (2015) use a cluster-based method to achieve short-term (1 hour and 24 hours ahead) electricity demand forecasting. Some argue that Support Vector Regression (SVR) is one of the most effective models to forecast electricity consumption (Chen, Chang and Lin, 2004; Sapankevych and Sankar, 2009; Cao and Wu, 2016; Chen et al., 2017). Wijaya et al. (2015) compared different algorithms including SVR, linear regression, and cluster-based aggregate forecasting (CBAF) and they suggest there is no single best algorithm in forecasting and use their own algorithm for clustering. From a review of load forecasting studies, it can be seen that Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are the three most widely used metrics to evaluate the accuracy of forecasting algorithms. However, it should be noted that those effectiveness metrics are not the only basis for algorithm measurement. Data structure and other relevant practical issues should also be considered.

McLoughlin et al. (2015), also drawing on Irish CER trial data, use clustering approaches to explore household load profile information. Unlike forecasting studies, they focused on the effectiveness of customer clustering by their household characteristics. The diurnal, intra-daily and seasonal consumption patterns were all examined. In order to evaluate the different clustering techniques, three of the most widely used algorithms are investigated: k-means; k-medoid and Self Organising Maps (SOM). They use a Davies–Bouldin (DB) validity index to compare the effectiveness of the algorithms as well as to determine an appropriate number of

clusters. Across the three techniques, the number of clusters was varied between 2 and 16. The results show that SOM and K-means have similar higher clustering power and that 8 to 10 clusters are the optimal numbers in this case. It is important to note that the optimal cluster number can vary depending on the objectives of the research, the features of datasets and even the selection of validity indices. Hierarchical clusterings are another popular set of algorithms for residential customer segmentation (Chicco, Napoli and Piglione, 2006; Chicco, 2012). Al-Wakeel et al. (2017) use k-means cluster analysis for load estimation study on the CER trial data. They suggest that compared to other algorithms, the significant advantages of k-means are simplicity and efficiency, particularly when considering the computational cost. Four different varieties of K-means distance functions are compared: Average Euclidean distance, Average Manhattan distance, Average Canberra distance, and Average Pearson correlation distance. The results of MAPE and RMSE indexes reveal that Canberra produces more accurate forecasts and the smallest error distributions. Gouveia and Seixas (2016) employ hierarchical clustering using Ward's Method based on the Squared Euclidean distance. They tested a range of numbers for cluster from 3 to 12. Since they conclude that increasing the number of clusters captures more information, they opted for the 10 clusters variant. One difference in conducting their cluster analysis is that they used mean daily consumption data to create a year profile, rather than using hourly data for a daily profile in other studies. In addition, the raw data was not normalised and the shapes of the profiles mainly reflect the magnitude of the consumption.

In summary, the key process of applying clustering analysis is to determine suitable algorithms and the number of clusters. No single algorithm outperforms the others in all situations with regard to residential electricity customer clustering. Any decision or selection must be based on the aims of the research and nature of the data structures. In particular, K-means and its relevant algorithms are mainstream choices that have been abundantly discussed for the case of residential load profile clustering. However, clustering on the basis of weather sensitivities

has rarely been explored in the past. Traditional consumption load profiles mainly focus on load forecasting (and forecast accuracy). On the other hand, clustering on weather sensitivities, might offer new perspectives and approaches to customer behaviour pattern studies that the weather response may be a good indicator for behavioural patterns.

2.2 Weather factors in residential electricity demand

Due to the lack of high-resolution consumption data at household level, past weather studies have mainly been conducted at the regional level using aggregated data. There are two types of research typically incorporating the weather factors: (i) demand forecast models and (ii) econometric models to identify the effect of different factors on electricity demand. Taieb et al. (2016) conducted a quantile regression to improve forecasting accuracy based on the Irish CER trial. In their model, the only weather variable included is outdoor temperature to control its effect in forecasting demand. Beccali et. al (2008) assessed the weather sensitivity on short-term household electricity consumption using cooling and heating degree-days (CDDs and HDDs) as temperature proxies. These two proxies are commonly employed when a non-linear relationship between temperature and demand is assumed (Fan and Hyndman, 2011; Blázquez Gomez, Filippini and Heimsch, 2013). Other weather factors that have been considered are relative humidity, humidex index, global solar radiation, wind speed, and atmospheric pressure. Some researchers (Albert and Rajagopal, 2012; Fikru and Gautier, 2015) claim that the main contributors are humidity index, CDDs, and HDDs while other variables are negligible in affecting residential electricity consumption. Henley and Peirson (1998) modelled the relationship between residential demand and price and temperature in the UK using a fixed-effects model and found a negative correlation. The opposite result was also shown in Wangpattarapong et al. (2008) in examining the impacts of climatic and economic factors on energy consumption in Bangkok. They use cooling-degree days as their temperature variable and find that a significant

positive relationship exists. Although the results in these two studies seem contradictory, the different effects of temperature may be the product of geographical location or climate zone. Whereas peak consumption in Bangkok is from air conditioning on the hottest summer days, peak UK electricity demand comes in winter. Hor, Watson and Majithia (2005) find a very weak negative effect of rainfall for monthly demand from 1983 to 1995 in the UK. Apart from temperature, which is usually seen as the main driver of a weather effect, humidity, wind speed, degree of cloudiness, and barometric pressure are also often discussed in related studies (Pardo, Meneu and Valor, 2002; Albert and Rajagopal, 2013).

However, to the extent they are considered, weather conditions have been treated as exogenous variables to control the effects of interest. Perhaps surprisingly, investigating the impacts of weather changes on households' daily life patterns through the electricity demand response to weather variables have been rarely seen. In the next section on Methodology, a novel approach for using weather sensitivities of consumption as proxies for household behaviour pattern will be explored in detail.

3 Data and methodology

In this study, three datasets are used for the clustering analysis of customers' life patterns: meter readings and survey results from the Customer Behaviour Trials (CBT) conducted by the Commission for Energy Regulation (CER) in Ireland, and hourly weather data collected by the Irish Meteorology Office. We begin with an overview of the data sources followed by a discussion of the data pre-processing needed. Finally, the algorithms and the performance measures used are presented.

3.1 Data preparation

The metering data from the CBT contains 15-minute consumption data from over 4,000 respondents during the period from July 2009 to December 2010. Since around 1,000 commercial customers participated, we only selected the sample of 3000 which are defined as residential customers.

Before the cluster analysis for load profiles, higher-resolution data (e.g. quarter-hourly) is often aggregated into hourly consumption profiles of 24 hours as part of data pre-processing. However, in this study, due to the different objectives, the data cleaning process is different than in most previous studies. The weather sensitivities of electricity consumption in households are not real-time responses but involve lags. Therefore, the quarter-hourly data is aggregated into larger chunks of time to accommodate the lag effects. Another consideration is the effect of various time-of-use (TOU) tariffs during the trial. The periods of the tariffs are: off-peak (8:00-17:00 and 19:00-23:00 weekdays and 17:00-19:00 weekends and bank holidays), peak (17:00-19:00 Monday to Friday, excluding bank holidays), and super off-peak (23:00-8:00). To control for the effects of the tariffs, all data aggregation should be within the period division with the same TOUs. Apart from the two-hour peak period, the off-peak and super-off-peak periods are much longer, which might hide some weather effects in specific sub-periods. For example, the demand response at lunch time might be clearer or stronger if the lunchtime from 12.00-14.00 were to be analysed separately. Otherwise, the demand change would seem minimal during the longer 9 hour off-peak period from 8.00 to 17.00. As a result, subdivisions of these two periods are created by considering the time use of the households during different periods of a day: Morning (6:00-8:00), Day_1 (8:00-10:00), Day_2 (10:00-12:00), Day_3 (12:00-15:00), Day_4(15:00-17:00), Peak (17:00-19:00), Evening_1 (19:00-21:00), Evening_2 (21:00-23:00) and Night (23:00-3:00). We exclude the period 3:00-6:00 from the analysis for two reasons: 1) From the Irish time use survey, it can be seen that over 99% of the households are sleeping after

2:00. 2) The demand response is minimal in that period. The metering data are then aggregated accordingly and transformed to indexes for the clustering.

Since the location of each household is not provided for confidentiality reasons, it is impossible to match exact local weather data with the households. The half-hourly weather data at Dublin airport from the Irish Meteorology Office is used because the participants were concentrated around Dublin according to the CER report (CER, 2012). Moreover, Ireland is a relatively small country and the weather variations across the country are limited (Ben Taieb et al., 2016). We assume that the weather at Dublin airport is sufficient to be representative for the weather elsewhere in the country at any given time. A weighted-average approach was also explored using several Irish weather stations but the differences from the Dublin-only approach were relatively minor (the weather data from the two datasets was compared by the t-test and the results is shown in Appendix. Three weather variables in the downloaded dataset, temperature, precipitation, and sun duration, are selected for the weather sensitivity estimations. As discussed in the literature review, the impacts of other weather variables, such as wind speed and humidity, might be negligible for which are not included in the analysis. The weather data is then aggregated and prepared for the clustering model.

3.2 Clustering input

To identify how people respond differently to weather variations we use a novel index to measure electricity demand changes under different weather conditions, including temperature, rainfall, and sun duration. Considering the behavioural response to weather may vary seasonally and on different days of week, we explore four combinations for each weather variable: summer workdays (SW), summer rest-days/weekends (SR), winter workdays (WW), and winter rest-days/weekends (WR). To ensure seasonality is as representative as possible, we define summer as from May to August and the winter from December to February.

In order to ensure a wide-enough fluctuation in weather variables while maintaining a relatively robust sample size, we selected the top 20% and bottom 20% of days in each period for all weather variables. It should be noted that in choosing days of two ends of temperatures, we excluded the days with medium or heavy rain before the day selection to control the precipitation effect. In addition, Christmas and New Year holidays are not included in the selection pools for weather variations in the winter rest-day scenarios, since it is expected that the behavioural sensitivity to weather would be completely different than usual weekends. The statistical summaries of the three weather variables are shown as below in Table 1 .

		Morning	Day_1	Day_2	Day_3	Day_4	Peak	Evening_1	Evening_2	Night
Temperature (°C)	Mean	7.23	8.64	10.18	11.03	10.75	9.83	8.75	7.91	7.20
	Std. Dev.	5.42	5.56	5.37	5.33	5.71	5.88	5.63	5.32	5.25
	Min	-7.49	-7.49	-5.16	-2.84	-3.60	-4.98	-5.20	-5.85	-6.33
	Max	17.52	18.09	19.94	20.57	20.78	19.82	18.69	17.52	17.09
Rel. Humidity (%)	Mean	89.26	84.43	77.98	74.21	74.97	78.61	83.04	86.48	88.75
	Std. Dev.	5.08	7.54	9.74	10.48	11.10	10.39	8.42	6.21	4.75
	Min	64.67	62.83	51.77	46.53	47.57	51.18	57.25	64.74	71.44
	Max	97.52	97.14	96.15	96.32	96.53	96.91	97.27	97.22	97.28
Wind speed (knots)	Mean	8.24	9.04	10.06	10.67	10.46	9.60	8.61	8.05	8.02
	Std. Dev.	3.56	3.55	3.72	3.77	3.84	3.87	3.82	3.89	3.78
	Min	2.03	2.54	2.91	3.46	2.66	2.86	1.75	1.47	2.06
	Max	27.16	27.87	27.07	29.03	28.93	29.79	28.83	27.93	24.34
Sun duration (% per hour)	Mean	0.14	0.36	0.48	0.46	0.33	0.20	0.07	0.00	0.00
	Median	0.04	0.31	0.50	0.45	0.25	0.03	0.00	0.00	0.00
	Std. Dev.	0.23	0.30	0.31	0.30	0.31	0.29	0.15	0.00	0.00
	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max	0.97	0.98	0.98	0.98	0.98	0.98	0.91	0.03	0.00
Rainfall (mm)	Mean	0.08	0.07	0.08	0.08	0.10	0.13	0.12	0.10	0.10
	Median	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.01
	Std. Dev.	0.19	0.18	0.18	0.18	0.23	0.31	0.29	0.23	0.22
	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max	1.40	1.34	1.32	1.71	1.96	2.27	1.93	1.84	1.82

Table 1 Descriptive Statistics for the weather variables

The new index for clustering is defined as follows and is calculated for each scenario separately:

$$I_{w,p,i} = \frac{\overline{E}_{w,p,i,low} - \overline{E}_{w,p,i,high}}{\overline{E}_{w,p,i,high}} \times 100\% \tag{1}$$

$I_{w,p,i}$ denotes the revised demand change index for the i_{th} household for the weather variable

w in period p. It shows by what percentage that energy demand changes towards the weather changes in the p_{th} period. $\overline{E}_{w,p,i,low}$ indicates the average electricity demand for household i in the p_{th} period on days with the bottom 20% of values for weather variable w. For example, for temperature in the morning period, we first selected the bottom 20% of the days in terms of temperature value and then calculated the average morning demand for those selected days based on household consumption. Similarly, $\overline{E}_{w,p,i,high}$ represents the demand of the p_{th} period for the top 20% of days for weather variable w. Therefore, the vector $C_{w,i}(I_{w,p_1,i}, I_{w,p_1,i}, \ldots, I_{w,n,i})$ consists of the weather sensitivities of household i of periods of day in certain scenario and the vectors are then directly used as inputs for the household clustering.

3.3 Algorithms and performance measures

The aim of cluster analysis is to identify weather sensitivity patterns. The sensitivity of the three weather variables can be regarded as different proxies for households' daily patterns:

1. The demand response to temperature may indicate the seasonality of activities during a certain period

2. The sensitivity to rainfall may imply that regular outdoor activities occur in that period or that the household is used to going out during that period

3. The sun sensitivity can be seen as an indicator of whether the household has spare time and to what extent their behaviour or activities are sensitive to sunshine over a given period

The weather sensitivity profile of a day for each household is obtained and consists of nine coefficients for each period of a day. The cluster analysis generates representative pattern curves for each weather variable.

The literature review shows that K-means is among the most widely used techniques for analysing load profiles. K-means have significant advantages in terms of being simpler and demanding less computational capacity. In addition, cluster analysis using K-means on index-based clustering results have been widely discussed in the past studies and have proved efficient. Hierarchical algorithms can be helpful to determine cluster numbers and also as cross-validation.

There are two important issues which must be addressed before clustering: how to decide on the number of clusters for the algorithms, and the effectiveness of the data partitioning. Performance can be measured using different clustering validity indicators but the indices used in previous studies vary. Moreover, Chicco (2012) finds that no single measure consistently prevails over the others. Therefore, most previous research into electricity consumption clustering adopts at least two different indices in order to address concerns over robustness and obtain a reliable and valid result (Yang and Sun, 2013; Räsänen et al., 2010; Ramos et al., 2015). We use two indicators, the Silhouette score and Davies–Bouldin index (DBI), which are widely used in electricity demand clustering studies, to assist in the selection. Silhouette score is defined by the mean intra-cluster distance and the mean nearest-cluster distance for each observation. A higher Silhouette score means clusters are farther apart and less dispersed, while values near 0 indicate overlapping clusters. The DBI score measures the average similarity of each cluster with its most similar cluster. The similarity is calculated using the ratio of within-cluster distances to between-cluster distances. A lower value indicates a better clustering.

Considering the load profile clusterings in the literature (Gouveia and Seixas, 2015), a series of numbers of clusters from 5 to 15 are examined. It should be noted that no absolute optimal number exists in the cluster analysis. The choice of the final number of clusters is based on the indicators and the practical experience.

3.4 Statistical inferences

After the representative weather sensitivity curves are created by the clustering algorithms, we want to investigate the relationships between household background variables and clustering of weather sensitivity patterns from two perspectives: 1) whether the household features could affect the clustering in each scenario, in other words, whether the clusters are independent of each household feature; and 2) whether certain dominated profiles are correlated with a particular cluster of weather sensitivity or daily behaviour patterns. To answer the first question, Chi-square tests of independence are employed to identify whether in a certain group the distribution/structure of one social variable is different from each other. On the second question, Chi-square goodness of fit test is adopted to examine whether the clusters statistically differ from that of the population as a whole. Effect sizes are calculated for both to compare which variable potentially has more effect on the sensitivity pattern segmentations. The demographic variables we are interested in are gender, age, employment status, social class, whether they live with other people, how many adults/children are in the household, education, and income (see Table 2).

4 Results and Discussion

In this section, we start by comparing algorithms and selecting a suitable number of clusters for each weather sensitivity analysis. The results for sunshine duration, rainfall and temperature are discussed separately and followed by the results of the statistical tests for the relationships between demographics and weather. With the assistance of DBI and Silhouette analysis, we chose seven as the cluster number for the workday scenarios for sun as well as all the scenarios for temperature, while six was the optimal number for the sun and rain weekend scenarios. In order to stay focused on the results here, the DBI and Silhouette results are included in the

Appendix.

4.1 Clustering results

In each sub-section, the weather sensitivity patterns for both workdays and weekends are briefly described. In the legends, the first number describes the cluster number, the second is how many households are categorised into that group and the third reflects the ratio of the households in that group to the whole sample.

Sun

The usage change in the Figure 1 represents the percentage change in electricity demand changes from the bottom 20% to the top 20% of sunny days. If the number is positive, it indicates that households use less electricity on a sunny days. As discussed under Methodology, the sun duration sensitivities represent the availability of discretionary time in certain periods. The sensitivity curves therefore can be seen as indicators of the extent to which households are able to allocate their time freely through a day. The sensitivity patterns are presented in Figure 1. Given the shortened days, in Winter (December to February), there is no direct sunlight for any periods after the Peak. Therefore, the sensitivity curves for sun duration only include periods from Morning through Day_4 in the winter scenarios.

In general, for all four scenarios, afternoons (Day_3 and Day_4) are more responsive to sunlight than mornings, indicating that people tend to have more discretionary time during the afternoon. In terms of the seasonal difference between workdays and weekends, mornings in summer are more sensitive to sun duration changes, while households are more responsive during afternoons in winter than in summer. The only exception is the morning period where responses are drastically stronger in winter. It could be explained by people tending to get up earlier on sunny days, since most of winter mornings would be dark and sunny mornings

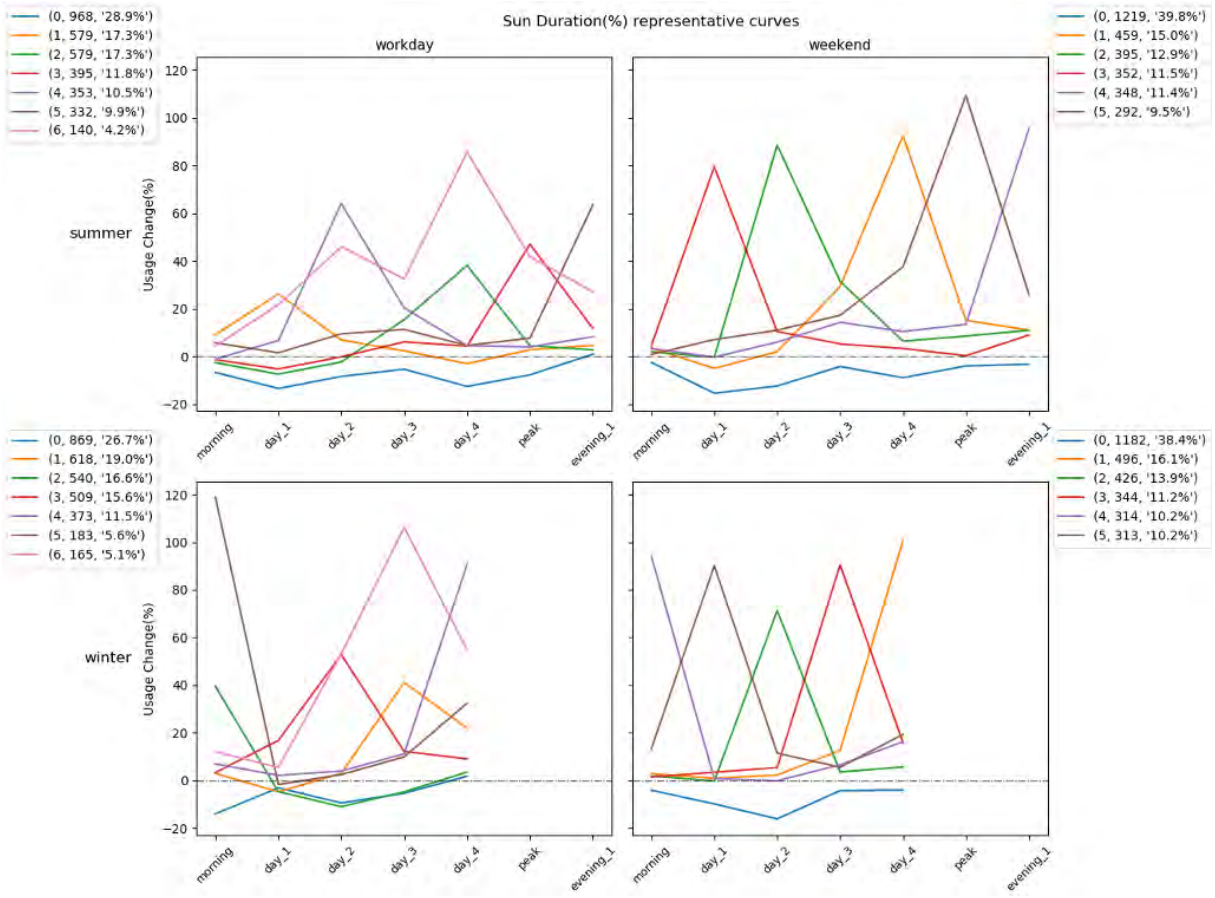

Fig. 1 Sun duration sensitivity

might wake people up earlier. In summer, especially on weekends, households respond even more dramatically in the evenings. In terms of winter, the trends are generally similar to the shapes in summer, although the responsive curves in the mornings of summer are more diverse. The possible explanation could be that people have less variety of outdoor/indoor activities in winter in general. For example, even if the weather conditions are good, households would be relatively unlikely to plan a picnic for weekend mornings in winter.

From the distribution/number of households in the clusters, we found that the segmentation is more dense and less balanced when dividing up into clusters on weekends. Group 0, the largest and least sensitive group, accounts for around 40% of households in the clusterings for

both summer and winter. However, Group 0 on weekends is still less flat and more sensitive than its counterpart group on workdays, especially for the mid-day periods (Day_2 to Day_4). It could be caused by less flexibility on workdays for employed households. Although in both summer and winter the largest group gives a negative score, which indicates the households use more electricity during a sunny day, the reason or behaviour behind it could be different. For winter, the increase in energy demand could be driven by the sun-related indoor activities, such as laundry or car washing or gardening. Because sunny days are rarer in winter people would take advantage of the weather to plan for weather-related activities. For summer with plenty of sunlight and better weather, the increase is likely to be from hosting parties especially on weekends or enjoying sunlight at home, rather than rushing to arrange the chores because of a good weather. And it should also be noted that even as the largest group, it still only accounts for 30% of the whole sample and the majority is positively sensitive to sunlight and would prefer outdoor activities in a sunny day.

Rainfall

From the negative response to rainfall one can infer whether a period is normally occupied by outdoor activities. The positive response reflects households using more electricity during the bottom 20% of rainy days (normally non-raining days). Thus, one can identify whether the period is typically used for rain-sensitive activities. Figure 2 shows that on summer workdays there is no single preferred outdoor period for all groups. The preferences are more evenly spread throughout the daytime, although there are slightly more groups affected during the before and after lunchtime periods (Day_2 and Day_4). People in winter workdays are clearly more responsive to rain. And the majority of groups, except Groups 4 and 5, prefer to go out during the midday periods of 10:00-15:00. One possible reason to prefer the mid-day periods may be that for stay-at-home family members those periods are more flexible and less likely to

be occupied by fixed house-bound activities, such as picking up children and preparing meals for families. The sensitivity decreases as it is getting late and it could be argued that as evening approaches, more indoor activities/house chores take place and households are less likely to choose these later periods to go out.

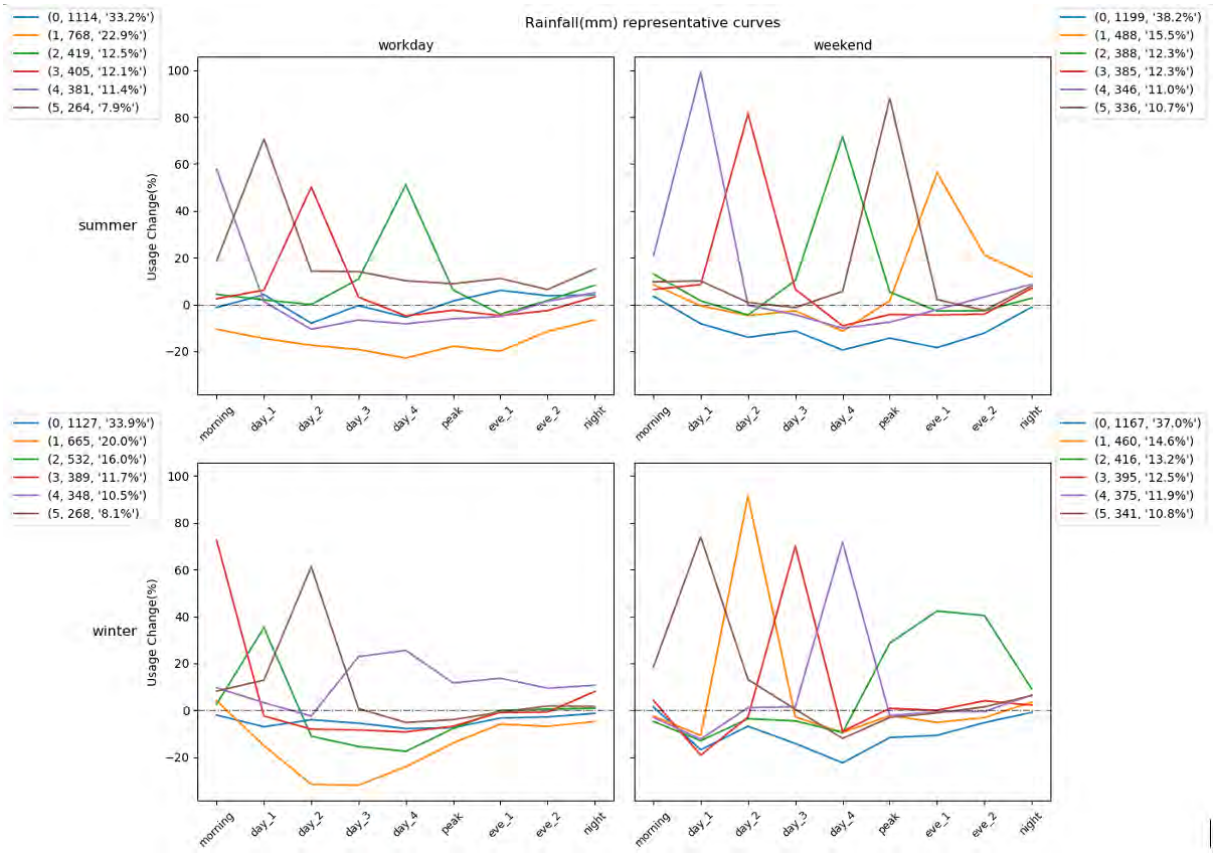

Fig. 2 Rain sensitivity

In terms of seasonal differences, it can be seen that all groups in the mornings of winter workdays have non-negative responses, while the changes in summer varies. One possible reason could be that people are more likely to be affected by rain on winter workdays and tend to leave home earlier to avoid possible traffic jams. However, on winter weekends almost all groups, apart from Group 5, display slightly non-positive changes. This could be caused by those who typically have outdoor morning plans, such as jogging, where rainfall interrupt their

schedules, although the responses are still minimal compared to Day_1. The magnitude of the sensitivity demonstrates that Day_1 is the period most affected, which indicates a preference of going out during mornings on winter weekends. It is clear from the Figure that workdays evenings in summer are slightly more likely to be affected. This may be because that people usually tend not to go out on workday evenings, especially in winter, and rain would be less disruptive in the winter evenings. The greater fluctuation in responses on summer weekends could reflect the fact that more outdoor activities are planned at those times and people would be more willing to go out than winter. Relatively bad weather in winter is expected and households would not easily alter their behaviour due to rainy weather.

To explore the differences between workdays and weekends, we examined the household distributions into clusters in all the scenarios. One finding is that, regardless of season, people are much more densely segregated into one group on weekends. With over 37% of households clustered into one group, which is also the most sensitive group, it could reflect the fact that most families would generally spend some time outdoors during weekends, since weekends are more flexible than workdays. However, it can be seen that there is no obvious peak/preferred period for the majority group since the curve is smoother than its counterpart in winter. On the other hand, Day_2 and Day_4 are preferred by many households in the largest group on winter weekends. The more significant positive responses at different periods of time could mean that many households may prefer specific time periods for outdoor activities on weekends, especially rain-sensitive work, for instance, washing cars or gardening.

Temperature

In general, response on workdays are less sharp than on weekends due to more limited flexibility (as can be seen in Figure 3). The sensitivities in winter are more vibrant than those in summer. The differences among clusters in winter is much bigger – for example, the difference

between the top 20% and the bottom 20% of day in summer is much smaller than in winter. The temperature difference is around 3-4 ℃ in every period in summer versus 7-8 ℃ in winter. As an island in the North Atlantic with mild summers and moderate winters, the maximum summer temperature in Ireland is only above 23 ℃, while the minimum in winter is -8℃. It might be imagined that such a narrow range of summer temperature fluctuations would result in fairly limited behaviour changes moderating any swings in electricity demand. As expected, temperature response curves in winter are much more stronger than in summer. The majority group (G0) in summer show a relatively flat response. Meanwhile, the household distribution in the clusters confims the hypothesis that individuals are less likely to be affected by temperature changes in summer, notably the largest and the least sensitive groups on both workday and weekends accounts for over 32% of all households. The situation in winter is more evenly spread and even the largest group is more responsive than in summer.

Another seasonal difference is that whereas the peaks/the most sensitive periods in winter fall during midday periods in winter, the counterparts in summer do not see a clear trend and can occur at any periods throughout the day. Unlike summer, almost all groups are sensitive to temperature in winter, especially the periods from Day_3 onward. For winter workdays, Day_2 and Day_4 appear to be the most responsive periods where people tend to go out if it is not extremely cold; On the other hand, we see all groups respond to Evening_1 on summer workdays. In addition, the sensitivities of temperature show a non-increasing trend after Evening_1 (19:00-21:00) in summer, but the responses in winter reflect an opposite tendency of non-decreasing. This contrary result is even more non-considerable on weekends. The possible explanation could be that due to the limited change of temperature in summer, the chance that temperature affects peoples' outdoor plans at night could be minimal, compared to the effect in winter. People would be more willing to go out during a warm day in winter.

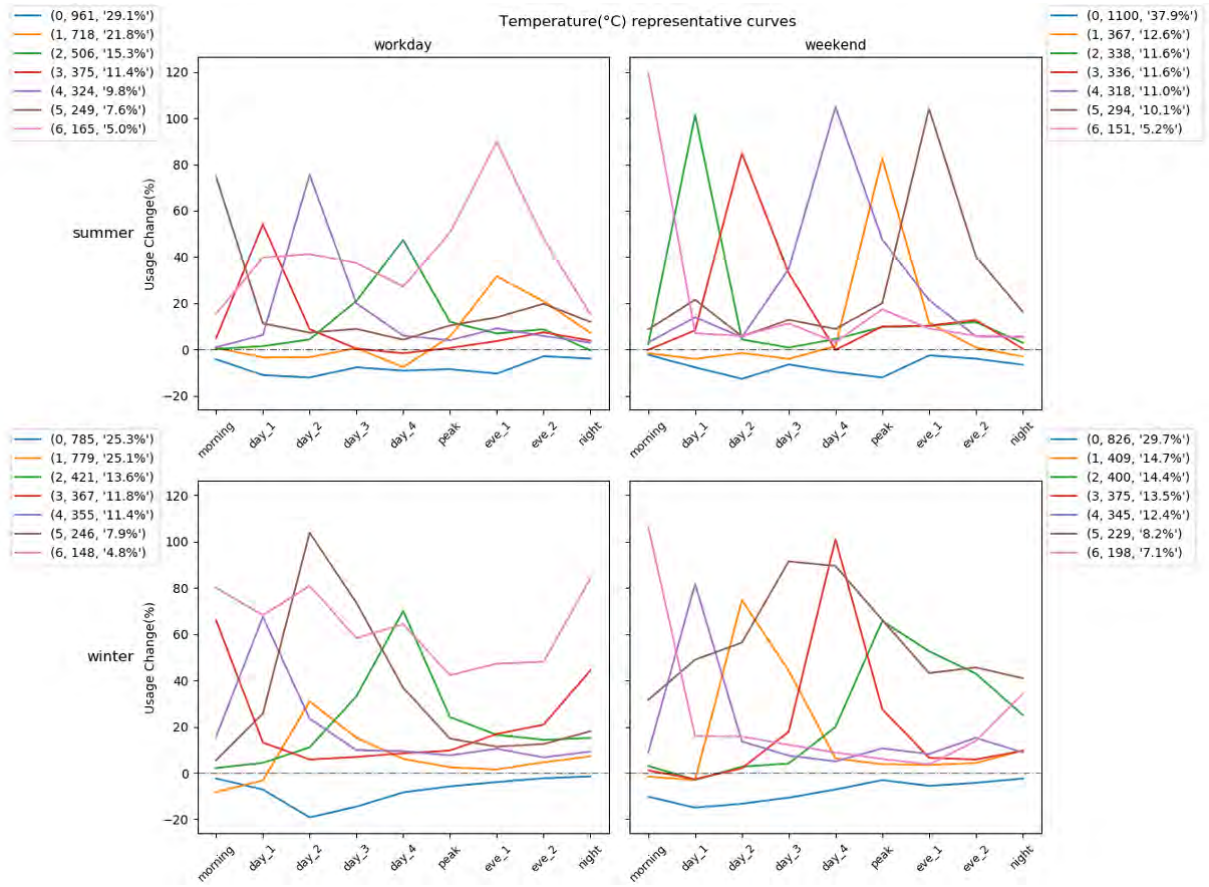

Fig. 3 Temperature sensitivity

4.2 Statistical results

In this part, we focused on the questions at two levels: 1) In general, whether the clustering is associated with some of the selected socio-economic variables and whether the variables are more connected to a season or workdays/weekends. 2) At the cluster level, is there clear household profiles behind some groups?

4.3 Features reflecting weather sensitivity clustering

To answer the first question, we used chi-square tests of independence as well as effect sizes to identify the variables that affect the clustering.

The Chi-squared test of independence is used to determine whether a relationship exists between two nominal/categorical variables. The frequency of each category for one nominal variable is compared across the categories of the second nominal variable. Here, we compare the frequency distribution of each social-economic variable for each cluster category separately. The observed frequencies are the total counts for each level of one variable at each level of the cluster category. The expected frequency counts are computed separately for each level of one categorical variable at each cluster[1]. The chi-square test is then performed based on the expected and observed frequencies[2]. For example, to investigate the relationship between education levels and clusters, the observed frequencies are the counts of each education level for each cluster. The expected frequencies are the calculated frequencies of each education level at each cluster based on the distribution of the total sample.

The list of variables we tested can be seen in Table 2. Table 3 shows the results that with p-value lower than 0.05 and the questions are ranked in descending order of effect size, which is indicated in parentheses.

From a quick glance at the number of variables in each column in Table 3, it can be seen that in general more socio-economic features are associated with rain sensitivities. In other words, compared to other weather variables, the behaviour patterns affected by precipitation are more related to multiple household characteristics. In terms of seasonal differences, the clustering in the workday scenarios are associated with more household demographic variables during winter although a number of the significant variables overlap. By contrast, for rest-days there is no consistent pattern or many significant variables that repeat for different seasons.

[1] $E_{r,c} = \frac{n_r * n_c}{n}$, where $E_{r,c}$ is the expected frequency count for level r of a social-economic variable A and level c of Cluster C, n_r is the total number of sample observations at level r of Variable A , n_c is the total number of sample observations at level c of Cluster C, and n is the total sample size.

[2] The test statistic is a chi-square random variable (χ^2) defined as: $\chi^2 = \sum \frac{(O_{r,c} - E_{r,c})^2}{E_{r,c}}$, where $O_{r,c}$ is the observed frequency count at level r of Variable A and level c of Cluster C, and $E_{r,c}$ is the expected frequency count at level r of Variable A and level c of Cluster C.

Code	Variables
Q300	**Age**: 18-25, 25-35, 36-45, 46-55, 56-65, 65+, Retired
Q310	**Employment Status:** Employee, Self-employed, Unemployed, Retired, Carer: Looking after relative family
Q401	**Social Class:** AB, C1, C2, DE and below, F(Record all farmers)
Q402	**Income level:** Less than €15k,15k to 30k, 30k to 50k, 50k to 75k, 75k+
Q410	**Living status:** Live alone, All people over 15yrs, Both adults and Children
Q420	**How many people over 15?**: 1, 2, 3, 4+
Q430	**How many people over 15 in house during day time?** (If Q410 is not "Live alone") 1, 2, 3+
Q4312	**How many under 15 in house during day time?** (If Q410 is "Both adults and Child") 1, 2, 3+
Q5418	**Education level:** Primary and below, Secondary to Intermediate Cert, Secondary to Leaving Cert, Third level

Table 2 Code list

Regarding the specific variables, living status (Q410) is the most relevant variable to affect clustering and is statistically significant in 10 out of 12 scenarios (across all weather variables in summer and winter workdays). For sun sensitivity clustering, how many people in the household are over 15 years old (Q420), whether they are at home during daytime (Q430), and employment status (Q310) are the next three most commonly significant variables. For rain sensitivities, age (Q300) affects all scenarios, although Q310 and Q420 also often play roles. Likewise, age (Q300) and employment status(Q310) are significant for certain temperature profiles. However, the effect of living status dominates other variables with the highest effect sizes regardless of seasons or workdays/weekends.

The effects of income-related variables are not as significant as the occupancy-related variables in the clustering of weather sensitivity. Yet with lower differentiating power, it stills has a role in pattern segmentations. In general, social-class variables are least likely to affect temperature sensitivity: Education level (Q5418) is the only income-related variable that is linked

	Sun	Rain	Temp
SW (Summer Workdays)	**Q410**(0.085), Q430(0.066), Q420(0.056), Q310(0.055)	**Q410**(0.076), Q420(0.068), Q430(0.066), **Q300**(0.059), Q310(0.058)	**Q410**(0.076), Q310(0.067), **Q300**(0.064), Q420(0.057), Q5418(0.057)
SR (Summer Rest-days)	**Q410**(0.06), Q420(0.058)	Q310(0.059), Q401(0.057), **Q410**(0.056), **Q300**(0.051),	**Q410**(0.078), **Q300**(0.07), Q310(0.069), Q5418(0.061)
WW (Winter Workdays)	Q310(0.094), Q430(0.081), Q401(0.084), **Q300**(0.073), **Q410**(0.07), Q420(0.058)	**Q410**(0.072), Q402(0.07), Q430(0.063), **Q300**(0.061), Q310(0.058), Q401(0.055), Q420(0.055),	**Q410**(0.11), Q420(0.078), **Q300**(0.074), Q310(0.074), Q430(0.071)
WR (Winter Rest-days)	Q402(0.076), Q5418(0.063), Q430(0.061), Q401(0.056), Q310(0.053),	**Q300**(0.066), **Q410**(0.06), Q401(0.055), Q420(0.053)	Q420(0.064), **Q300**(0.061)

Table 3 Significant variables for each scenario

with the temperature sensitivity clustering and even then only exists in summer scenarios. Rain patterns, on the other hand, are frequently associated with a number of variables of this kind, for example social class (Q401) and income level (Q402), especially Q401, which affects three out of four rainfall scenarios. In terms of sun clustering, it appears that only winter scenarios are related to income-related variables. However, it should be highlighted that the scenario of sun sensitivity in winter weekends is mostly associated with socio-economic variables, notably income, social class, and education level.

4.4 Clusters with clear background profiles

To investigate if any groups are dominated by specific demographic profiles of households, we look into the distributions of each variable to identify the differences in household characteristics between each group and the population using Chi-square tests. The variables we chose to test are the same as used in the last section and listed in Table 2. We selected the top two distinguished groups (the representative curves shown in in Figure 4) that have at least three profile variables statistically significant (p values less than 0.05) for the weather sensitivity clusterings.

The meaning of the labels in the legends is as follows: SW for summer workdays, WW for winter workdays, and WR for winter weekends. The latter part starting with "g" indicates the group number in that scenario.

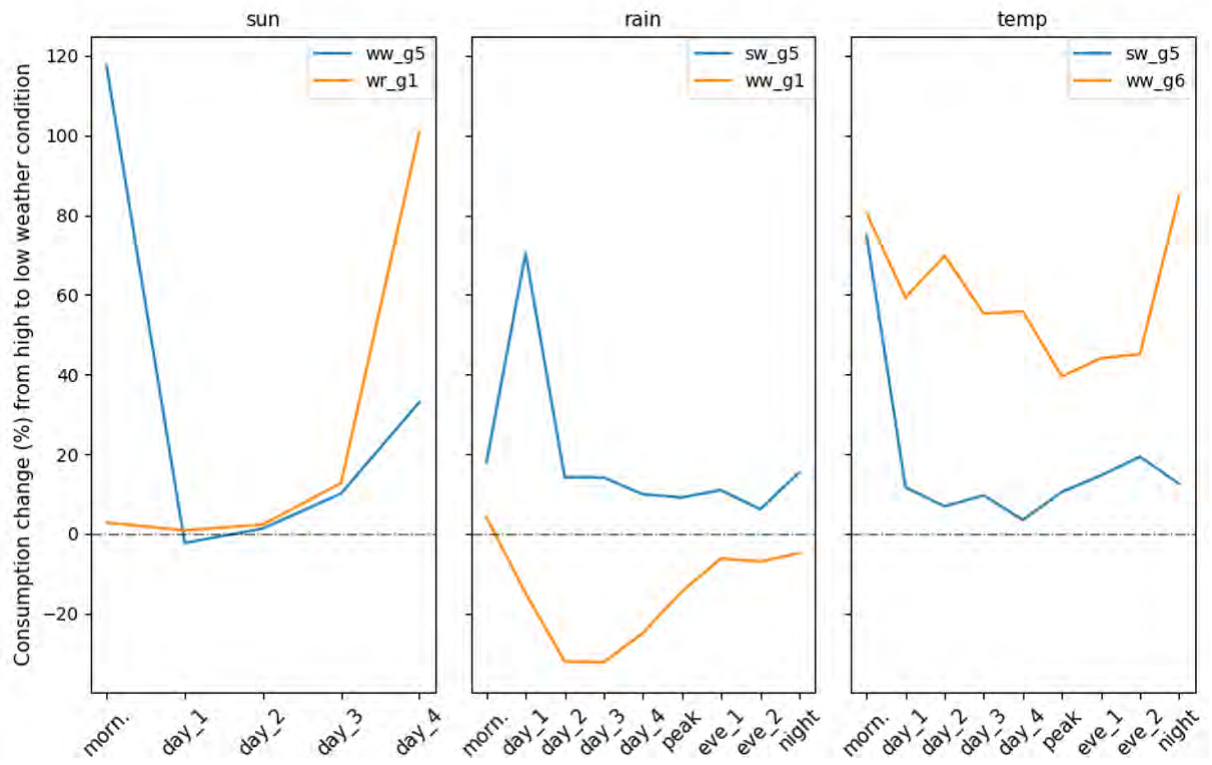

Fig. 4 Representativ curves for selected groups

Table 4 shows which questions are statistically significant for each group. We further examined how the group differs from the overall sample by comparing the distribution of the variables that are listed in Table 4. In terms of sun sensitivity, Group 5 on the winter workdays is notably sensitive to sunlight during early mornings. In looking at the demographic break-down, this group has the largest number of full-time workers. In addition, it includes the highest share among all groups of those belonging to higher managerial and professional class (AB) and supervisory and junior managerial (C1), as well as lowest percentage of those in the DE social class. The group also includes both moe 18-35 and 36-45 year olds and is

also the group with largest percentage of households with no person staying in the house during the daytime. Meanwhile, Group 1 on the winter weekends seems more likely to have their entertainment/spare time during the later periods of the day. Indeed, G1 has the highest ratio of younger employed/self-employed individuals. In addition, it is also a relatively affluent group, since it includes the largest percent of people in AB and C1 social class, as well as the lowest in DE.

	Sun duration	Rainfall	Temperature
SW		G5: Q430(0.041); Q420(0.052); **Q410**(0.061); Q310(0.053); Q300(0.11)	G5: Q5418(0.042); Q420(0.061); **Q410**(0.069); Q401(0.032); Q310(0.043); Q300(0.063)
WW	G5: Q430(0.14); Q401(0.12); Q310(0.17); Q300(0.088)	G1: Q5418(0.01); Q401(0.011); Q310(0.013); Q402(0.012)	G6: Q420(0.18); **Q410**(0.21); Q310(0.16) Q300(0.23)
WR	G1: Q401(0.026); Q310(0.017); Q300(0.02)		

Table 4 Statistically significant variables

For the rainfall clustering, Group 5 is more likely to include employed young families or singles (within the 18-35 or 36-45 age categories) compared to other groups on summer workdays. The demographic analysis shows that the households include mainly those who live alone or live with children, but with the fewest number of families where all people are over 15 years old. Moreover, it has the highest percent of respondents where no one or at most one adult person remain in the house during the daytime. It also appears that this group would prefer to arrange their external activities during the mornings. On winter workdays, Group 1 have a wider and more sensitive response to rainfall, which could indicate more regular outdoor time for those households. This group is also more likely to have a higher educational degree and higher social class (AB and C1) with the fewest in the lowest social class (DE). In addition, it has the highest percentage of households that are employed or self-employed. Group 1 is most likely to include those with incomes above €50,000 and least likely to include those with

incomes of less than €15,000.

There are also some interesting differences between groups for the temperature clustering. Group 5 in the summer workday scenarios appears to be most sensitive during the early mornings. This group is most likely to have the highest education level, a family structure with significantly lower possibility of living alone or consisting of less than 2 adults and a greater likelihood of children and adults living together. Looking to age and employment status, it is seen that the group has an extremely high ratio of middle-age people (36-45) and fewer older members, as well as a much larger proportion of members of households bing employed full time. By contrast, Group 6 is sensitive throughout the day on winter workdays. Households in this group have a greater possibility of including those living alone and in low income groups. The ratio of being older than 65 years old and retired is dramatically higher in Group 6 and it also shows the highest number belonging to the DE social class. The structure explained why the group could be sensitive to temperature changes all day, since they are mainly staying at home.

5 Conclusion

The introduction of smart meters has brought opportunities for both utilities and policymakers to understand residential electricity consumption in greater depth. Due to the extremely large volume of high-resolution data, machine learning techniques have been used to investigate the information buried in metering data. Most studies have focused on load management, especially for demand forecasting and customer load profile segmentation and most implementation of clustering algorithms has been applied directly to metering data. There have been, however, few studies using the techniques to study daily life patterns within households. We introduce a novel method to detect household behaviour/daily patterns using clustering algorithms applied to weather variables. Our analysis proposes using the weather sensitivities as proxies for the

household daily life patterns, for instance, when a household tends to go out and at which periods of a day they have more spare time. The clusterings are not applied to meter readings but to the weather sensitivity coefficients. To reflect the differences in behaviour patterns between workdays and weekends in different seasons, the clusterings were conducted separately for seasons as well as for weekends versus workdays and for three weather variables – sun duration, temperature, and rainfall.

We are able to characterize clear differences in the daily patterns between workdays and weekends in summer and winter and how households respond to changing weather patterns. Based on the sun sensitivities, households are found to be less flexible and have less spare time during the middle of a workday while enjoying greater freedom during the afternoons. Households are more responsive to sun on summer weekends, which indicates greater discretionary time and outdoor activities during summer. The rain sensitivity profile curves tell us that stay-at-home family members tend to go out during late mornings and early afternoons regardless of season, probably due to having fewer fixed housework commitments such as cooking dinner and picking up children. Meanwhile, people are more likely to arrange outdoor activities in the evenings on summer weekends compared to workdays. The profiles which yield the fewest noteworthy differences are the temperature sensitivity curves. The statistical tests suggest that demographic features are most connected to rain sensitivities. In terms of seasonal differences, the clustering in workday scenarios reflect more about the household features in winter.

Loooking across all factors, the effect of social class in the clustering of weather sensitivity was not as significant as the occupancy-related variables. Living status, employment status, and the number of adults of the household are the main classifiers for all the clusterings. Among all weather variables, rain patterns are relatively more associated with variables of this kind, especially social class and income level.

This analysis could also serve as a starting point for classifying customers by their daily life

patterns. Understanding during which periods individuals may prefer to be outside of the home and when they are more likely to have spare time or be more flexible in their behavior patterns could be important when designing customised electricity price schemes. This work and the methods presented herein could be the basis of a new prediction model to classify existing or new customers' behaviour patterns and responses to weather conditions.

Appendix

Weather data comparison

We tried two weather datasets, 1) from the Dublin Airport station; 2) from the weighted weather dataset from four weather observatory stations (see Section 2.3.2). We examined the data similarity with the t-test. The result in Figure 1 demonstrated that the two datasets were highly similar. We further looked at the representative curves generated from these datasets and found that the trends are similar and comparable. Therefore, we decided to use the Dublin dataset for the following reasons:

- The differences between using the two datasets were not statistically significant.
- The Dublin dataset could retain the information of extreme weather conditions, while this information could be cancelled out when calculating the weighted dataset.

Weather variables	p-value
Temperature	0.998
Wind speed	0.995
Rain	0.998
Sun Duration	0.998
Relative humidity	0.998

Fig. 1 Weather data t test results

Cluster number selection

To cross-validate the suitable cluster numbers for different scenarios, we used combined techniques, including silhouettes scores, DBI scores, and silhouettes analysis. In general, the higher the silhouette score/the lower the DBI score it is, the better the clustering performance it means (see below from Fig 2 to Fig 4). It should be noted that no absolute optimal cluster number exists and it largely depends on the objectives of the research as well as the selection of validity

indices. Based on these rules, we chose seven as the cluster number for the workday scenarios for the sun duration as well as all the scenarios for temperature, while six was the optimal number for the sun and rain weekend scenarios.

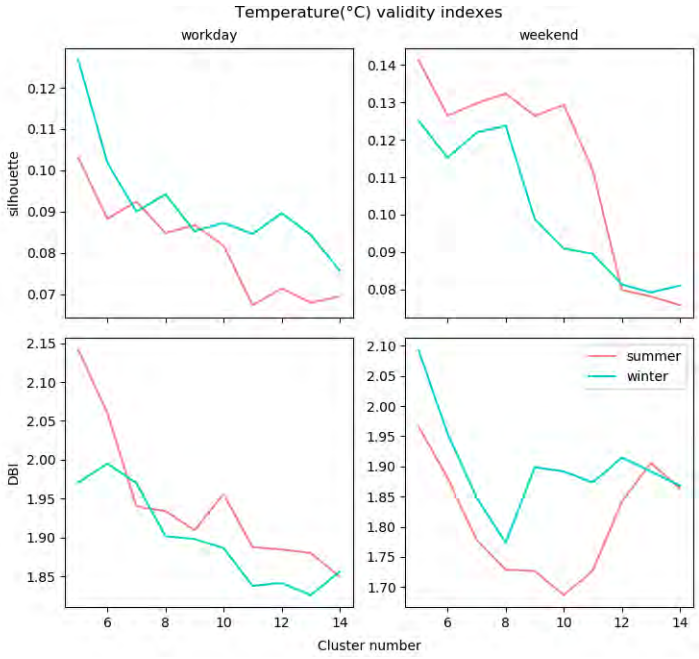

Fig. 2 Clustering validity indexes for temperature

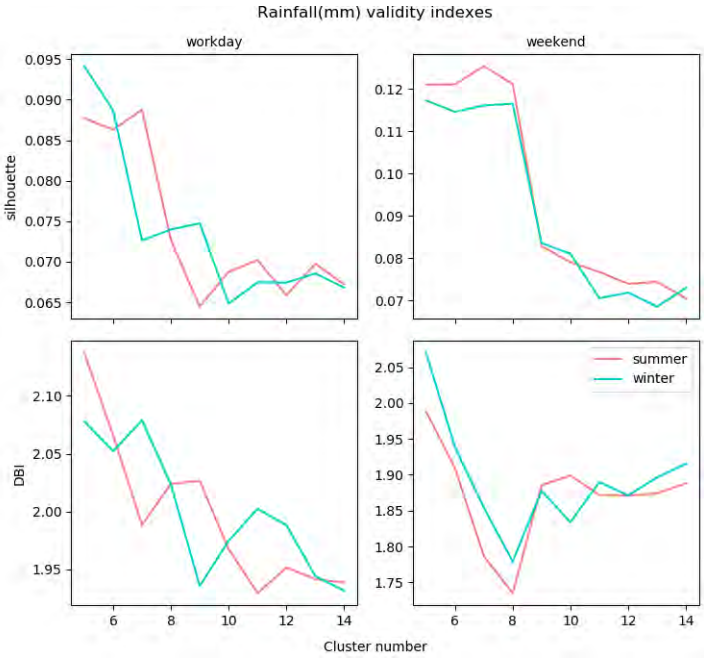

Fig. 3 Clustering validity indexes for rainfall

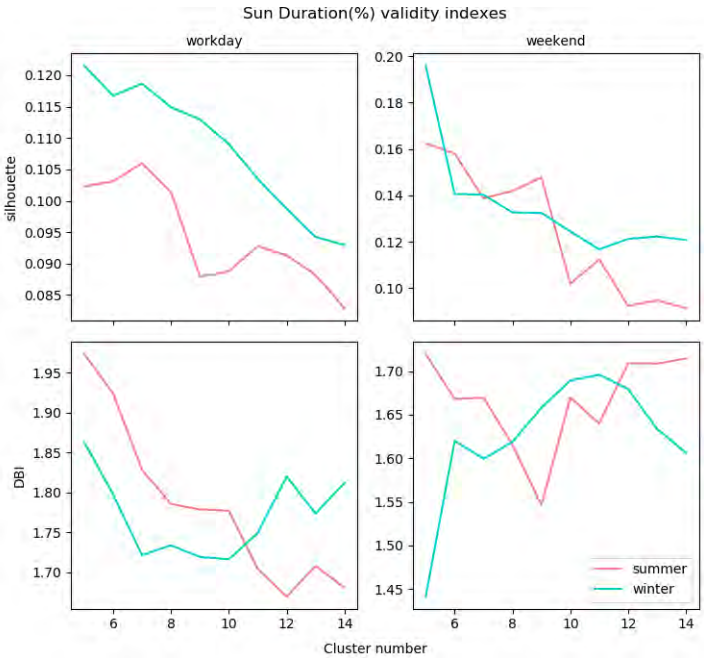

Fig. 4 Clustering validity indexes for sun duration

We only showed one of the silhouette analysis as an example here (Figure 5), since the actual analysis was significantly longer and remained irrelevant to the objective of this paper. In principle, a silhouette plot with evenly distributed areas across clusters and a high silhouette coefficient would be ideal. In this case, the right plot (cluster=7) would be better than the left (cluster = 6).

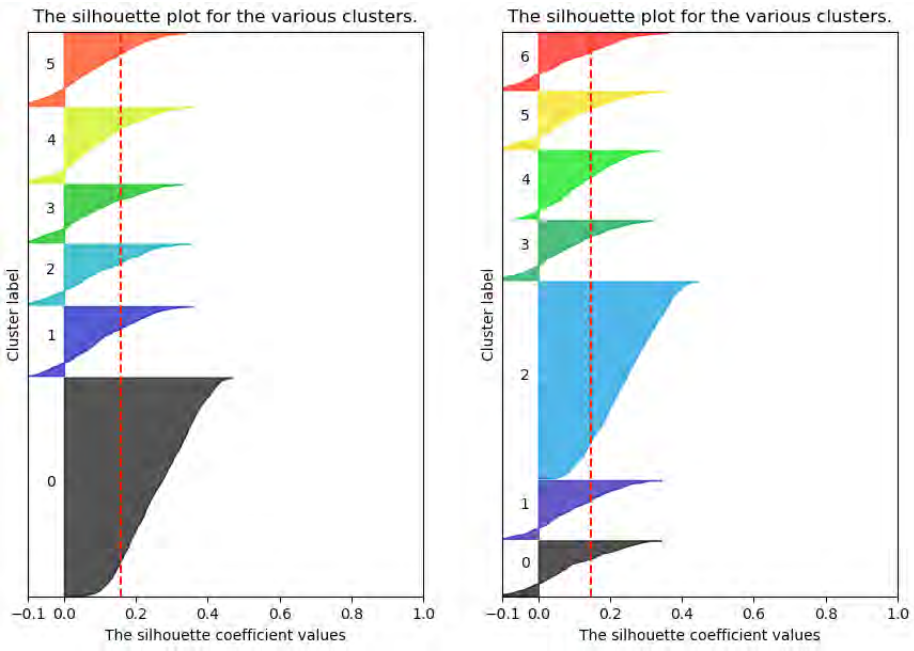

Fig. 5 Silhouette analysis for daily profiles, when cluster = 6 (left) and 7 (right)

Reference

Al-Wakeel, A., Wu, J. and Jenkins, N. (2017) 'k-means based load estimation of domestic smart meter measurements', Applied Energy. 194, pp. 333–342. doi: 10.1016/J.APENERGY.2016.06.046.

Alberini, A. and Filippini, M. (2011) 'Response of residential electricity demand to price: The effect of measurement error', Energy Economics, 33 (5), 889–895. doi: 10.1016/j.eneco.2011.03.009.

Albert, A. and Rajagopal, R. (2013) 'Building dynamic thermal profiles of energy consumption for individuals and neighborhoods', in Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013. pp. 723-728. doi: 10.1109/BigData.2013.6691644.

Beccali, M. et al. (2008) 'Short-term prediction of household electricity consumption: Assessing weather sensitivity in a Mediterranean area', Renewable and Sustainable Energy Reviews. 12(8), pp. 2040–2065. doi: 10.1016/J.RSER.2007.04.010.

Beckel, C. et al. (2014) 'Revealing household characteristics from smart meter data', Energy. 78, pp. 397–410. doi: 10.1016/J.ENERGY.2014.10.025.

Beckel, C. et al. (2015) 'Automated customer segmentation based on smart meter data with temperature and daylight sensitivity', 2015 IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 653–658. doi: 10.1109/SmartGridComm.2015.7436375.

Bianco, V., Manca, O. and Nardini, S. (2009) 'Electricity consumption forecasting in Italy using linear regression models', Energy. 34(9), pp. 1413–1421. doi: 10.1016/J.ENERGY.2009.06.034.

Blázquez Gomez, L. M., Filippini, M. and Heimsch, F. (2013) 'Regional impact of changes in disposable income on Spanish electricity demand: A spatial econometric analysis', Energy Economics. 40, pp. S58–S66. doi: 10.1016/J.ENECO.2013.09.008.

Cao, G. and Wu, L. (2016) 'Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting', Energy., 115, pp. 734–745. doi: 10.1016/J.ENERGY.2016.09

CER(Commission for Energy Regulation). (2012). CER Smart Metering Project -Electricity

Customer Behaviour Trial, 2009-2010 [dataset]. 1st Edition. Irish Social Science Data Archive.
SN: 0012-00. www.ucd.ie/issda/CER-electricity

Chen, B.-J., Chang, M.-W. and Lin, C.-J. (2004) 'Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001', IEEE Transactions on Power Systems, 19(4), pp. 1821–1830. doi: 10.1109/TPWRS.2004.835679.

Chen, Y. et al. (2017) 'Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings', Applied Energy. 195, pp. 659–670. doi: 10.1016/J.APENERGY.2017.03.034.

Commission for Energy Regulation (CER). (2012). CER Smart Metering Project - Electricity Customer Behaviour Trial, 2009-2010 [dataset]. 1st Edition. Irish Social Science Data Archive. SN: 0012-00. www.ucd.ie/issda/CER-electricity

Chicco, G. (2012) 'Overview and performance assessment of the clustering methods for electrical load pattern grouping', Energy. 42(1), pp. 68–80.
doi: 10.1016/j.energy.2011.12.031.

Chicco, G., Napoli, R. and Piglione, F. (2006) 'Comparisons Among Clustering Techniques for Electricity Customer Classification', IEEE Transactions on Power Systems, 21(2), pp. 933–940. doi: 10.1109/TPWRS.2006.873122.

Cramer, J. C. et al. (1984) 'Structural-behavioral determinants of residential energy use: Summer electricity use in Davis', Energy, 9(3), pp. 207–216. doi: 10.1016/0360-5442(84)90108-7.

Druckman, A. and Jackson, T. (2008) 'Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model', Energy Policy. 36(8), pp. 3177–3192. doi: 10.1016/J.ENPOL.2008.03.021.

Fan, S. and Hyndman, R. J. (2011) 'The price elasticity of electricity demand in South Australia', Energy Policy. 39(6), pp. 3709–3719. doi: 10.1016/J.ENPOL.2011.03.080.

Faruqui, A. and Sergici, S. (2010) 'Household response to dynamic pricing of electricity: A survey of 15 experiments', Journal of Regulatory Economics. 38, 193–225. doi: 10.1007/s11149-010-9127-y.

Faruqui, A. and Malko, J R. (1983). The residential demand for electricity by time-of-use: a survey of twelve experiments with peak load pricing. Energy. 8(10). pp.781-795.

Firth, S. et al. (2008) 'Identifying trends in the use of domestic appliances from household electricity consumption measurements', Energy and Buildings. 40(5), pp. 926–936. doi: 10.1016/J.ENBUILD.2007.07.005.

Ghofrani, M. et al. (2011) 'Smart meter based short-term load forecasting for residential customers', in 2011 North American Power Symposium. IEEE, pp. 1–5. doi: 10.1109/NAPS.2011.6025124.

Gouveia, J. P. and Seixas, J. (2016) 'Unraveling electricity consumption profiles in households through clusters: Combining smart meters and door-to-door surveys', Energy and Buildings. 116, pp. 666–676. doi: 10.1016/j.enbuild.2016.01.043.

Hackett, B. and Lutzenhiser, L. (1991) 'Social structures and economic conduct: Interpreting variations in household energy consumption', Sociological Forum. 6(3), pp. 449–470. doi: 10.1007/BF01114472.

Haider, H. T., See, O. H. and Elmenreich, W. (2016) 'A review of residential demand response of smart grid', Renewable and Sustainable Energy Reviews. 59, pp. 166–178. doi: 10.1016/J.RSER.2016.01.016.

Henley, A. and Peirson, J. (1998) 'Residential energy demand and the interaction of price and temperature: British experimental evidence', Energy Economics. North-Holland, 20(2), pp. 157–171. doi: 10.1016/S0140-9883(97)00025-X.

Herter, K. (2007) 'Residential implementation of critical-peak pricing of electricity', Energy Policy, 35 (4), pp. 2121–2130. doi: 10.1016/j.enpol.2006.06.019.

Herter, K., McAuliffe, P. and Rosenfeld, A. (2007) 'An exploratory analysis of California residential customer response to critical peak pricing of electricity', Energy, 32(1), pp. 25-34. doi: 10.1016/j.energy.2006.01.014.

Hor, C. L., Watson, S. J. and Majithia, S. (2005) 'Analyzing the impact of weather variables on monthly electricity demand', IEEE Transactions on Power Systems, 20(4), pp. 2078-2085. doi: 10.1109/TPWRS.2005.857397.

Irish Social Science Archive. (2005). 'Irish National Time Use Survey 2005'. Accessed from: https://www.ucd.ie/issda/data/irishnationaltimeusesurvey.

Jones, R. V., Fuertes, A. and Lomas, K. J. (2015) 'The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings', Renewable and Sustainable Energy Reviews., 43, pp.901–917

doi: 10.1016/j.rser.2014.11.084.

Karanfil, F. (2009) 'How many times again will we examine the energy-income nexus using a limited range of traditional econometric tools?', Energy Policy, 37(4), pp. 1191–1194. doi: 10.1016/j.enpol.2008.11.029.

Kavousian, A., Rajagopal, R. and Fischer, M. (2015) 'Ranking appliance energy efficiency in households: Utilizing smart meter data and energy efficiency frontiers to estimate and identify the determinants of appliance energy efficiency in residential buildings', Energy and Buildings. 99, pp. 220–230.
doi: 10.1016/J.ENBUILD.2015.03.052.

Kaza, N. (2010) 'Understanding the spectrum of residential energy consumption: A quantile regression approach', Energy Policy, 38(11), pp. 6574–6585.
doi: 10.1016/j.enpol.2010.06.028.

McLoughlin, F., Duffy, A. and Conlon, M. (2012) 'Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study', Energy and Buildings. 48, pp.240-248.
doi: 10.1016/j.enbuild.2012.01.037.

McLoughlin, F., Duffy, A. and Conlon, M. (2015) 'A clustering approach to domestic electricity load profile characterisation using smart metering data', Applied Energy. 141, pp. 190–199. doi: 10.1016/J.APENERGY.2014.12.039.

Newsham, G. R. and Bowker, B. G. (2010) 'The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: A review', Energy Policy. 38(7), pp. 3289–3296. doi: 10.1016/J.ENPOL.2010.01.027.

Pardo, A., Meneu, V. and Valor, E. (2002) 'Temperature and seasonality influences on Spanish electricity load', Energy Economics. 24(1), pp. 55–70. doi: 10.1016/S0140-9883(01)00082-2.

Quilumba, F. L. et al. (2015) 'Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities', IEEE Transactions on Smart Grid, 6(2), pp. 911–918. doi: 10.1109/TSG.2014.2364233.

Räsänen, T. et al. (2010) 'Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data', Applied Energy. 87(11), pp. 3538-3545. doi: 10.1016/j.apenergy.2010.05.015.

Razavi, R. et al. (2019) 'Occupancy detection of residential buildings using smart meter data: A large-scale study', Energy and Buildings. 183, pp. 195–208. doi: 10.1016/J.ENBUILD.2018.11.025.

Sanghvi, A. P. 1989. Flexible strategies for load/demand management using dynamic pricing. IEEE Transactions on Power Systems. 4(1), 83-93.

Sanquist, T. F. et al. (2012) 'Lifestyle factors in U.S. residential electricity consumption', Energy Policy. 42, pp. 354–364. doi: 10.1016/J.ENPOL.2011.11.092.

Sapankevych, N. and Sankar, R. (2009) 'Time series prediction using support vector machines: A survey', IEEE Computational Intelligence Magazine, 4(2), pp. 24-38. doi: 10.1109/MCI.2009.932254.

Silk, J. I. and Joutz, F. L. (1997) 'Short and long-run elasticities in US residential electricity demand: a co-integration approach', Energy Economics, 19(4), pp. 493–513. doi: 10.1016/S0140-9883(97)01027-X.

Ben Taieb, S. et al. (2016) 'Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression', IEEE Transactions on Smart Grid, 7(5), pp. 2448–2455. doi: 10.1109/TSG.2016.2527820.

Valor, E., Meneu, V. and Caselles, V. (2001) 'Daily Air Temperature and Electricity Load in Spain', Journal of Applied Meteorology, 40(8), pp. 1413-1421. doi: 10.1175/1520-0450(2001)040¡1413:DATAEL¿2.0.CO;2.

Viegas, J. L. et al. (2015) 'Electricity demand profile prediction based on household characteristics', in International Conference on the European Energy Market, EEM. August. doi: 10.1109/EEM.2015.7216746.

Wangpattarapong, K. et al. (2008) 'The impacts of climatic and economic factors on residential electricity consumption of Bangkok Metropolis', Energy and Buildings, 40(8), pp. 1419–1425. doi: 10.1016/j.enbuild.2008.01.006.

Weiss, M. et al. (2012) 'Leveraging smart meter data to recognize home appliances', in 2012 IEEE International Conference on Pervasive Computing and Communications, 2012, pp. 190–197. doi: 10.1109/PerCom.2012.6199866.

Wijaya, T. K. et al. (2015) 'Cluster-based aggregate forecasting for residential electricity demand using smart meter data', in 2015 IEEE International Conference on Big Data (Big Data). IEEE, pp. 879–887. doi: 10.1109/BigData.2015.7363836.

Wooldridge, J. M. (2013) Introductory Econometrics, Cengage Learning.
doi: 10.1016/j.jconhyd.2010.08.009.