

Capturing variation in daily energy demand profiles over time with cluster analysis in British homes (September 2019 – August 2022)

Martin Pullinger^{a,b,*}, Ellen Zapata-Webborn^a, Jonathan Kilgour^b, Simon Elam^a, Jessica Few^a, Nigel Goddard^b, Clare Hanmer^a, Eoghan McKenna^a, Tadj Oreszczyn^a, Lynda Webb^b

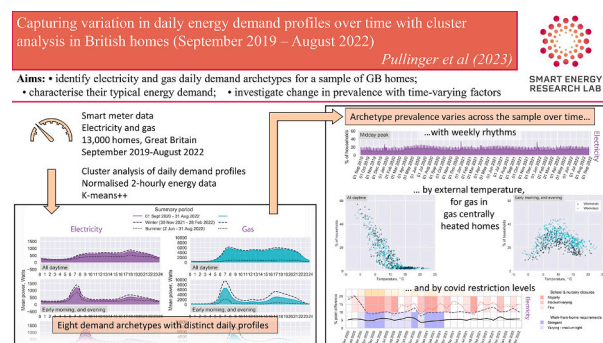
^a UCL Energy Institute, 14 Upper Woburn Place, London WC1H 0NN, United Kingdom

^b University of Edinburgh, School of Informatics, 10 Crichton St, Edinburgh EH8 9AB, United Kingdom

HIGHLIGHTS

- Eight typical domestic energy demand profiles identified using cluster analysis.
- Gas and electricity demand archetypes characterised for 13,000 homes over 3 years.
- Archetypes include 'All daytime' and 'Early morning, and evening' usage patterns.
- Seasonal variations in archetype energy demand, particularly for gas.
- Variations in prevalence on weekends, by temperature and during COVID-19 pandemic.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

Domestic energy demand profiles
Cluster analysis
Electricity and gas data
Temporal variation
Temperature variation
COVID-19

ABSTRACT

This study investigates typical domestic energy demand profiles and their variation over time. It draws on a sample of 13,000 homes from Great Britain, applying k-means cluster analysis to smart meter data on their electricity and gas demand over a three-year period from September 2019 to August 2022. Eight typical demand archetypes are identified from the data, varying in terms of the shape of their demand profile over the course of the day. These include an 'All daytime' archetype, where demand rises in the morning and remains high until the evening. Several other archetypes vary in terms of the presence and timing of morning and/or evening peaks. In the case of electricity demand, a 'Midday trough' archetype is notable for its negative midday demand and high overnight demand, likely a combination of the effects of rooftop solar panels exporting to the grid during the day and overnight charging of electric vehicles or electric storage heating. The prevalence of each archetype across the sample varies substantially in relation to different temporally-varying factors. Fluctuations in their prevalence on weekends can be identified, as can Christmas Day. Among homes with gas central heating, the prevalence of gas archetypes strongly relates to external temperature, with around half of homes fitting the 'All daytime' archetype at temperatures below 0 °C, and few fitting it above 14 °C. COVID-19 pandemic restrictions on work and schooling are associated with households' patterns of daily demand becoming more similar on weekdays and weekends, particularly for households with children and/or workers. The latter group had still not returned to pre-pandemic patterns by March 2022. The results indicate that patterns of daily energy demand vary

* Corresponding author at: UCL Energy Institute, 14 Upper Woburn Place, London WC1H 0NN, United Kingdom.

E-mail address: m.pullinger@ucl.ac.uk (M. Pullinger).

<https://doi.org/10.1016/j.apenergy.2024.122683>

Received 23 October 2023; Received in revised form 5 January 2024; Accepted 15 January 2024

Available online 5 February 2024

0306-2619/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

with factors ranging from societal weekly rhythms and festivals to seasonal temperature changes and system shocks like pandemics, with implications for demand forecasting and policymaking.

1. Introduction

Domestic energy demand in many countries is in a state of substantial flux, driven by ongoing changes to behaviour induced by the COVID-19 pandemic and subsequent exceptional energy price fluctuations, set against the longer-term spread of new low carbon technologies such as electric vehicles, heat pumps and rooftop solar panels. This is in addition to patterns driven by societal rhythms of work and schooling, and seasonal variation in demand for heating and cooling, among other factors. Policymakers and energy system engineers are faced with supplying secure, affordable and increasingly sustainable energy in response to uncertain changes in demand. Good data on patterns of energy demand can be of value to help reduce this uncertainty, by allowing changes in demand over time to be better observed, and how that demand responds to such rapidly-changing social and environmental factors to be investigated.

In the context of Great Britain, the growing availability of data from smart meters has facilitated a growth in research into such patterns. This has enabled domestic energy demand to be investigated in greater detail, including change over time and for different household types [1,2]. Research has further investigated how demand has varied in relation to specific major events such as the COVID-19 pandemic [3] and the recent cost of living crisis [4], as well as some of the behavioural factors underlying such variation [5,6].

To date, the majority of this research has focused on daily average demand, and in some cases average demand profiles, i.e. the average timings and sizes of peaks and troughs in demand over the day for a group of households over a period of time. This paper aims to contribute to this field of research by focusing on demand profiles in more detail, to investigate common demand profiles occurring behind such group averages, and the changes in their prevalence over time, for a broadly representative sample of British households, through the application of cluster analysis to their smart meter data. A substantial body of research has arisen that applies cluster analysis to granular smart meter data to identify distinct but commonly occurring demand profiles in the domestic, as well as industrial and commercial, sectors [7–10]. The aim of cluster analysis is to take a set of cases and segment them into a number of groups (clusters), such that cases within a group are more similar to each other than they are to those in other groups. In the case of daily energy demand profiles, this “reveals characteristic customer load profiles within the heterogeneous population” [7]; that is, the process identifies typical patterns of energy demand over the day (such as the timings and sizes of peaks and troughs), and allocates each case (for example, the demand profile from a given day and for a given home) to one of the identified patterns based on similarity. A range of clustering approaches exist for identifying ‘similarity’, many of which have been applied in the energy profile literature [7,8]. In the case of domestic energy demand, the outputs can be thought of as a set of typical energy demand ‘archetypes’, grounded in the empirical data, along with information on their prevalence across the sample being analysed. With the addition of linked contextual data, the factors shaping their prevalence can also be investigated.

There are several notable gaps in the literature to date relating to the clustering of domestic daily demand profiles. Firstly, most studies utilise datasets that are static and relate to periods prior to the COVID-19 pandemic, so the impacts on demand profiles of the extensive pandemic-induced changes in levels of working from home and other occupancy patterns have not been investigated. Secondly, existing studies deal almost exclusively with electricity demand. Only one research group was identified that has published peer-reviewed cluster analyses of gas demand profiles [11]. No studies were identified that

have included both fuels, electricity and gas, in the same analysis. This means that there is currently no research that has developed clustering results that allow direct investigation of the relationship between domestic electricity and gas demand archetypes.

In this paper, we aim to address these gaps. The paper has the following research aims:

- To identify and present domestic daily demand archetypes for both electricity and gas profiles for the same sample of homes, drawing on a dataset that includes periods before, during, and after the COVID-19 pandemic.
- To describe the characteristics of the resultant daily demand archetypes: their typical energy profiles, their average prevalence across the sample, and the relationships between electricity and gas archetypes.
- To investigate how the prevalence of these archetypes changes in the sample with time-variant factors, including weekly and annual societal rhythms and festivals, external temperature, and over the course of the COVID-19 pandemic.

We draw on the Smart Energy Research Lab (SERL) Observatory dataset [12–14] in this research. This is a longitudinal dataset of electricity and gas smart meter data and linked contextual data including weather, survey and EPC data, from a sample of 13,000+ consenting households that is broadly representative of the Great Britain¹ (GB) population. We draw on the full sample and three full years of data, from September 2019 to August 2022, so spanning periods prior to, during and after the main phases of the COVID-19 pandemic. In the current study, we focus on *daily* demand archetypes, i.e. taking each day's profile from each home and fuel as a separate case. In forthcoming work, we will be investigating *household* demand archetypes, i.e. identifying the typical daily demand profiles of households averaged over extended periods of time, and identifying how these vary with household characteristics.

The paper addresses the following research questions to achieve the above aims:

1. What are the typical energy profiles of the daily demand archetypes identified through the analysis?
2. How do the observed archetypes relate to the full sample's average demand profile?
3. Does the occurrence of the gas and electricity archetypes relate to each other?
4. How prevalent are the demand archetypes across the sample, and how does their prevalence vary over time in relation to different time-varying factors?

This paper therefore contributes to the existing literature by providing the first analysis of daily demand archetypes for both electricity and gas from the same sample of households, allowing them to be compared and contrasted. Secondly, it contributes by decomposing the average electricity and gas demand profiles of the sample to identify how they are composed of demand from households exhibiting the different archetypes. Finally, it provides a detailed investigation into how the prevalence of the demand archetypes for both fuels varies over time, including before, during and after the COVID-19 pandemic, and how such variation correlates with several different time-varying

¹ Great Britain consists of the United Kingdom excluding Northern Ireland, i.e. England, Wales and Scotland.

factors.

The paper has the following structure: [Section 2](#) reviews related literature, focusing on the potential end-uses of demand archetype analysis and existing findings on the predictors of demand archetypes. [Section 3](#) describes the SERL Observatory dataset in more detail. [Section 4](#) describes the methodology used to identify daily demand archetypes. [Section 5](#) presents results, structured around the research questions described above. [Section 6](#) discusses the results and concludes.

2. Literature review

In the context of Great Britain, existing studies have described average domestic demand profiles and their variation with household characteristics and time-varying factors. The Energy Follow Up Survey (EFUS), an extension of the English Housing Survey, describes heating season gas profiles based on data from 143 homes in England from October 2018 to April 2019. The profiles show very low overnight demand, and morning and evening peaks. The size of the peaks and relative size of drop in demand during the middle of the day is found to vary with building characteristics (floor area, and energy efficiency as measured for Energy Performance Certificates), occupant characteristics (number of occupants, daytime occupancy, fuel poverty status) and duration and pattern of heating use [2]. The SERL Statistical Report [1] presents average domestic demand profiles for the whole of 2021 for both gas and electricity, drawing on the same smart meter dataset from around 13,000 homes from Great Britain as used in the current study (and with many of the same authors). That finds similar morning and evening peaks in demand for gas to those found in the EFUS study, with demand falling during the middle of the day and being low overnight. In the case of electricity, a morning rise in demand from an overnight low is sustained through the middle of the day, then rises further in the evening before falling again. In most cases, breaking down by household group reveals similar average patterns, with only the sizes of peaks and troughs varying, e.g. demand is higher throughout the day at lower outdoor temperatures. However, for some groups, demand profiles are substantially different to the sample average. Homes with rooftop solar panels, for example, have, on average, morning and evening peaks in demand, and negative electricity demand during the middle of the day, indicating that the home is exporting electricity to the grid [1].

In terms of the existing literature that aims to identify common demand archetypes using cluster analysis, studies vary in terms of whether the target for clustering is taken to be demand profiles from individual days from each home, or the households themselves, in which case the typical demand profile of homes over periods of time are identified and clustered on. As our focus in the current study is on daily demand archetypes, in the literature review here we focus on publications relating to this level of analysis.² Literature dealing with household demand archetypes will be considered in our forthcoming paper that focuses on that level of analysis.

Three recent papers (from 2019 and 2020) already contain substantial reviews and evaluations of existing literature and methods [7,8,15]. We draw on these, supplemented with papers published subsequently (2020 onwards). The published studies of daily demand archetypes that we identified focus exclusively on electricity demand (the work on gas demand mentioned in the introduction takes households as the unit of analysis). We focus primarily on papers with an applied focus and/or using datasets of a similar scale to the one used in this current work, with household sample sizes into the thousands, and a duration of data of one or more full years.

Satre-Meloy et al. [15] reviewed 27 published articles in the field, with a mix of daily and household-level analyses, but all dealing with

electricity demand. Among the most common applications of the approach at the daily level were:

- to test and compare clustering approaches (i.e. methodological studies);
- to assess the stability and variability of daily demand profiles over time for individual households;
- and to investigate “the variability in timing of peak demand, the contributions of different customer segments to peak demand, or related time-of-day and seasonal effects on electricity consumption patterns”.

These different forms of analysis can have a variety of practical end uses for energy system actors, particularly energy companies managing the energy network. Several are discussed in the literature we reviewed, in many cases in the form of proposals rather than evaluated examples. These include:

- To better understand variation in patterns of customer demand [7];
- To improve the performance of load forecasting algorithms [7,8];
- To support “the detection of non-technical losses” [7,16], that is, electricity that is consumed but not billed for reasons such as inaccurate recording of consumption, defective appliances or deliberate fraud.

The end uses typically involve two stages of analysis: firstly, the identification of typical daily demand archetypes using cluster analysis, and secondly, the identification of factors that predict the archetype to which a particular case belongs.

While several of the reviewed papers mention that there is often high variability in demand profiles from day to day even for the same home, alluding to societal and weather-related factors, only one of the reviewed papers presented empirical analyses of the relationships between daily demand profiles and other factors. Czétány et al. [17] clustered daily electricity profiles from nearly 1000 homes from Hungary for January 2017 to December 2018, and found the prevalence of each cluster across the sample on a given day correlated with type of day (weekdays vs weekends) and season, along with minor differences along the lines of a combined building type/number of occupants variable. Peak demand was also found to vary by settlement type (village, town, city).

2.1. Key points

A variety of end uses for clustering daily energy demand profiles have been proposed in the existing literature. The existing research also points to several factors that vary over time that could impact on the chances of a particular home's electricity usage on a given day fitting a particular demand archetype, notably weekly societal rhythms and seasonal weather variation. However, among these studies, there is little empirical research looking at this relationship between the prevalence of daily demand archetypes and such temporally varying factors.

None of the studies reviewed uses energy data from during or after the COVID-19 pandemic, which would allow recent demand archetypes to be investigated. Gas archetypes have only been studied at the household level, and no studies have combined analyses of both electricity and gas. The current paper therefore contributes to the literature with its analysis of daily demand archetypes for both electricity and gas from before, during and after the main phases of the COVID-19 pandemic in Great Britain.

3. Dataset

The data used in this paper is the Smart Energy Research Lab (SERL) Observatory dataset [14,18]. The dataset contains half-hourly smart meter readings for electricity and for gas, where available, for a broadly

² Published research uses various terminology for daily demand archetypes, including substituting the words daily for diurnal, demand for load or consumption, household for home or dwelling, and archetype for profile or cluster.

representative sample of 13,292 homes recruited into the project from across Great Britain. Linked to the smart meter data are hourly-resolution weather variables from the ERA5 reanalysis climate data [19], along with survey data from the participants relating to building and occupant characteristics, among other variables. For more information on the SERL datasets see [14,20,21]. Data extends back for some households to August 2018. The 5th edition of the SERL Observatory dataset was used for this article [22], the most recent release of the dataset available at the time of analysis, which contains smart meter data up to the end of August 2022, and weather data up to the end of June 2022. New smart meter and linked contextual data continue to be collected, cleaned and periodically released by the SERL team, with the dataset made available to approved projects in the UK research community for research in the public interest. This provides scope for the current work to be updated over time.

The dataset has already been utilised to produce statistical reports of demand patterns and trends over time [1] and to investigate the range of predictors of those patterns [23], as well as the impacts of specific events such as the COVID-19 pandemic [3] and of household characteristics such as their EPC rating [24].

This article focuses on analysis of gas and *net* electricity demand. Net electricity demand is a household's demand from the electricity grid minus any electricity that the household generates, such as from rooftop solar panels. In households without such microgeneration, net electricity is the same as their total consumption from the grid. In households that do have microgeneration, it is their consumption minus their production. In the large majority of the homes with microgeneration, it takes the form of rooftop solar voltaic panels (solar PV), and the effect on their demand profile is to create a characteristic dip in net consumption in the middle of the day, frequently into net negative values.

3.1. Characteristics of the sample used in this study

The SERL Observatory sample was recruited in three waves between August 2019 and February 2021 using a stratified random sampling approach, with stratification along the lines of geographic region and Index of Multiple Deprivation (IMD) quintile (IMD is a common indicator in the UK of the relative level of deprivation of small geographic areas, based on measures of multiple dimensions of deprivation). As such, the sample is approximately representative of households in Great Britain in terms of their distribution by both these variables (with a slight overrepresentation of Wales and underrepresentation of Yorkshire) [14]. The sample is also approximately representative along the lines of several other characteristics (compared to the census and national surveys), with some biases as follows:

- Numbers of occupants: slight³ overrepresentation of 2-person households.
- Tenure: overrepresentation of owner-occupiers (by 15.7 percentage points) and slight underrepresentation of private renters and social renters.
- Managing financially (self-reported): slight overrepresentation of those reporting living comfortably.
- Property characteristics: slight overrepresentation of detached houses and underrepresentation of flats and tenements; overrepresentation of large properties (with 8+ rooms) (by 11.2 percentage points) and underrepresentation of medium-sized properties (4–5 rooms).

More detailed sample characteristics are available and published in the dataset's data descriptor [14].

For this study, we drew on the full sample of homes and on three full

years of data up to the most recently available time point, i.e. 1 September 2019–31 August 2022. Full years of data were used to help account for annual seasonal variations in climate in the analysis. Half-hourly smart meter data was collected from participants from up to three months preceding their date of joining the study (the data being stored locally on their smart meters for at least this long), however the recruitment timeline for the SERL project means that for many homes smart meter data does not extend back to September 2019. The number of households with available half-hourly data in the SERL sample increases in blocks, reaching the full sample in the third quarter of 2020. The early waves of recruitment were less representative of Great Britain, being focused on the south of England due to regional delays in the national smart meter rollout. However, as a proportion of all individual 'home-fuel-days' of data, data from this period represents only a small proportion of the full dataset analysed, and so any distortionary effects on the cluster results is likely to be small. The benefit of including data from this period was that it enabled analysis of the prevalence of demand archetypes before, during and after the main periods of the COVID-19 pandemic and associated restrictions.

Fig. 1 below shows the number of homes that had sufficient half-hourly smart meter data to identify their demand archetypes on each given day, separately for electricity and for gas. Counts reached a maximum of 12,020 homes with sufficient electricity data and 8930 homes with sufficient gas data.⁴ Fewer homes have available gas data because not all homes have gas supplies, and not all homes that have electricity smart meters have gas smart meters, even if they have gas supplies (while all homes with gas smart meters do have electricity smart meters, as the latter are required for the gas meters to communicate with the national smart meter infrastructure). In total, 18,890,620 data points (each data point being a single 'home-fuel-day' – a day of data for a given home and for a given fuel, either electricity or gas) were labelled with their demand archetypes across the full sample and three years covered.

Some of the analyses presented in the results section use data from a subset of these homes. Analyses in section 5.2.2 draw on data from 8480 homes that self-reported having gas central heating and had sufficient gas smart meter data to identify demand archetypes. Analyses in section 5.2.3 draw on data from 3650 homes in England and Wales that had sufficient electricity and/or gas data smart meter data to identify demand archetypes from January 2020 onwards.

4. Methodology

Fig. 2 below provides a graphical overview of the stages of data analysis used in this research. The following sections describe the details of the approach, starting first with how archetypes were generated, and then describing the approach taken to analysing the characteristics of the archetypes and factors correlating with their prevalence.

4.1. Generation of daily electricity and gas demand archetypes

As a first stage of data preparation, we cleaned the Observatory dataset following the same approach used for the SERL Statistical Report Volume 1 [1], such that half-hourly readings were retained only where they were not flagged as potentially erroneous or anomalous in the dataset, e.g. with implausibly high values or incorrect time stamps. Where necessary, gas usage measurements in cubic metres were converted to kWh following the same calculation used in the SERL Observatory dataset itself (which is also the national standard to convert meter readings to kWh), i.e. cubic metres * correction factor (1.02264) * calorific value (39.5) / conversion factor (3.6) = kWh.

Energy demand archetypes were then generated from the cleaned

³ Slight is used here to signify under- or overrepresentation of less than 10 percentage points. The sizes of larger differences are stated in the text.

⁴ Counts here and throughout this article are rounded to the nearest 10 for purposes of statistical disclosure control.

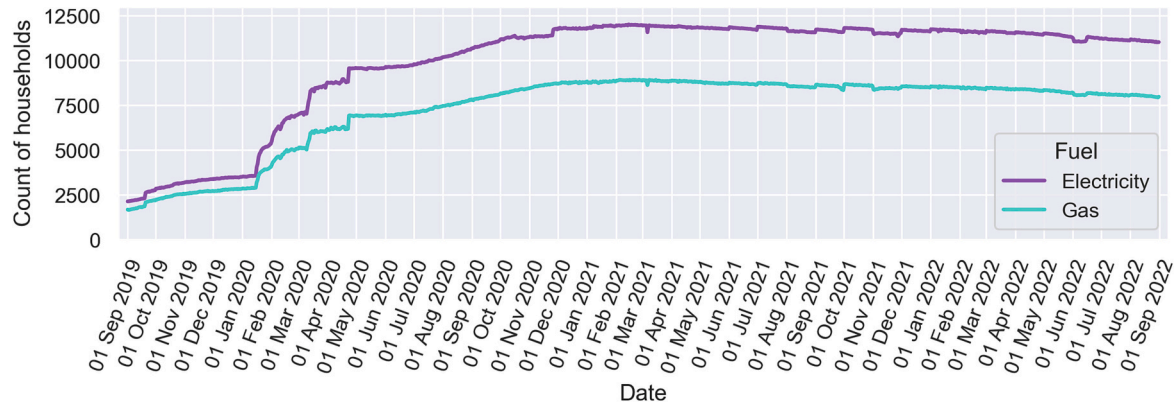


Fig. 1. Count of homes in the SERL Observatory dataset with sufficient smart meter data to calculate demand archetypes for each day from 1 September 2019 to 31 August 2022

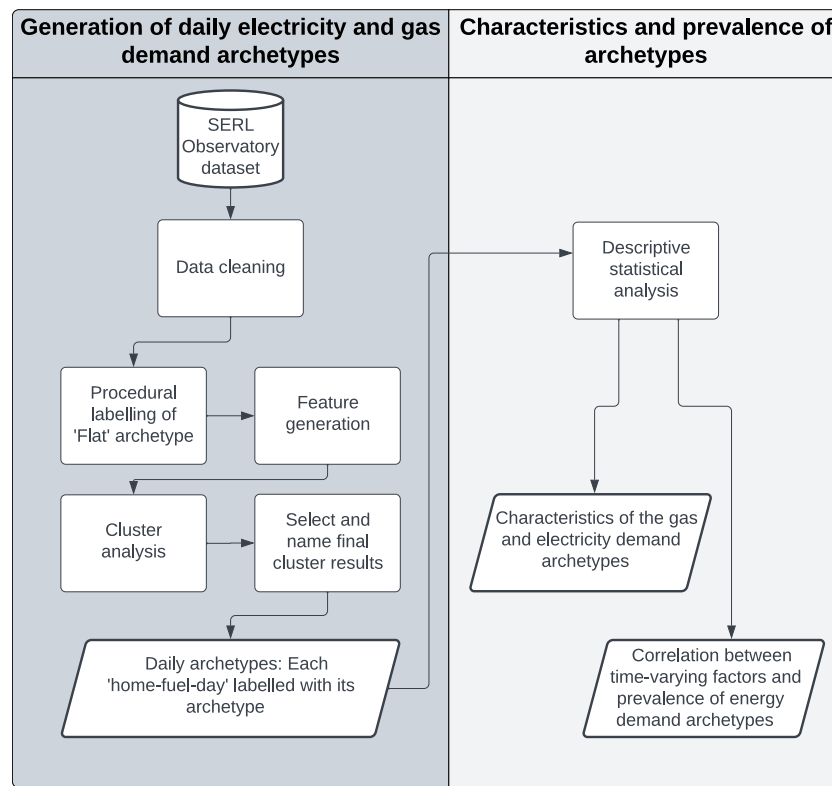


Fig. 2. Outline of the data preparation and analysis methodology used in this research. After initial data cleaning, daily demand archetypes are identified through initial procedural labelling and then using cluster analysis, so that each 'home-fuel-day' (each day of data for each home and fuel) is treated as a separate case and labelled. Analyses are then undertaken of the characteristics of these archetypes, and their relationship to time-variant factors. Arrowed lines indicate data flow; rectangles indicate stages of analysis; rhomboids indicate outputs of the analyses.

energy data through several stages of analysis. Each separate day of data for each home and each fuel (a 'home-fuel-day') was taken as a separate case to which to allocate an archetype. The details of the methodologies used in existing published research vary substantially (see, for example, the review in [7]); however, the approach we developed and applied in this study adapts the general stages commonly found whilst also resulting in archetypes for electricity and for gas that can be directly compared to one another. We also aimed to meet a further set of criteria for the resultant archetypes:

- Archetypes should be distinguished by the time of day of peaks and troughs, rather than their scale in energy terms. This reflects a substantive interest in whether the timing of a household's energy

demand coincides with system-wide peak times, and provides results that complement existing research that focuses instead on the scale of total daily demand rather than the timing of demand within the day.

- A relatively low number of archetypes was aimed for, with a target in the range of five to ten. This was to capture the diversity in patterns of demand in the data whilst producing results that could still be clearly communicated – fewer archetypes would likely miss important diversity, whilst higher numbers would become difficult to present and discuss.
- Given the large dataset and interest in the approach being relatively accessible for reuse, a computationally efficient method was preferred.

Given the wide range of clustering approaches that have been utilised in the existing literature, a range of input features and clustering methods were tested before selecting the final approach used in this research. The approaches selected for testing were initially expert driven, based on the existing literature and researcher experience, and were then vetted against the above criteria until a suitable approach was identified. Hierarchical cluster analysis and hdbscan clustering were both tested with various input features but were too computationally intensive to apply to the full dataset (in the case of the former), and unable to yield a substantively useful number of clusters despite testing a range of hyperparameters (in the case of the latter). K-means cluster analysis was eventually selected for use in this study, a commonly applied method in the field. The sections below detail the approach selected for use in this paper.

4.1.1. Procedural definition of a 'Flat' archetype

Before performing cluster analysis, we first defined a 'Flat' archetype as a home-fuel-day in which there was little variation in energy use over the course of the day. Such home-fuel-days are of low substantive interest as they contribute little to overall network peaks and troughs in demand. They also indicate little variation in the energy using activities in the home for that day. Such low fluctuations in energy use may have several origins. In the case of electricity, this includes appliances such as fridges and freezers that cycle periodically between low levels of energy use, and appliances left charging or on standby. In the case of gas, this might include low-level use by boilers, e.g. by combi-boilers that maintain a small reserve of hot water, pilot flames in some remaining older boilers, use for hand washing, etc. Such cases suggest few other appliances are used that day, and/or space heating has not been used.

We used a data-driven approach to identifying which home-fuel-days to label 'Flat'. We plotted histograms of the daily differences between minimum and maximum half-hourly values for each home-fuel-day in the cleaned dataset, which revealed a tri-modal distribution of the values for both fuels, i.e. three peaks, with two troughs between them. The initial trough in the histograms occurred at around 100 Wh for electricity and 300 Wh for gas. We took these values as the thresholds below which to label home-fuel-days 'Flat'. That is, individual home-fuel-days with differences in minimum and maximum half-hourly values below the threshold for the fuel in question were labelled as belonging to the 'Flat' archetype. These were also excluded from the subsequent cluster analysis, as to include flat profiles would have skewed the results due to the normalisation process, which stretches peaks and troughs, however small.

4.1.2. Feature creation and data tidying

The features used as inputs into the cluster analysis were then created for the remaining unlabelled home-fuel-days, again taking each home-fuel-day of data as separate cases. The following approach was taken:

1. Smart meter data were downsampled from the initial 30-min resolution to a two-hour resolution, taking the mean of the available readings for each two-hour block starting from midnight. Down-sampling is used to avoid 'the curse of dimensionality', where the more features involved, the more sparsely populated the feature space and the less distinguishable cases become from one another [8,25]. It is also used to reduce data size and the associated computational requirements [7], and can help to smooth shorter-term fluctuations in demand that amount to 'noise' [8]. A minimum of 3 of the 4 valid readings for each 2-h block for each home-fuel-day was needed, otherwise the value was set as missing. On days when the time zone changed (which occurs at 2 a.m. in the UK), the feature corresponding to either 12–2 a.m. (when clocks were changed back at 2 am by one hour to Greenwich Mean Time, the winter time zone in the UK) or 2–4 a.m. (when clocks were changed forwards at 2 am by one hour to British Summer Time, the summer

time zone in the UK) was based on the average of the six or two available readings, respectively.

2. For each home-fuel-day, the 12 two-hour values were then normalised by subtracting the minimum two-hour energy use for that home-fuel-day ('deminning') and dividing by the difference between the maximum and minimum two-hour energy use for the home-fuel-day. As such, each home-fuel-day of data had a minimum value of zero and maximum value of one, with intermediate values scaled linearly between zero and one. Deminming is used in the literature to enable a focus on 'discretionary' consumption, i.e. not baseload (for example, [15]). While some previous studies use input features that capture both the timing and size of demand (e.g. [26]), most normalise so that each home-fuel-day has a comparable scale. Such an approach enables a focus on the timing of demand and resultant shape of the demand profile rather than the magnitude of consumption over the day [8].
3. Home-fuel-days with any missing feature values were omitted from clustering, and their demand archetype was set to 'Missing'.

4.1.3. Cluster analysis

The cluster analysis method used in this research was k-means [27]. This is relatively light on computational resources, and also has the benefit that in future new data points can be accurately classified to the existing clusters based on which centroid they are closest to. This is useful in a case such as ours where the source data is updated periodically with new smart meter data, as in any future research, new points can be accurately classified to the existing archetypes, and hence compared to earlier data points and research based on them.

The number of clusters must be specified in advance for k-means cluster analysis. The optimal number can be identified through data-driven techniques such as identifying the number of clusters with the maximum silhouette coefficient, or the 'elbow' in the value of the sum of squared distances plotted against the number of clusters. In our case, these tests indicated the use of a substantially larger numbers of clusters than aimed for. As such, we identified an optimum lower number of clusters by producing dendrograms from the hierarchical cluster analysis approach that was also tested, using the same input features. Data size and resource constraints meant hierarchical cluster analysis could only be run on a small fraction of the full dataset at any one time, and so repeat runs using random samples of 35,000 home-fuel-days were performed. These all generated dendrograms indicating that 6 or 7 clusters were optimal, with the modal value being 7, within our target range of 5 to 10. K-means cluster analysis was therefore performed on the full dataset (just over 18 million home-fuel-days) using seven as the input parameter for the number of clusters. k-means++ initialisation [28] was used, with the best performant results of 10 runs of the algorithm taken, each run having a maximum number of iterations set to 300 (the best performant iteration converged on the solution in 32 iterations) [29].

The feature values of each cluster is presented in the appendix, section 10. To make the seven resultant clusters easier to communicate, we gave them descriptive names, which were based on each cluster's average demand profile. The names are: "All daytime", "Early morning, and evening", "Evening", "Late afternoon", "Mid morning", "Midday peak" and "Midday trough". The characteristics of each of these archetypes are described in the results, section 5.1. These names are used throughout the rest of the paper to refer to them.

4.2. Analysis of archetype characteristics and prevalence

The rest of the methodology describes analyses conducted on the resultant daily energy demand archetypes. The analyses comprise a description of their energy demand characteristics, their relationships to each other, and how different factors correlate with their prevalence over time across the sample of households in the SERL Observatory dataset. More details are provided below.

4.2.1. Characteristics of daily demand archetypes

The results section begins with descriptive analyses of the characteristics of the daily demand archetypes: their average prevalence over the two full years to August 2022; the average energy demand over the day for each archetype, varying by season; and the contribution of each archetype to the full sample's average energy demand profile. The prevalence of each daily demand archetype across groups of homes in the sample was calculated for each day, separately for electricity and gas. To do this, for each day and fuel, for each archetype, the number of households labelled with that archetype was divided by the total number of households with valid archetype labels on that day, and multiplied by 100 to give the percentage prevalence. Averages over periods of time weight each day's average value equally, to control for variation in levels of missing data over time. Average energy use for each half-hour was calculated for each group of households in a similar way, taking the average of the available labelled home-fuel-days each day, and weighting each day equally when calculating averages over time. We also use a chi-squared test of independence to investigate the correlation between the electricity and gas archetypes.

4.2.2. Relationship between daily demand archetype prevalence and time-variant factors

The literature review indicated that the likelihood of a home exhibiting a particular energy demand archetype for a particular day was correlated with factors which can, broadly, be characterised as either contextual factors that vary over time, or household factors that are typically invariant or at least stable over significant periods. Time-variant factors include the weather and patterns of work and schooling. Household-level factors include the age of the building, the type of heating fuel, and number of occupants. For time-variant factors, we investigate their relationship to the prevalence of each daily demand archetype across the sample. Household-level factors are considered in a forthcoming paper.

We plot the prevalence of each archetype in the sample over time for the three years 1 September 2019 to 31 August 2022. We then present the prevalence of each archetype each day plotted against mean external temperature that day, focusing on gas data for homes with gas central heating, as temperature primarily influences heating energy use. More details of the approach to each analysis is given along with the results to aid in their interpretation.

The final time-variant factor we investigate is the relationship between COVID-19 restrictions and daily demand archetypes. For this analysis, we calculated a metric based on the prevalence values that we called a 'weekday-weekend difference score'. This is an aggregate measure of the difference in the prevalence of each archetype between weekdays and weekends, summed across all the archetypes, calculated per calendar month. More details of the rationale for the score are given in the relevant section of the results.

The score was calculated per fuel and per calendar month, and separately for each of three groups of households defined by their family status: those with workers (people 16 or over in formal paid or unpaid employment) and no children (defined as occupants aged below 16 years), those with children (with or without workers), and those with no workers or children. These groups were defined based on the age bands of the occupants and work status data that was self-reported by participants when joining the SERL panel, and follows the approach used by Webb et al. (2023) [3].

The formula for calculating the weekday-weekend difference score for a given month, fuel and family status is presented below:

$$wwds = \frac{1}{2} \sum_{x \in M} |\bar{P}_{wd,x} - \bar{P}_{we,x}|$$

where:

- $wwds$ is the weekday-weekend difference score for a given month, fuel and family status;
- M is the set of all archetypes (excluding 'Missing');
- $\bar{P}_{wd,x}$ is the mean prevalence, P , of archetype x on weekdays, wd ;
- $\bar{P}_{we,x}$ is the mean prevalence, P , of archetype x on weekend days, we .

In short, the absolute difference in prevalence of each archetype for weekdays compared to weekends is calculated and the values for each archetype summed. Finally, the value is divided by two to give a maximum value of 100 and a minimum of 0 for a given month, fuel and family status. Change in these values over time is presented in the results and discussed in more detail there.

4.3. Software

Data analysis was performed using Python version 3.7.6 [30] and the following libraries: *pandas* 1.0.1 [31], *numpy* 1.18.1 [32], *scikit-learn* 0.22.1 [33], *scipy* 1.4.1 [34], Jupyter Lab 1.2.6 [35], *matplotlib* 3.1.3 [36] and *seaborn* 0.10.0 [37].

5. Results

The results presented below are grouped into two sections. [Section 5.1](#) presents analyses of the characteristics of the electricity and gas demand archetypes, including the relationships between them, addressing research questions 1 to 3 presented in the Introduction. [Section 5.2](#) addresses the fourth research question, presenting analyses of the prevalence of the daily demand archetypes over time, and the relationship between their prevalence and different time-variant factors.

5.1. Characteristics of the electricity and gas daily demand archetypes

5.1.1. Energy demand

[Fig. 3](#) below presents the average energy demand profile for each archetype, separately for electricity and for gas, at a half-hourly resolution. Given the large variation in demand over the course of a year, we present values for summer and winter periods as well as annual averages. Annualised values are based on the means for the two full years from 1 September 2020 to 31 August 2022 (we omit data from 1 September 2019 to 31 August 2020 because the majority of homes do not have data available for much of that period, and because of the disruptive effects of the COVID-19 pandemic). Winter and summer energy demand are each based on three months of data (13 weeks/91 days): 30 November 2021 to 28 February 2022 for winter, and 2 June to 31 August 2022 for summer.

The archetypes vary from each other primarily in terms of the presence and exact timing of morning and evening peaks, and the extent to which demand through the middle of the day dips or is maintained relative to those peaks. For the 'All daytime' archetype, average demand begins to rise from an overnight low at around 6 am, then remains higher with some variation until the late evening. The 'Early morning, and evening' archetype shows a more pronounced morning peak and a smaller but distinct evening one, with lower energy use during the middle of the day. The next four archetypes in the figure can be differentiated by variation in the timing of a single peak, with relatively lower energy use for the rest of the day. The profile of the 'Midday trough' archetype shows a distinct pattern, defined by lower energy use during the middle of the day compared to overall higher use in the evening and overnight. Average energy demand for the 'Flat' archetype, as per its initial definition, remains nearly constant, and is also very low.

Comparing electricity and gas archetypes, electricity demand is overall lower in every archetype than it is for gas. This reflects the different uses of the two fuels, particularly that gas is the primary heating fuel in most homes in the source dataset, as it is for Great Britain generally. The electricity archetypes also exhibit generally less variation

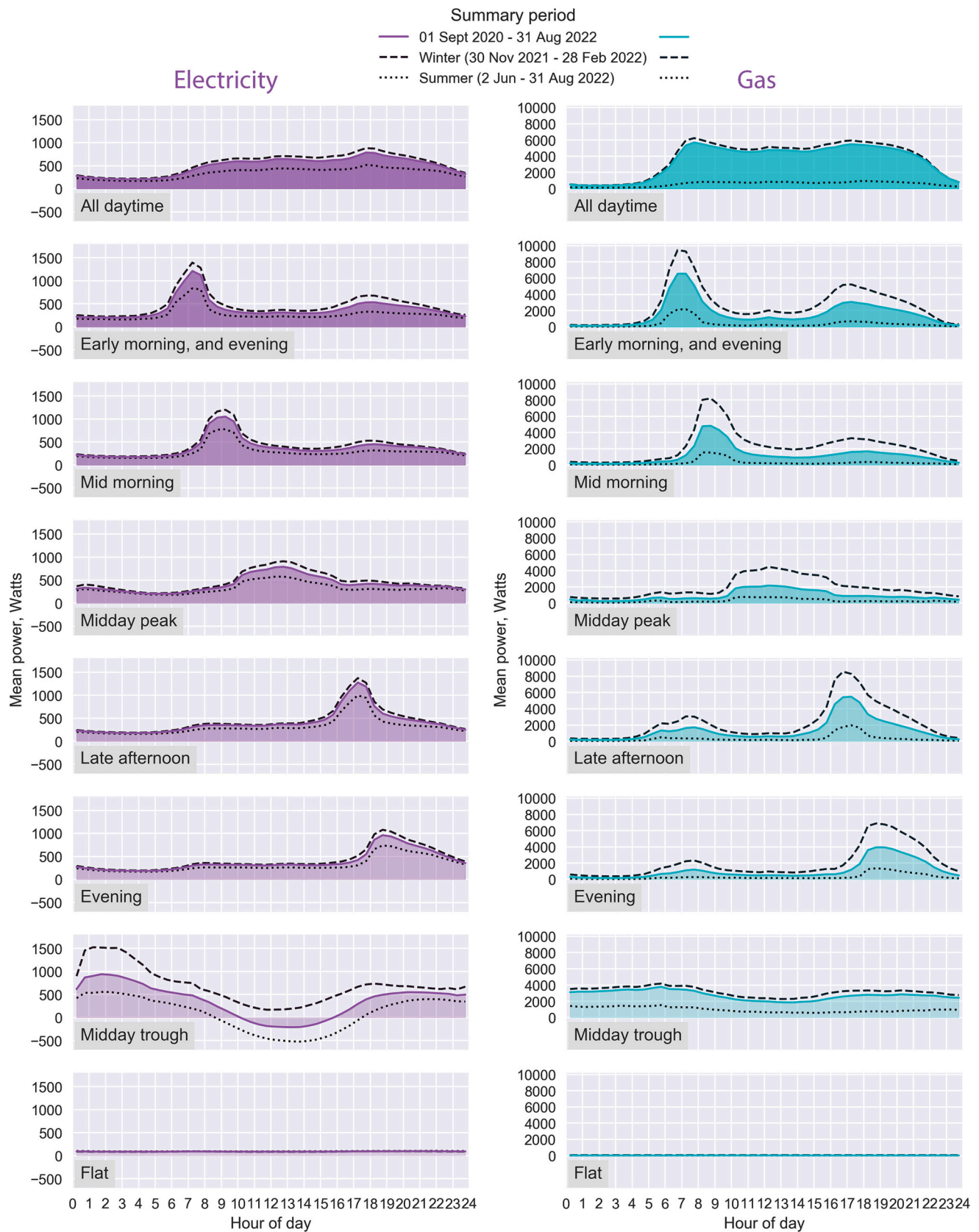


Fig. 3. Mean energy use over the day for each energy demand archetype, for each of: the two years 1 September 2020–31 August 2022; Winter 2021–2022 (31 November 2021–28 February 2022); Summer 2022 (2 June - 31 August 2022). Electricity and gas shown separately. Note different y-axis scales for the two fuels.

than gas archetypes over the course of the day; in particular, overnight demand for electricity is higher as a proportion of daytime demand than for gas.

The use of gas as a heating fuel results in much higher seasonal variation in demand in the gas archetypes than is seen for electricity. For gas, average summer demand does not exceed 2 kW in any given half hour for any of the archetypes; by contrast, it reaches close to 10 kW in several archetypes in winter.

The 'Midday trough' archetype is somewhat distinct from the others for both fuels. In the case of electricity, it shows much higher seasonal variation than the other archetypes, higher overnight demand, reaching over 1.5 kW in winter, and negative demand during the middle of the day in summer and on average over the year. This pattern can be explained by the fact that the 'Midday trough' electricity demand archetype occurs primarily in homes with rooftop solar photovoltaics, and represents net grid demand (i.e. consumption minus generation) – in Great Britain, excess household generation can typically be exported back to the grid, explaining the average negative values for this archetype during the middle of the day. We speculate that the high overnight demand may be the result of a combination of overnight charging of electric vehicles and electric storage heaters; future work could investigate this further. In the case of gas, the 'Midday trough' archetype is the only one with high overnight demand, similar to daytime levels. This and the relatively high summertime demand is consistent with it being exhibited by homes that keep their heating turned on overnight as well as during the day, and throughout the year. Again, future work could investigate this further.

5.1.2. Average prevalence of the archetypes

Table 1 presents the average prevalence of each archetype across the sample of households over the two years 1 September 2020 to 31 August 2022, split by fuel.

In the case of electricity, on average (mean), 20% of homes on any given day exhibit the 'All daytime' archetype. Aside from 'All daytime', the most common archetypes are those with higher energy use towards the middle or later part of the day, i.e. 'Midday peak', 'Late afternoon' and 'Evening'. 'Midday trough' occurs in 8% of homes on an average day. Variation in the prevalence of the archetypes from day to day across the sample is generally moderate, as indicated by the standard deviations, which vary from 1.5 to 5.0 percentage points.

Gas archetypes, by contrast, have rather different patterns of

Table 1
Percentage prevalence of each archetype across the sample, by fuel, for 1 September 2020 to 31 August 2022

	Mean	Median	Standard deviation	Minimum	Maximum
Electricity					
All daytime	19.6	18.2	5.0	12.8	34.2
Early morning, and evening	5.0	5.2	1.5	1.3	8.6
Mid morning	8.7	8.7	1.5	4.6	13.5
Midday peak	15.4	14.1	3.6	9.8	34.0
Late afternoon	15.3	14.8	2.6	7.9	24.3
Evening	23.1	23.5	3.3	7.6	30.5
Midday trough	8.2	8.8	2.1	3.3	12.1
Flat	4.7	4.4	2.2	1.4	13.0
Gas					
All daytime	12.0	6.3	12.4	1.5	56.3
Early morning, and evening	25.5	25.3	6.5	9.4	43.7
Mid morning	12.8	12.7	3.0	6.6	19.3
Midday peak	11.2	10.3	4.7	3.7	23.5
Late afternoon	13.6	13.6	2.5	5.9	23.3
Evening	13.4	14.2	3.9	3.1	23.2
Midday trough	3.7	4.0	2.5	0.2	10.3
Flat	7.6	5.7	4.5	2.3	20.4

prevalence, which is to be expected given the substantially different end uses of the two fuels. Across the two years, the most common gas archetype was 'Early morning, and evening', with just under 26% of homes exhibiting this archetype on average. The 'All daytime' archetype, and the archetypes representing a single peak at some point during the day, are all approximately equally common: 12–14% of homes exhibit each of them, on average. The prevalence of a few of the archetypes varies substantially over the period, most notably 'All daytime', with a standard deviation of over 12 percentage points, and which occurs in over 56% of households on its most prevalent day, more than four times as many as the average over the period. 'Early morning, and evening' also has a high maximum prevalence, at 44% of households. This variability in the prevalence of the gas demand archetypes is returned to in section 5.2.2.

5.1.3. Decomposition of average daily demand profiles

Fig. 4 plots the average daily energy demand profile of the full sample of households for 1 September 2020 to 31 August 2022. This is decomposed to show how the different energy demand archetypes combine together to generate this average profile. In essence, this stacks the individual demand archetypes from Fig. 3 multiplied by their mean prevalence across the sample from Table 1.

The figure reveals how for both fuels, the average daily energy demand from all the homes in the sample is the result of energy demand archetypes that each appear quite different from that average. For example, two archetypes, 'Late afternoon' and 'Evening', do much to create the high peak in evening demand for both fuels.

In the case of electricity, the 'Midday trough' archetype contributes much to increasing overnight demand. It contributes over a third of total nighttime demand after midnight, despite only 8% of homes on average exhibiting this profile. Similarly, it accounts for much of the midday dip that is observed in the full sample. The 'Midday peak' archetype, conversely, substantially reduces the size of the dip in average demand that there would otherwise be during the middle of the day.

For gas demand, most overnight demand is due to homes in the 'Midday trough' archetype, despite less than 4% of homes fitting that profile on average. The 'All daytime' archetype contributes a large proportion of total daytime demand, while the 'Midday peak' archetype contributes to smoothing the trough in demand in the middle of the day.

5.1.4. Relationship between electricity and gas archetypes

It might be expected that a home's electricity demand archetype on a given day is correlated with its gas demand archetype: patterns of occupancy and sleep, for example, can be expected to influence both. We therefore tested for relationships between electricity and gas daily demand archetypes, using a chi-squared test of independence. We tested for relationships separately for the winter 2021–22 and the summer 2022 periods defined earlier. This allowed us to investigate relationships in different weather and seasonal conditions.

In both periods, the tests indicated that the occurrence of the archetypes was highly statistically significantly correlated. However, residuals indicated that this was entirely due to a strong relationship between the fuels in the occurrence of the Flat archetype. As a confirmation, rerunning the chi squared tests with the Flat archetype omitted from the contingency table yielded non-significant results for both winter 2021–22 and summer 2022, i.e. apart from the Flat archetype, no statistically significant relationships were found between homes' electricity and gas demand archetypes on a given day.

5.2. Variation in the prevalence of daily demand archetypes with time-variant factors

5.2.1. Chronological variation

The prevalence of each demand archetype over time is plotted in Fig. 5 (for electricity) and Fig. 6 (for gas). These show the percentage of households exhibiting each daily demand archetype each day from

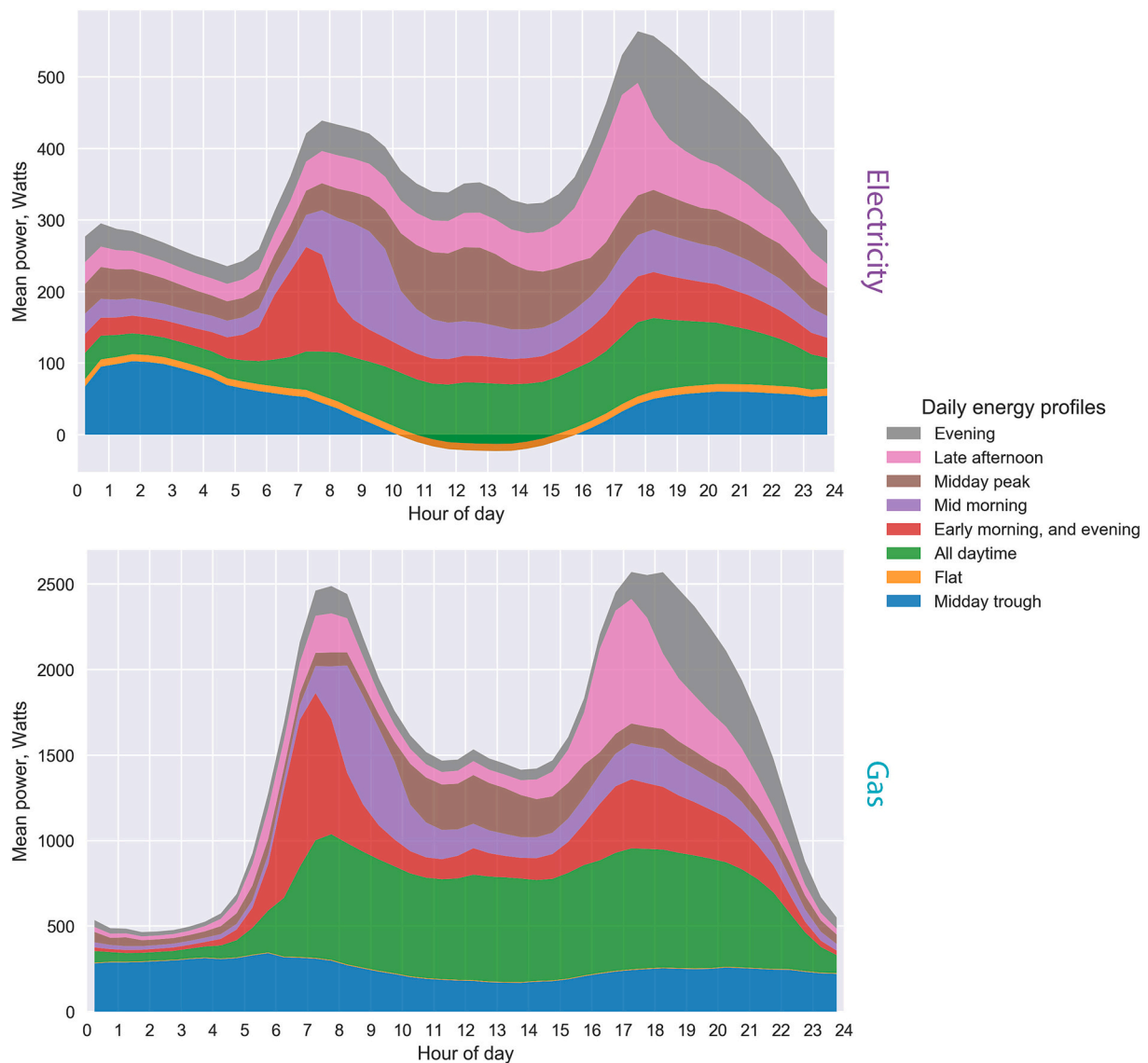


Fig. 4. Stacked area plots indicating how different archetypes contribute to the overall daily profiles seen for electricity and gas demand across the full sample, averaged for 1 September 2020 to 31 August 2022.

September 2019 to August 2022. (Note that due to the timing of participant recruitment waves, sample sizes are small until February 2020, reaching about 75% of the full sample by May 2020, so comparisons of prevalence before and after approximately mid-2020 should be treated with caution). Patterns of change in the prevalence of certain archetypes can be discerned. For example, for electricity, the 'Midday peak' and 'Evening' archetypes show consistent differences in their prevalence between weekdays and weekends (visible as frequent and regular spikes and dips in the figure). This weekly rhythm is consistent with changing weekday and weekend patterns of occupancy and time use, and hence timing of energy using activities, likely driven by standard daytime Monday-Friday work and school patterns.

Seasonal variation can be readily seen for both fuels in the prevalence of the 'All daytime' archetype, among others. In the case of gas, the primary space-heating fuel for most homes in Great Britain, this is highly influenced by external temperature, which accounts for much of the day-to-day fluctuation apparent in the figure (see [section 5.2.2](#) for more analysis of this relationship).

One-off events, notably Christmas Day (25th December), commonly observed in Great Britain, can be seen in the form of spikes or dips in the prevalence of several archetypes on that day each year, notably the

electricity 'Midday peak', 'Late afternoon' and 'Evening' archetypes. We speculated that this was driven by the use of electric ovens to cook Christmas dinners, traditionally prepared for consumption as a midday meal, with a concomitant reduction in cooking for an evening meal (evening is the most common time for eating a hot meal in the majority of UK homes on most other days). We reproduced the plots omitting households that self-reported having electric ovens (as part of the initial recruitment survey), and found the spikes in prevalence in the 'Midday peak' archetype to be almost absent, while the dips in the 'Late afternoon' and 'Evening' archetypes were smaller (plots not presented here for space purposes). Also visible around Christmas Day each year in [Fig. 5](#) and [Fig. 6](#) is a spike in the prevalence of the 'Flat' archetype, likely from people staying with relatives or friends for a day or two, leaving their homes unoccupied (note also that the prevalence of the 'Flat' archetype is not substantially affected by omitting homes with electric ovens, which is what would be expected if the spikes around Christmas Day are due to an increase in unoccupied homes).

There are also some changes in archetype prevalence in the plot that appear consistent with impacts of COVID-19 restrictions, particularly around the start of the first lockdown, from 23 March 2020. For electricity, there is a reduction in the prevalence of the 'Early morning, and

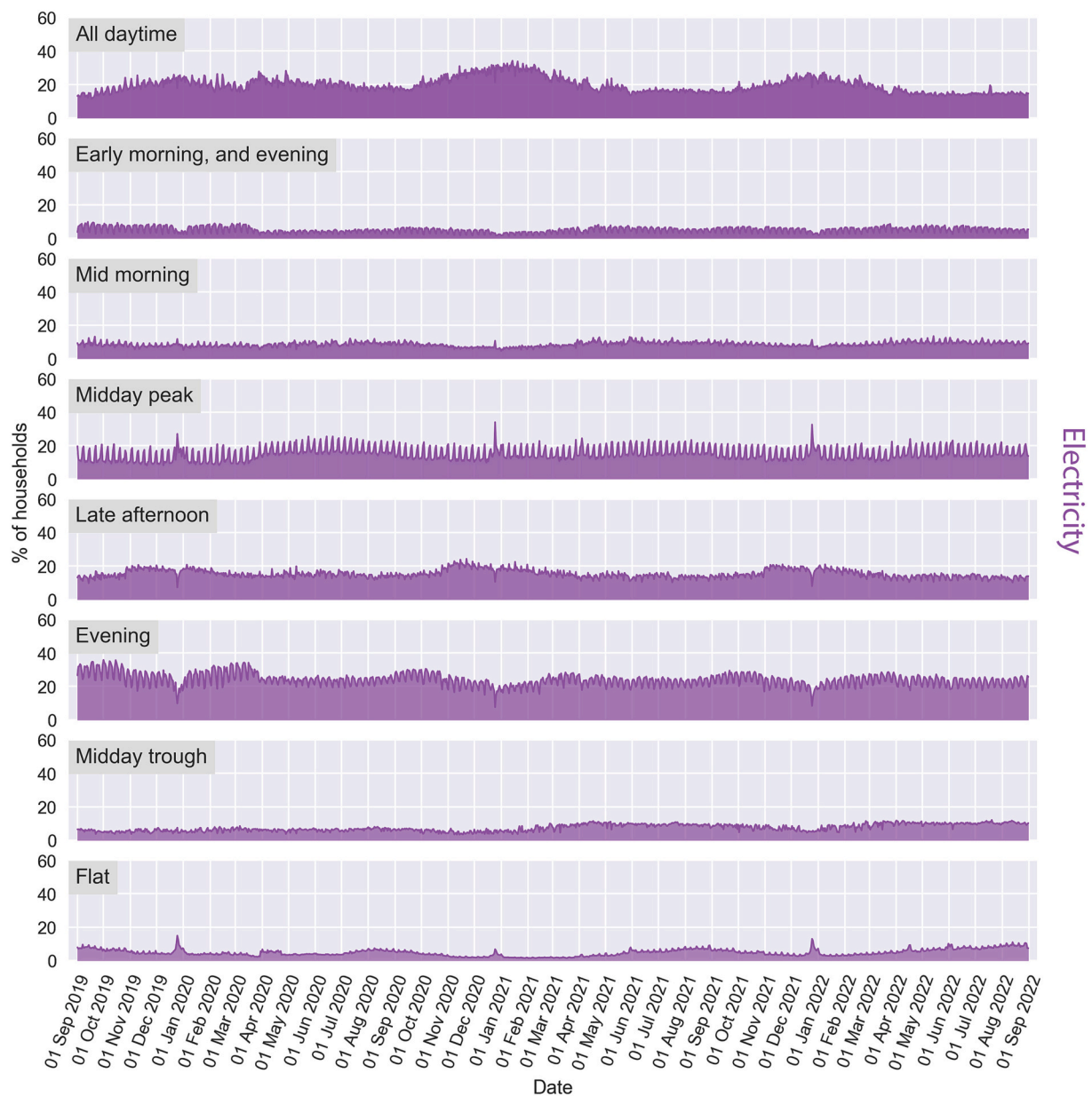


Fig. 5. Line graph of the prevalence of the electricity demand archetypes each day from 1 September 2019 to 31 August 2022
n homes = 2140–12020 (varies by data point)

evening' and the 'Evening' archetypes on weekdays around the start of the first lockdown, and an increase in the prevalence of the 'All daytime' and 'Midday peak' archetypes, particularly on weekdays, which could be indicative of increased daytime occupancy. The 'Flat' archetype also increases for a period in the electricity data, possibly as some people opted to move out and stay with others temporarily. For the gas data, the 'Mid-morning' archetype becomes relatively more common on weekdays compared to weekends. Other changes around the time, such as in the prevalence of the 'All daytime' archetype for gas, are likely the result of changes in external temperature and associated space heating demand. The changes in many cases seem quite subtle, so the impacts of COVID-19 restrictions are investigated with additional analyses in [section 5.2.3](#) below.

5.2.2. Variation in gas archetype prevalence with external temperature

Here we investigate the relationship between the prevalence of daily demand archetypes and external temperature. As gas is the primary central heating fuel in Great Britain, the analysis focuses on gas

archetypes only, and on the subsample of homes from the SERL Observatory dataset that self-reported having gas central heating systems.

[Fig. 7](#) shows scatter plots of the relationships between the prevalence of each gas daily demand archetype and mean external temperature. Each point represents the prevalence of the given archetype on a single day across the sample of homes with gas central heating, for the two years 1 July 2020 to 30 June 2022.⁵ Complete years of data were preferred for the current analysis, as external temperature follows annual cycles. Colour coding distinguishes weekdays from weekends, to reveal the interaction between temperature and this driver of occupancy.

All the archetypes show a clear relationship between their prevalence across the sample on a given day and mean external temperature that day. The prevalence of the 'All daytime' archetype in particular is

⁵ Although the smart meter data in the dataset runs until August 2022, June 2022 is the last month in the dataset that also has weather data.

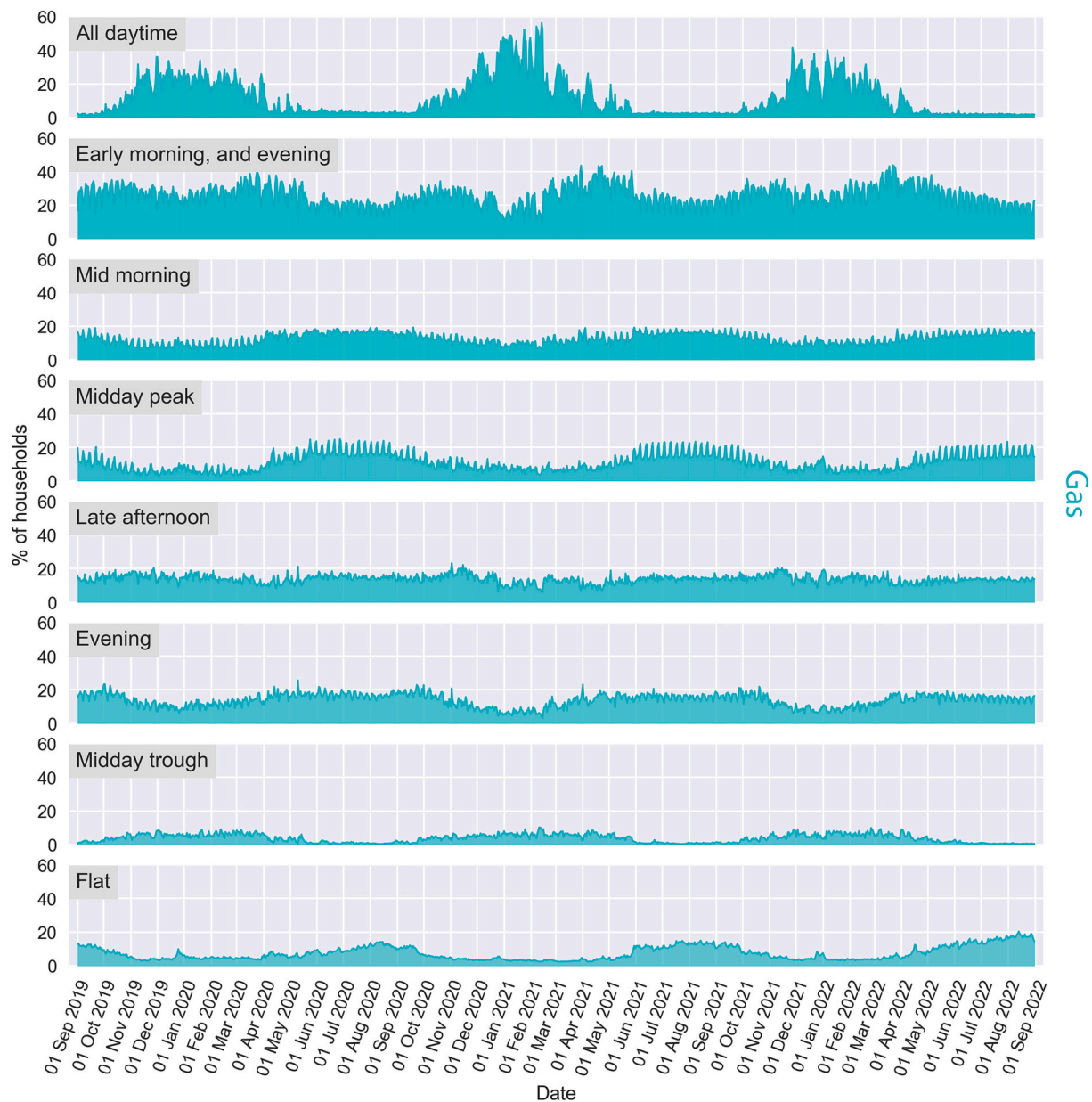


Fig. 6. Line graph of the prevalence of the gas demand archetypes each day from 1 September 2019 to 31 August 2022
n homes = 1660–8930 (varies by data point).

strongly inversely correlated with external temperature up until about 14 °C is reached. Below 0 °C, the prevalence exceeds half of the sample on some days, while it drops close to zero from around 14 °C and above. This is an archetype that is consistent with the central heating being on throughout the waking hours of the day (e.g. because it is programmed to maintain either a comfortable indoor temperature or a lower set-back temperature), a pattern that would be expected to be found in more households as temperature decreases.

The ‘Early morning, and evening’ archetype also shows a strong relationship with external temperature, although this time a curvilinear one. This is consistent with heating being used intermittently in the shoulder seasons (spring and autumn), where heating is only required for part of the day in order to maintain comfortable indoor temperatures. At lower temperatures, this pattern gives way to the ‘All daytime’ archetype, while at higher temperatures it gives way to other archetypes. The interaction with occupancy is clear in this plot – during the week, when more homes are unoccupied during the day, the ‘Early morning, and evening’ archetype is more common than it is on

weekends. The ‘All daytime’ archetype is on average more common at weekends than on weekdays at any given temperature, most likely as more homes are occupied through the day, although the pattern is less strong than for the ‘Early morning, and evening’ archetype.

The remaining archetypes, except ‘Midday trough’, become more common as temperature increases – each would be consistent with a single period of heating, either in the morning or the evening, at external temperatures when less heating is required to maintain a comfortable indoor temperature. The consistently higher prevalence of the ‘Mid morning’ and, in particular, ‘Midday peak’ archetypes on weekends compared to weekdays is again consistent with higher daytime occupancy across the sample on weekends.

5.2.3. System shocks: The case of COVID-19

The changes in the prevalence of demand archetypes over time presented earlier in Fig. 5 suggest that there were changes related to the COVID-19 pandemic, at least around the period of the first lockdown, from 23 March 2020. However, as discussed above, the changing sample

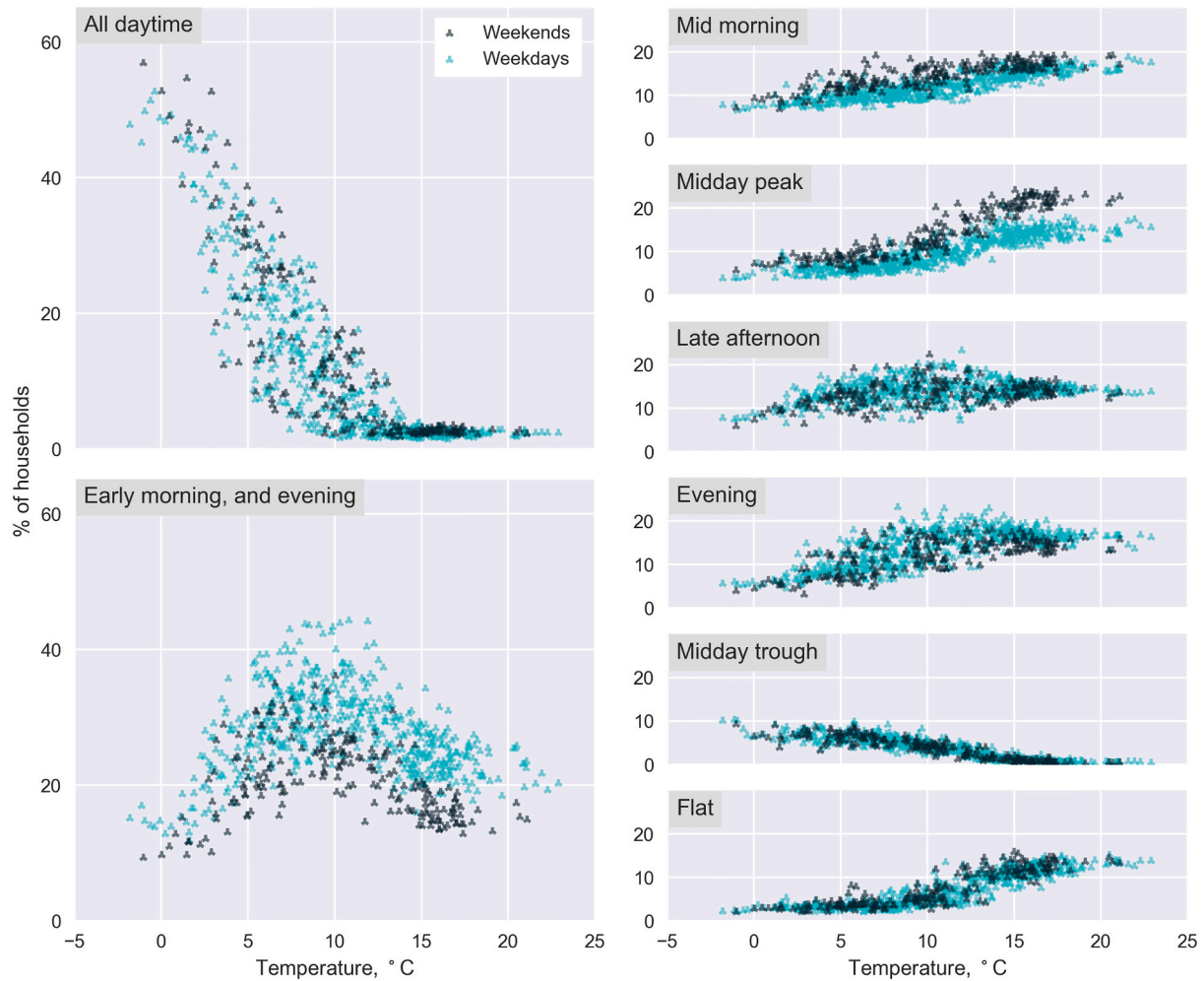


Fig. 7. Scatter plot of the relationship between daily demand archetype prevalence against mean external temperature, for gas, for homes with gas central heating, for 1 July 2020 to 30 June 2022. Each point indicates the prevalence of the archetype across the selected homes for one day. n homes = 6720–8480 (varies by data point).

numbers over that period add a confounding factor to the observations. Here, we refine the analytical tools to attempt a more powerful investigation of possible COVID-19-related impacts on the prevalence of the different daily demand archetypes.

Firstly, we take a fixed subsample of the households that had data throughout a period from shortly before the start of the pandemic to beyond the end of the last major restrictions on behaviour. We selected January 2020 to March 2022 for the analysis. We also focused on England and Wales, removing Scottish households for two reasons: there were very few Scottish households in this subsample due to the timetable of recruitment of participants into the SERL Observatory panel; and Scotland had substantially differing COVID-19-related restrictions. The sample consists of 3650 households in total: 3540 of which had electricity data, and 2850 of which had gas data.

Secondly, COVID-19-related restrictions impacted different groups to different extents. Whilst some restrictions on leaving home applied to the whole population, such as ones related to socialising and attending leisure venues, two major sets of restrictions varied by household type: work-from-home rules impacted households with workers that were deemed ‘non-essential’, whilst home-schooling rules impacted all households with school-attending children. We therefore expect that any impacts on the prevalence of demand archetypes would vary by household type: those with and without adults in formal work (paid or unpaid), and those with and without children. Furthermore, as schooling and, for many, formal work, occur primarily on weekdays, we

would expect different impacts in these groups on weekdays to weekends.

Here we formalise these expectations into three related testable hypotheses regarding specific changes in the prevalence of different demand archetypes that we might expect due to COVID-19-related restrictions for different segments of the population:

- The sample of households with children and/or at least one adult in work will have seen their weekday demand archetypes become more similar to their weekend demand archetypes during periods of stronger restrictions on schooling and/or work, respectively.
- Households with differing mixes of children and work will have responded differently to the changing restrictions (e.g. households with adults in work but no children will have been less affected by school closures).
- Households with no children and no adults in work will have had more similar demand archetypes on weekdays and weekends throughout, and the difference between weekdays and weekends will have been less affected by restrictions in place for this group.

Fig. 8 presents line plots of the change over time in a ‘weekday-weekend difference score’ for different family statuses, separately for electricity and for gas, to test the above hypotheses. If the hypothesis holds, we would expect this score to fall during periods with lockdowns and restrictions, and to be higher at other times, particularly for

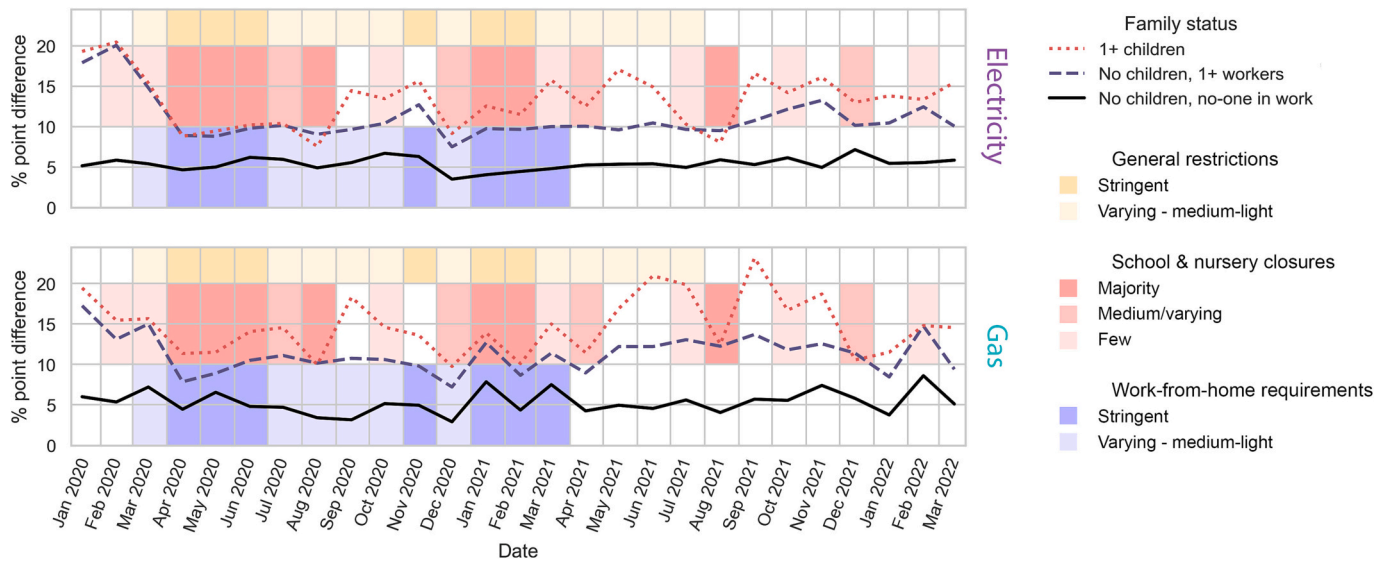


Fig. 8. Line plots of ‘weekday-weekend difference scores’, being the percentage point differences in the prevalence of demand archetypes between weekdays and weekends, by month, for different family statuses (see text for details). Background shading represents the levels of general restrictions, school and nursery closures (including holidays), and work-from-home/furlough requirements in place for each month.

households with children and/or with workers. The score is a derived measure that compares, for each group and for each month, the prevalence of the archetypes on weekdays versus weekends. A lower value indicates that weekdays and weekends are more similar in the spread of archetypes found in that group for that month, with 0 indicating that each archetype occurs equally frequently on weekdays as it does on weekends for that month, while 100 indicates no overlap at all, i.e., archetypes occurring on weekdays are absent on weekends and vice versa during that month for that group. The formula for calculating the weekday-weekend difference score is presented in the methods, [section 4.2.2](#).

The coloured bars in [Fig. 8](#) indicate the average level of restrictions in place month by month, with darker shading representing higher levels of restrictions. Yellows indicate general lockdowns and restrictions, which in the UK included varying restrictions on leaving the home, meeting people from other households, and attending social events and leisure venues (drawn from dates provided in [\[38\]](#)). Reds indicate nursery and school closures, including regular holiday periods as well as closures related to COVID-19 restrictions. Blues indicate work-from-home and furlough requirements. The values are approximate, as rules frequently changed over time and often varied by different regions of England and Wales, as well as by different school year groups and categories of worker.

Although causality cannot be determined with this approach, the results are nonetheless consistent with the hypotheses presented above. Households with no children and no adults in work (solid black line) have consistently more similar weekdays and weekends before and throughout the pandemic than the other groups (indicated by a lower difference score that shows no discernible changes related to the levels of COVID-19 restrictions). In contrast, households with adult workers and/or children show a substantial drop in their difference scores at the start of the first lockdown, for both electricity and gas. For households with children (dotted red line) (90% of which also have at least 1 adult in work in the sample), subsequent change in this weekday-weekend difference score over time is also largely consistent with patterns of school openings and closures in England and Wales, with the score generally falling with greater levels of school closures, including holidays, and rising as schools reopened.

For households with adults in work but no children (dashed blue line), the difference score again drops substantially at the start of the first lockdown, and subsequently stays lower overall and more constant

than for households with children (dotted red line). This pattern is consistent with a substantial level of self-imposed home-working continuing even during periods when it was no longer government-mandated.

6. Discussion and conclusion

This paper set out to investigate the diversity of daily electricity and gas demand profiles from GB households between September 2019 and August 2022, using data from a broadly representative sample of homes. Analyses drew on the SERL Observatory dataset of smart meter and linked contextual data from 13,000 consented households, and applied a methodology based on k-means cluster analysis to identify eight demand ‘archetypes’, i.e. typical daily energy demand profiles. The cluster analytical approach was designed to focus on the shape of the energy demand profiles - the pattern and timing of peaks and troughs, rather than their size in kWh terms, by using normalised energy demand profiles for each day for each home and fuel as input features into the cluster analysis. The characteristics of these archetypes and their prevalence across the sample over the period was analysed.

Four research questions were presented in the introduction, and how the results relate to each of these is discussed below.

- 1. What are the typical energy profiles of the daily demand archetypes identified through the analysis?** Four of the archetypes, which we named ‘Mid morning’, ‘Midday peak’, ‘Late afternoon’ and ‘Evening’, each exhibit a single dominate peak in demand, which varies between them in its timing over the day, from mid morning to evening. One other, ‘Early morning, and evening’, exhibits a more bimodal distribution with an early morning and an evening peak, while ‘All daytime’ shows a morning rise in demand that is then sustained through to the evening. A ‘Flat’ archetype remains low and flat through the day. A final archetype, ‘Midday trough’, is particularly distinctive for electricity, where average demand is negative during the middle of the day and high overnight in the early hours of the morning, which we speculate results from a combination of generation from rooftop solar photovoltaic panels, and overnight charging of electric vehicles and electric storage heaters. Electricity demand is consistently lower than gas demand across all the archetypes, reflecting the primary heating role of gas in most homes. Gas demand exhibits significant seasonal variation across the archetypes,

with low summer demand and substantial winter demand. Electricity demand varies much less with season, which is consistent with electricity use being primarily on non-heating end uses that are less affected by weather. The exception is the 'Midday trough' archetype; the relatively higher winter usage throughout the day, particularly from just after midnight until the evening, is again consistent with overnight charging of storage heating (higher in the winter) and reduced winter electricity generation from solar PV. Future work could investigate this in more detail by analysing patterns for this archetype just for homes in the dataset that have self-reported having these technologies.

2. **How do the observed archetypes relate to the full sample's average demand profile?** The daily demand archetypes identified by the clustering method were used to 'decompose' the average daily electricity and gas demand profiles of the full sample, revealing how the average pattern of peaks and troughs in demand over the day across the sample is built up from a mix of demand archetypes that exhibit a range of substantially different profiles. This result highlights the diversity of demand profiles exhibited by households that is not apparent from average demand data, even when broken down by consumer segments. Some of the results have particular relevance for policy. For example, the 'Midday trough' archetype contributes over a third of nighttime demand despite only 8% of homes exhibiting this profile on average, while for gas, most overnight demand is due to just 4% of homes that are on average exhibiting this same archetype. This suggests that small groups of homes with certain electrical appliances and technologies, and having particular heating patterns (with heating left on overnight as well as during the day) have substantial impacts on average domestic demand patterns over the day. Further research could look into the relationships between technologies, heating behaviours and the demand archetypes.
3. **Does the occurrence of the gas and electricity archetypes relate to each other?** Certain factors such as patterns of occupancy and sleep might be expected to influence the timing of demand for both gas and electricity and so lead to co-occurrence of their demand archetypes. As such, the relationship between gas and electricity archetypes in the sample was investigated using chi-squared tests of independence. However, aside from a strong relationship in the occurrence of the 'Flat' archetype between the two fuels, no other statistically significant relationships were found in either winter (2021–22) or summer (2022). The 'Flat' archetype occurring in homes for both fuels on the same days could indicate times when homes are left unoccupied for the day, although this could not be confirmed with the current dataset.
4. **How prevalent are the demand archetypes across the sample and how does their prevalence vary over time in relation to different time-varying factors?** For electricity, the most common demand archetype on average was found to be 'All daytime' (20% of homes), followed by archetypes with higher energy use in the middle or latter part of the day: 'Midday peak,' 'Late afternoon,' and 'Evening'. 'Midday trough' occurs in 8% of homes on average. Gas usage patterns were found to be quite different, with the 'Early morning and evening' archetype being most common (about 26% of homes on average), followed by 'All daytime' and archetypes with a single peak during the day (all around 12–14% of homes). Variation in the prevalence of the archetypes over time is moderate for the electricity archetypes, but more substantial for the gas archetypes, especially 'All daytime,' with a high standard deviation of over 12 percentage points. Analysis of change over time revealed that several factors could be seen to relate to prevalence. Weekly rhythms (likely related to work and school schedules) were visible as peaks or dips in the prevalence of several archetypes on weekends compared to weekdays, in line with past work that has investigated this [17]. Annual events, notably Christmas Day, could be seen in the data, with variations in archetype prevalence consistent with increased oven use for preparing the traditional midday meal, as well as with people

temporarily leaving their homes unoccupied, as would happen if visiting others over the festival.

In the case of gas archetypes, among gas centrally-heated homes, analysis of their prevalence by external temperature revealed a clear relationship, consistent with heating being used for more of the day at lower temperatures. Furthermore, after controlling for external temperature, distinct weekday and weekend variations in the prevalence of several gas archetypes were identified that are consistent with higher occupancy and hence daytime heating use across the sample on weekends. Conversely, archetypes consistent with heating being used primarily during peak morning and evening times was more prevalent on weekdays, consistent with lower occupancy during the middle of the day. Further research could investigate the relationship between these energy archetypes and patterns of heating use and achieved indoor temperatures, both of which have been found in other research to exhibit similar kinds of diversity in their profiles over the day [25,39]. The relationship between energy demand, heating and achieved temperatures has implications for models of energy demand such as the UK's Standard Assessment Procedure (SAP), a widely used model for assessing the energy efficiency of homes, that assumes standard weekday and weekend patterns of occupancy and heating when estimating a property's energy demand [40].

The evidence from the analyses (for England and Wales) was also found to be consistent with the hypotheses that the COVID-19 pandemic, and associated restrictions on movement, led households to have demand patterns that were more similar on weekdays and weekends than previously. This is consistent with the more similar occupancy patterns between weekdays and weekends that were enforced by lockdowns and other restrictions on leaving the home. The results were consistent with the hypotheses that changes in demand patterns from restrictions were stronger in households with children and those in formal work; the latter appears to have still not returned to pre-pandemic patterns by March 2022, despite work-from-home requirements having been ended around a year prior to that. These results are interesting to compare with previous work on the impacts of the pandemic on total daily demand using the same dataset: Zapata-Webb et al. (2023) [3] found generally higher daily demand for both fuels during lockdowns compared to modelled predictions, returning to predicted levels, or even slightly below them in the case of gas, by Q1 2022. Changes varied by the same family types as used in this research, but there was little detected difference in changes in daily demand on weekdays and weekends.

6.1. Limitations and future work

The current research did include some limitations that future research may be able to address. Firstly, even though the clustering approach helps reveal distinct daily demand archetypes, the k-means method inherently allocates all home-fuel-days in the data to the nearest cluster no matter how different they are. Some previous research has used methods that include a concept of 'noise', which in this case means profiles that are far from any cluster centre, in relatively sparsely populated regions of feature space. Such home-fuel-days are likely to account for some proportion of the variation in prevalence of the archetypes observed between days in the current study; assessing the extent of this would be of value, as it could lead to more robust identification of archetypes and stronger certainty about the relationships between demand archetypes and other factors. Future work could therefore investigate cluster methods that include a concept of 'noise', such as revisiting the hdbSCAN approach tested for this work, or Gaussian Mixture Models, which attribute a probability to each case of its belonging to each cluster.

The current paper uses data up until August 2022, the most recent

available at the time of analysis. The SERL Observatory dataset used in this research continues to be collected and periodically released for use by GB researchers on approved projects [14]. Future work could therefore utilise future releases to continue the observational analysis of change in demand archetype prevalence over time, and to study the ongoing impacts of system-wide changes that are expected to affect domestic energy demand. The impacts of the large increases in energy prices over the winter of 2022–2023, and wider cost of living crisis, would be of particular interest and policy relevance to investigate.

Future work will also use the same dataset to investigate energy demand archetypes at the level of the household, based on their typical demand profiles over extended periods of time. As discussed elsewhere in the paper, household-level analysis has also been the subject of previous research and has its own particular applications, distinct from the day-level analysis presented in this paper.

7. Conclusion

The analytical approach used in this paper has enabled the characterisation of distinct energy demand archetypes for both electricity and gas demand, for a GB-wide sample of over 13,000 homes with data spanning three years. The approach taken maintains comparability between fuels (electricity and gas). Although several papers have analysed demand profiles in similar ways using cluster analysis of smart meter data, to our knowledge no other published research to date has done this for a sample of households in the thousands with data from before, during and after the COVID-19 pandemic, and for both gas and electricity from the same households, nor explored in such detail how the prevalence of demand archetypes varies with different temporally-varying factors.

The analyses have demonstrated that there are detectable relationships between energy demand archetypes for both fuels and multiple factors that vary on different timescales (from weekly to annually). These could be of substantive interest for forecasting demand profiles, for targeting interventions and for informing policymaking.

CRediT authorship contribution statement

Martin Pullinger: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Ellen Zapata-Webborn:** Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis. **Jonathan Kilgour:** Writing – review & editing, Software, Methodology, Formal analysis. **Simon Elam:** Project administration, Funding acquisition. **Jessica Few:** Writing – review & editing. **Nigel Goddard:** Writing – review & editing, Funding acquisition. **Clare Hanmer:**

Writing – review & editing. **Eoghan McKenna:** Writing – review & editing, Conceptualization. **Tadj Oreszczyn:** Writing – review & editing, Funding acquisition. **Lynda Webb:** Writing – review & editing, Funding acquisition.

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 SERL Observatory dataset used in this research continues to be collected and periodically released for use by UK researchers on approved projects [14,18].

Data: Elam, S., Webborn, E., Few, J., McKenna, E., Pullinger, M., Oreszczyn, T., Anderson, B., Ministry of Housing, Communities and Local Government, European Centre for Medium-Range Weather Forecasts, Royal Mail Group Limited. (2023). *Smart Energy Research Lab Observatory Data, 2019–2022: Secure Access*. [data collection]. 5th Edition. UK Data Service. SN: 8666, DOI: <https://doi.org/10.5255/UKDA-SN-8666-6>

Data descriptor: Webborn E, Few J, McKenna E, Elam S, Pullinger M, Anderson B, et al. *The SERL Observatory Dataset: Longitudinal Smart Meter Electricity and Gas Data, Survey, EPC and Climate Data for Over 13,000 GB Households*. *Energies* (Basel) 2021;14. <https://doi.org/10.3390/en14216934>.

Acknowledgements

This article is based on research funded by UKRI, the Engineering and Physical Sciences Research Council project Smart Energy Research Lab, grant number EP/P032761/1. Particular thanks go to the SERL technical team at the UK Data Archive: Darren Bell, Deirdre Lungley, Martin Randall and Jacob Joy for software, data curation and dataset development, and James O'Toole at UCL for project administration. Support from the SERL Advisory Board, Data Governance Board and Research Programme Board played a critical role in the establishment and ethical operation of SERL. We would also like to thank the 13,000+ SERL Observatory households who have consented access to their smart meter data, without whom it would not have been possible to undertake this research.

Neither the European Commission nor the European Centre for Medium-Range Weather Forecasts is responsible for the use in this study of the Copernicus information or data it contains.

Appendix A. Feature values for the daily demand archetypes

Fig. 9 presents the distribution of feature values for the daily energy demand archetypes. As described in the methodology, section 4.1, for each day of data per fuel and per home, the highest 2-h period of energy use was normalised to a value of 1, and the lowest point to 0. The blue lines in Fig. 9 indicate the trend over the day in the mean value for each archetype, while the red shading is a heat map indicating the proportions of the home-fuel-days in each archetype that fall into each banded value for each feature. The 'Flat' profile is not presented; this was procedurally defined, and as it includes home-fuel-days in which there is no variation in demand over the day, it is not possible to calculate feature values for it (the calculation method requires some variation).

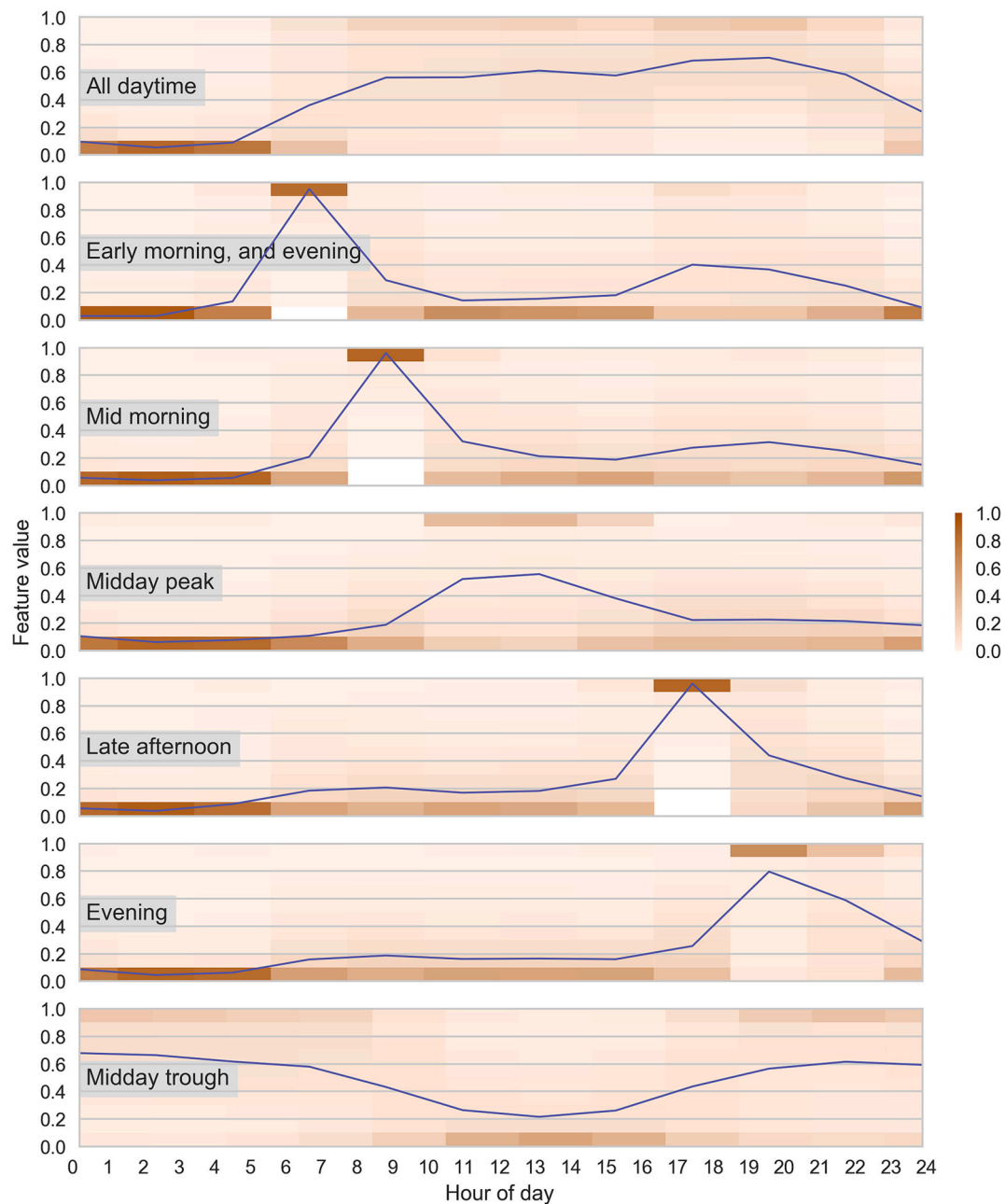


Fig. 9. Average feature values for each daily energy demand archetype, at a two-hour granularity, based on data from 13 January 2019 to 31 August 2022.

References

- [1] Few J, Pullinger M, McKenna E, Elam S, Webborn E, Oreszczyn T. Smart energy research lab: Energy use in GB domestic buildings 2021. In: Smart energy research Lab (SERL) statistical reports. Vol. 1; 2022. London, UK.
- [2] BEIS. Energy follow up survey: Household energy consumption and affordability. Final report. London. 2021.
- [3] Zapata-Webborn E, McKenna E, Pullinger M, Cheshire C, Masters H, Whittaker A, et al. The impact of COVID-19 on household energy consumption in England and Wales from April 2020 to march 2022. *Energ Buildings* 2023;297:113428. <https://doi.org/10.1016/J.ENBUILD.2023.113428>.
- [4] Zapata-Webborn E, Hanmer C, Oreszczyn T, Huebner GM, McKenna EJ, Few J, et al. Winter demand falls as fuel bills rise: understanding the energy impacts of the cost-of-living crisis on British households. *OSF Preprints* 2023. <https://doi.org/10.31219/osf.io/83a4t>.
- [5] Huebner GM, Watson NE, Direk K, McKenna E, Webborn E, Hollick F, et al. Survey study on energy use in UK homes during Covid-19. *Build Cities* 2021;2:952. <https://doi.org/10.5334/bc.162>.
- [6] Huebner G, Hanmer C, Zapata-Webborn E, Pullinger M, McKenna E, Few J, et al. Self-reported energy use behaviour changed significantly during the cost-of-living crisis in winter 2022/23: insights from cross-sectional and longitudinal surveys in Great Britain. *SocArXiv* 2023. <https://doi.org/10.31235/osf.io/984yh>.
- [7] Rajabi A, Eskandari M, Jabbari M, Li L, Zhang J. A comparative study of clustering techniques for electrical load pattern segmentation. *Renew Sustain Energy Rev* 2020;120. <https://doi.org/10.1016/j.rser.2019.109628>.
- [8] Yilmaz S, Chambers J, Patel MK. Comparison of clustering approaches for domestic electricity load profile characterisation- for demand side management. *Energy* 2019;180:665–77. <https://doi.org/10.1016/j.energy.2019.05.124>.
- [9] Li H, Wang Z, Hong T, Parker A, Neukomm M. Characterizing patterns and variability of building electric load profiles in time and frequency domains. *Appl Energy* 2021;291:116721. <https://doi.org/10.1016/j.apenergy.2021.116721>.
- [10] Räsänen T, Voukantis D, Niska H, Karatzas K, Kolehmainen M. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Appl Energy* 2010;87:3538–45. <https://doi.org/10.1016/j.apenergy.2010.05.015>.

- [11] Fernandes M, Viegas J, Vieira S, Sousa J. Segmentation of residential gas consumers using clustering analysis. *Energies* (Basel) 2017;10:2047. <https://doi.org/10.3390/en10122047>.
- [12] Smart Energy Research Lab. Welcome to the smart energy research Lab n.d.
- [13] Webborn E, Elam S, McKenna E, Oreszczyn T. Utilising smart meter data for research and innovation in the UK. In: *Proceedings of ECEEE 2019 Summer Study on energy efficiency*. Belambra Presqu'île de Giens, France: ECEEE; 2019. p. 1387–96.
- [14] Webborn E, Few J, McKenna E, Elam S, Pullinger M, Anderson B, et al. The SERL Observatory dataset: longitudinal smart meter electricity and gas data, survey, EPC and climate data for over 13,000 GB households. *Energies* (Basel) 2021;14. <https://doi.org/10.3390/en14216934>.
- [15] Satre-Meloy A, Diakonova M, Grünwald P. Cluster analysis and prediction of residential peak demand profiles using occupant activity data. *Appl Energy* 2020; 260:114246.
- [16] Alonso AM, Martín E, Mateo A, Nogales FJ, Ruiz C, Veiga A. Clustering electricity consumers: challenges and applications for operating smart grids. *IEEE Power Energy Mag* 2022;20:54–63. <https://doi.org/10.1109/MPE.2022.3150826>.
- [17] Czetányi L, Vámos V, Horváth M, Szalay Z, Mota-Babiloni A, Deme-Bélafi Z, et al. Development of electricity consumption profiles of residential buildings based on smart meter data clustering. *Energ Buildings* 2021;252:111376. <https://doi.org/10.1016/j.enbuild.2021.111376>.
- [18] Elam S, Webborn E, Few J, McKenna E, Pullinger M, Oreszczyn T, et al. Smart Energy Research Lab Observatory data, 2019–2022: Secure access. 6th ed. UK Data Service; 2023. <https://doi.org/10.5255/UKDA-SN-8666-6>.
- [19] ECMWF. ERA5: Data Documentation. ECMWF; 2021.
- [20] Webborn E, Elam S, McKenna E, Oreszczyn T. Utilising smart meter data for research and innovation in the UK. In: *Proceedings of ECEEE 2019 Summer Study on Energy Efficiency*. Belambra Presqu'île de Giens, France: ECEEE; 2019. p. 1387–96.
- [21] Webborn E, McKenna EJ, Elam S, Anderson B, Cooper A, Oreszczyn T. Increasing response rates and reducing bias: Learnings from the Smart Energy Research Lab pilot study n.d. doi:10.31219/OSF.IO/F82B7.
- [22] Elam S, Webborn E, McKenna E, Oreszczyn T, Anderson B, Ministry of Housing Communities & Local Government, et al. Smart Energy Research Lab Observatory Data, 2019–2022: Secure Access. UK Data Service. 6th Editio 2022. <https://doi.org/10.5255/UKDA-SN-8666-1>.
- [23] McKenna E, Few J, Webborn E, Anderson B, Elam S, Shipworth D, et al. Explaining daily energy demand in British housing using linked smart meter and socio-technical data in a bottom-up statistical model. *Energ Buildings* 2022;258:111845. <https://doi.org/10.1016/j.enbuild.2022.111845>.
- [24] Few J, Manouseli D, McKenna E, Pullinger M, Zapata-Webborn E, Elam S, et al. The over-prediction of energy use by EPCs in Great Britain: a comparison of EPC-modelled and metered primary energy use intensity. *Energ Buildings* 2023;288: 113024. <https://doi.org/10.1016/j.enbuild.2023.113024>.
- [25] Pullinger M, Berliner N, Goddard N, Shipworth D. Domestic heating behaviour and room temperatures: empirical evidence from Scottish homes. *Energ Buildings* 2022;254:111509. <https://doi.org/10.1016/j.enbuild.2021.111509>.
- [26] Trotta G. An empirical analysis of domestic electricity load profiles: who consumes how much and when? *Appl Energy* 2020;275:115399. <https://doi.org/10.1016/j.apenergy.2020.115399>.
- [27] scikit-learn developers. Clustering 2023. <https://scikit-learn.org/stable/modules/clustering.html#k-means> (accessed June 10, 2023).
- [28] Arthur D, Vassilvitskii S. K-Means++: The advantages of careful seeding. In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, USA: Society for Industrial and Applied Mathematics; 2007. p. 1027–35.
- [29] scikit-learn developers. sklearn.cluster.KMeans. 2023. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html> (accessed June 10, 2023).
- [30] Van Rossum G, Drake FL. Python 3 reference manual. Scotts Valley, CA: CreateSpace; 2009.
- [31] McKinney W. Data structures for statistical computing in Python. In: van der Walt S, Millman J, editors. *Proceedings of the 9th Python in Science Conference*; 2010. p. 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>.
- [32] Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature* 2020;585:357–62. <https://doi.org/10.1038/s41586-020-2649-2>.
- [33] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12:2825–30.
- [34] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPi 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17:161–272. <https://doi.org/10.1038/s41592-019-0686-2>.
- [35] Kluyver T, Ragan-Kelley B, Perez F, Granger B, Bussonnier MA, Frederic J, et al. Jupyter notebooks - a publishing format for reproducible computational workflows. In: Schmidt B, Loizides F, editors. *Positioning and power in academic publishing: Players, Agents and Agendas*; 2016. p. 87–90.
- [36] Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng* 2007;9:90–5.
- [37] Waskom ML. Seaborn: statistical data visualization. *J Open Source Softw* 2021;6: 3021. <https://doi.org/10.21105/joss.03021>.
- [38] Zapata-Webborn E, McKenna E, Pullinger M, Cheshire C, Masters H, Whittaker A, et al. The impact of COVID-19 on household energy consumption in England and Wales from April 2020 – March 2022. *Forthcoming*. 2023.
- [39] Huebner GM, McMichael M, Shipworth D, Shipworth M, Durand-Daubin M, Summerfield AJ. The shape of warmth: temperature profiles in living rooms. *Build Res Inf* 2015;43:185–96. <https://doi.org/10.1080/09613218.2014.922339>.
- [40] Henderson J, Hart J. BREDEM 2012: A technical description of the BRE Domestic Energy Model, Version 1.1. Watford, UK: BRE. 2015.