Energy Efficiency (2024) 17:86
https://doi.org/10.1007/s12053-024-10264-3

ORIGINAL ARTICLE

Physical, socio-psychological, and behavioural determinants of household energy consumption in the UK

Aven Satre-Meloy · Sam Hampton

Received: 18 March 2024 / Accepted: 22 September 2024

Abstract Determining which attitudes and behaviours predict household energy consumption can help accelerate the low-carbon energy transition. Conventional approaches in this domain are limited, often relying on survey methods that produce data on individuals' motivations and self-reported activities without pairing these with actual energy consumption records, which are particularly hard to collect for large, nationally representative samples. This challenge precludes the development of empirical evidence on which attitudes and behaviours influence patterns of energy consumption, thus limiting the extent to which these can inform energy interventions or conservation programs. This study demonstrates a novel methodology for estimating energy consumption in the absence of actual energy records by using a large, publicly available data set of energy consumption in the UK. We develop a predictive model using the Smart Energy Research Laboratory (SERL) data portal (with records from nearly 13,000 UK households) and then use this model to predict energy consumption (both electric and gas) for a sample of 1,000 UK householders for which we separately collect over 200 variables relating to climate change attitudes and practices. Our approach uses a set of over 50 independent variables that are shared between the data sets, allowing us to train a model on the SERL data and use it to analyse the relationship between energy consumption and the opinions, motivations, and daily practices of survey respondents. Results show that electricity consumption is influenced by a broader range of factors compared to gas. Household energy use is best explained by physical dwelling characteristics, socio-demographic variables, and certain behavioural and attitudinal measures. Notably, pro-environmental attitudes, frugality, and conscientiousness correlate with lower energy use, while income and consumerism are linked to higher consumption. We discuss how these findings can inform efforts to decarbonise home energy use in the UK.

Keywords Behaviour · Energy · Attitudes · Values · Modelling · Survey

Aven Satre-Meloy and Sam Hampton shared first authorship.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12053-024-10264-3.

A. Satre-Meloy
Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
e-mail: asatremeloy@lbl.gov

S. Hampton (✉)
University of Bath, Claverton Down, Bath BA2 7AY, UK
e-mail: sah53@bath.ac.uk

S. Hampton
University of Oxford, 3 S Parks Rd, Oxford OX1 3QY, UK

Published online: 15 October 2024

Introduction

In order to meet climate goals, it is imperative that emissions from household energy use are reduced. In the EU, domestic consumption accounts for 27% of final energy demand (Eurostat, 2023), and 26% of all greenhouse gas emissions in the UK are from households (ONS, 2023a). Promisingly, there is significant potential for emissions reductions from social and behavioural changes in the domestic setting (Dubois et al., 2019). In the US it is estimated that 20% reductions can be achieved through behavioural changes without sacrificing wellbeing (Dietz et al., 2009). Understanding the determinants of energy consumption (and, especially, the primary social and behavioural determinants that might be realised more easily and at lower cost than infrastructural or technological changes) is therefore essential to help accelerate the low-carbon energy transition.

While physical characteristics of buildings such as dwelling type, floor area and heating system type are the strongest predictors of energy use and have been shown to explain a large portion of observed variance in statistical studies of household energy consumption (Huebner et al., 2015; Satre-Meloy, 2019), occupant-related variables that represent social, psychological, and behavioural factors have also been shown to explain variability between houses of similar construction (Frederiks et al., 2015; Gram-Hanssen, 2013).

Unfortunately, evidence on social, psychological, and behavioural determinants of household energy consumption is limited. Research relies on survey methods that produce data on individuals' motivations and self-reported activities without pairing these with actual energy consumption records, which are particularly hard to collect for large, nationally representative samples. This is because when answering surveys, householders are typically poor at estimating their energy consumption (Attari et al., 2010; Marghetis et al., 2019). As such, most household social and behavioural surveys simply avoid direct questions about energy consumption. Further, energy records are otherwise hard to access directly from suppliers due to data protection laws.

This study aims to address this research gap by seeking to quantify the physical, social, psychological, and behavioural determinants of energy consumption in UK households. The UK represents a valuable case study due to its mixed housing stock (MHCLG, 2023), relatively advanced smart-meter roll-out programme (BEIS, 2023), and the socio-economic diversity of its population (DWP, 2023). Average household energy consumption (around 39 kWh/day) is close to the European average (BEIS, 2020). We address this gap by developing a novel methodology which combines two large datasets from the UK. The first is obtained from the Smart Energy Research Laboratory (SERL)[1], which includes energy consumption data (electricity and natural gas) from smart meters installed in nearly 13,000 UK households (Webborn et al., 2021). The second is a large survey of 1,001 UK householders with a total of 345 questions, spanning socio-demographic variables, knowledge, attitudes, and behaviours relating to climate change.

We use a specific class of regression models — those from the sub-field of statistical learning with sparsity — to develop predictive models trained on SERL energy consumption data (electricity and gas), which we then use to predict consumption for the 1,001 households that responded to the separate, detailed attitudinal and behavioural survey. Our approach relies on the fact that these datasets share many independent variables, allowing us to train a model on the SERL data, predict energy consumption for respondents who completed an entirely separate survey, and use the resulting predicted energy values to analyse the relationship between energy consumption and the opinions, motivations, and behaviours of our survey respondents.

The remainder of the article is structured as follows: Section 2 reviews approaches for modelling household energy consumption, evidence on the dwelling and socio-demographic predictors of energy use in households, and the literature from environmental psychology disciplines on attitudes and behaviours related to household energy consumption. Section 3 outlines the study's methodology, including data collection procedures and analysis methods. In section 4, we report model results for the various phases of our analysis, and in section 5, we discuss the implications for future research as well as domestic energy and climate policy.

[1] Available at: https://serl.ac.uk/.

Literature review

Modelling household energy usage using statistical methods

There are many different approaches from the literature on quantitative analyses of household energy consumption. We recommend Foucquier et al. (2013) for an overview of the primary physical, statistical, and hybrid modelling approaches that researchers have used to predict consumption. This paper focuses on the sub-field of statistical modelling and, specifically, regression analysis with sparsity (Hastie et al., 2015), to analyse patterns in household energy demand.

Fumo and Rafe Biswas (2015) provide a comprehensive overview of the various regression approaches that have been applied in previous studies of household energy consumption. The benefits of these methods include their simplicity (in terms of both understanding the outputs and application to numerous types of household and energy data), their flexibility (i.e., they can be adapted to many different contexts and can handle many different types of data), and their computational and data efficiency (in comparison to, for example, machine learning approaches that typically require large data sets and more computational time to train).

Regression approaches also present numerous disadvantages, however. Most traditional regression techniques are not appropriate to handle data contexts in which there are a large number of predictive variables that could influence the response. In the case of household energy consumption, this is particularly relevant given that the number of factors that could influence consumption are nearly limitless (Hsu, 2015). Another disadvantage is the challenge of collecting data on these factors — regression approaches typically rely on either existing data bases of energy consumption and household variables or specific data sets collected via surveys or other similar instruments. In both cases, lack of data availability can present challenges for researchers, and data can also be costly to access or collect. The result is often data sets that are small or specific to particular geographic or economic contexts, which limits their applicability to wider populations of interest.

One related challenge for regression techniques is that they tend to have poor predictive accuracy when applied to samples for which they were not originally developed (e.g., if a regression model is trained on a data set of household energy consumption and physical dwelling variables, it may not generalise well to new data sets in terms of its predictive accuracy). This is especially the case when the number of potential predictive variables is large, in which cases typical regression models often "overfit" the data, meaning they are too tuned to the data set on which they are initially trained (Lever et al., 2016). It is also the case that data sets with a large number of predictive variables suffer from cases of "multicollinearity", which occur when multiple predictors have large degrees of pairwise correlation (an example being dwelling size and income) (Farrar & Glauber, 1967). The presence of multicollinearity can cause model standard errors and coefficients to be inflated and lead to misinterpretation (Daoud, 2017).

There is a class of regression techniques that were developed, in part, to handle many of these challenges. These are a class of regression methods from the field of statistical learning with sparsity (Hastie et al., 2015). Also called "regularisation" methods, these approaches exploit the concept of "sparsity", which refers to models that have a small number of variables with non-zero coefficients. These models are much easier to estimate and interpret than dense models with many non-zero coefficients. Hsu (2015) and Satre-Meloy (2019) provide helpful summaries of the application of regularisation methods to regression models of household energy consumption and test a wide variety of approaches in terms of their accuracy in prediction tasks as well as their ability to identify key determinants of household energy consumption.

The ability to predict with high accuracy while also yielding sparse models that are straightforward to interpret is perhaps the clearest advantage of regularisation methods in the context of statistical modelling of household energy consumption. With traditional methods, there is a trade-off between these objectives, and one must determine prior to analysis whether the primary aim is to predict consumption well or yield interpretable models that have broader insights for researchers and/or policymakers. Shmueli (2010) discusses this distinction in depth and comments on practical considerations when undertaking statistical modelling.

In this paper, as will be discussed in detail in the Methods section, our aim is twofold: we both aim to develop models with high predictive accuracy that generalise well to out-of-sample data and can be used to reliably estimate energy consumption for a separate data set, and we aim to use statistical modelling to assess the relationships between physical dwelling and occupant social and behavioural factors and energy consumption. As such, regularisation methods are particularly apt to our modelling context.

Physical dwelling and socio-demographic determinants of household energy consumption

Physical dwelling characteristics, building appliances/equipment, and occupant socio-economic factors have been studied thoroughly in the household energy consumption literature. These factors tend to account for a large portion of observed variance in statistical models of household energy consumption (Huebner et al., 2016; Kavousian et al., 2013; Satre-Meloy, 2019; Satre-Meloy et al., 2020). Dwelling characteristics that have been studied the most frequently include dwelling floor area (or number of rooms/bedrooms as as a proxy used in occupant surveys), dwelling type (e.g., detached vs. attached), year of construction, presence or absence of different heating and cooling equipment, and thermal insulation properties of the dwelling. Of these, dwelling size and type tend to show the strongest associations with energy consumption (a significant, positive effect for larger dwellings and detached homes vs. semi-detached or attached homes) (Jones et al., 2015).

For models of household electricity usage, appliance-related variables are also shown to account for a large percentage of observed variance in consumption (Huebner et al., 2016). Although these variables are distinct from "dwelling" variables that do not directly involve occupant behaviour, previous studies have found that the number of major appliances present in the household has a strong, positive association with household electricity usage. More specifically, number of TVs, computers, refrigerators, and specific electric cooking appliances (e.g., electric ovens) all consistently demonstrate positive associations (Jones et al., 2015).

Socio-demographic variables also demonstrate strong associations with household energy consumption, though the percentage of observed variance these variables typically account for is somewhat less than physical dwelling and appliance variables (Huebner et al., 2016; Jones et al., 2015). Socio-demographic variables that have consistently shown significant positive effects in previous studies include number and age of occupant(s) (often a single representative for the home but in some cases all occupants' ages are collected), household income (and disposable income), and presence of teenagers; presence of young children and elderly occupants, household representative's education level, and tenure type show mixed results in terms of the direction of association, with some studies finding a positive association with energy consumption and others demonstrating a negative association (Jones et al., 2015).

Psychological and social influences on householders' environmental impacts

There is a long history of studies investigating the psychological determinants of householders' environmental impacts, and constructs such as attitudes, values, and perceived self-efficacy have been shown to have significant influence on pro-environmental behaviours such as activism, voting behaviours and food choices (Duncan & Stewart, 2007; Köster, 2009; Larson & Story, 2009; Roser-Renouf et al., 2014). However, evidence for psychological predictors of household energy use is more variable, and a consistent finding is that socio-demographic factors such as income and household size have stronger associations with consumption and conservation behaviours (Becker et al., 1981; Brandon & Lewis, 1999; Frederiks et al., 2015; Gatersleben et al., 2002; Seligman et al., 1979).

Various theories have been used to explain household behaviour, including the Theory of Planned Behaviour (Ajzen, 1991; Vining et al., 2002), theories of moral responsibility (Samuelson & Biek, 1991; Van Raaij & Verhallen, 1983), normative social influence (Midden & Ritsema, 1983), and habit (Huebner et al., 2013; Webb et al., 2022). There are also extensive debates over how to conceptualise and measure thermal comfort, which reveals the range of theoretical perspectives adopted in an extensive academic literature on household energy use (Cole et al., 2008; de Dear & Brager, 1997; Fanger, 1970; Gram-Hanssen, 2010). Perceptions of energy costs have a significant effect on conservation behaviours (Broberg &

Kažukauskas, 2021; Lesic et al., 2018), and there is robust evidence indicating that real-time feedback to householders can lead to energy savings (Abrahamse et al., 2007; Darby, 2006).

Social norms and peer influence are significant psychological factors affecting energy use. Norm-dynamics occur both within the household (Kleinschafer & Morrison, 2014), and in relation to wider society (Schultz et al., 2007). Knowing that energy-saving behaviours are practised by similar social or demographic groups can encourage similar behaviours. Social norms relating to responsibilities to reduce emissions have also been shown to drive conservation behaviour (Horne and Kennedy 2017). Feedback based on social norms is effective intervention, and messages comparing individual households with others of similar size and construction can induce behavioural change via injunctive norms (Mukai et al. 2022).

Carbon capability

The previous two sub-sections demonstrated that household energy consumption can be explained by a combination of structural, social, and psychological factors. Reflecting this, the theoretical framework which guides this study is one of relatively few within energy social science which deliberately integrates structure and agency, giving equal emphasis to psychological constructs as to infrastructures and systems of provision. Carbon capability emphasises the need for householders to make informed judgments about energy (with sufficient knowledge and information), but it also highlights the role of systemic actors such as energy suppliers and market design and the influence of social norms on household behaviour (Wei et al., 2016; Whitmarsh et al., 2011).

In a recent review of literature on individual choice for climate change, we (Hampton & Whitmarsh, 2023) build on the Carbon Capability approach to identify six domains of choice. Four relate to individuals' own environmental impacts: energy, transport, food, and shopping. These reflect the main components of an individual's carbon footprint, and there have been several recent studies which have quantified the impacts of adopting less-impactful behaviours in each of these domains, attracting widespread interest (Ivanova et al., 2020; Wynes & Nicholas, 2017). The remaining two domains relate to public sphere activities—namely, 'influence' and 'citizenship'. Influence relates to the ways in which individuals can impact others in personal and professional networks to raise awareness or concern about climate change or to take action to reduce their own carbon footprints. Activities include talking about climate change to friends and family (Corner & Clarke, 2016), influencing leaders in the workplace (Zibarras & Coan, 2015), or changing organisational norms and practices. Citizenship relates to individuals' political agency and activities include voting behaviours, engaging in community or charitable activities, or campaigning for pro-environmental policy (Curtin et al., 2010). These six distinct domains of Carbon Capability directly inform this study's design, as the next section explains.

Methodology

This section presents an overview of the data collection and analysis approaches used in the study. First, we describe data collection procedures for the national survey that was conducted as part of this study to understand UK householders' knowledge, attitudes toward climate change, as well as their emissions-related behaviours. Next, we describe an approach to estimate energy usage for this survey sample using an existing open-access data repository of electricity and gas consumption for the UK population. Finally, we present our methodology for assessing the relationships between predicted energy consumption and measures of Carbon Capability to explore the social and psychological determinants of energy consumption in the UK.

Data collection

UK household survey on Carbon Capability

To facilitate our study's objective of determining social and psychological determinants of household energy consumption, we designed a survey for UK respondents using the concept of Carbon Capability (Whitmarsh et al., 2011). Survey questions were split into six main sections, corresponding to the six domains described in the previous section. These examined householders' knowledge, opinions, and attitudes towards climate change, as well as energy use, transport, food, spending behaviours, willingness to change, and support with respect to climate policy. The survey also included question sets relating to personality traits using the Big Five Inventory, a 44-item questionnaire (Costa Mastrascusa et al., 2023), and personal values using the 21-item Portrait Values Questionnaire (Bouman et al., 2018).

A third-party survey agency, Dynata, delivered the survey to a nationally representative sample of UK households between March and April 2022. A single household representative was asked to complete the survey on behalf of the household. Given its length, the survey was split into two 20-minute questionnaires. Wave 1 ($n = 2{,}000$), consisting of 174 questions, focused on energy, transport, and policy, and included a range of socio-demographic indicators. Wave 2 ($n = 1{,}004$) included 171 further questions, focusing on food consumption, spending habits, and social capital. In order to develop our predictive model of household energy usage, our survey included 24 questions which used precisely the same wording as the SERL user survey (as explained further in Section 3.2). The Wave 1 sample was representative on the basis of age, gender, educational attainment, and ethnicity, based on UK census data. Around 50% of Wave 1 respondents also completed the Wave 2 survey, issued three weeks later.

Two trial questions were included in Wave 1 and one in Wave 2. Respondents failing either of these were removed ($n = 139$ in Wave 1 and $n = 114$ in Wave 2). Further quality assurance included checks for the number of children per household (>7 excluded), the number of heating systems (>3 excluded), and any "straight line" answers to survey Likert Scale questions. These checks identified three unreliable responses. The final sample size thus consists of 1,001 verified household surveys.

A review of sample socio-demographics compared to the UK population (based on national statistics from (ONS, 2021) is shown in Table 1. Due to the bias introduced by the fact that only 50% of respondents completed both surveys, some demographic variables show under- or over-representation in our final sample compared to the UK population. These differences include a slight under-representation of males and over-representation of females; an under-representation of 18–24 year-olds and an over-representation of those 65+; and an under-representation of individuals with lower levels of educational attainment. Chi-squared tests indicate that differences between observed and

Table 1 Sample socio-demographic statistics and comparisons to UK population

	Demographic variable	CC Survey sample	UK population (ONS, 2021)	Chi-Squared Statistic (χ^2)	Degrees of Freedom (df)	p-value
Sex	Male	46%	50%	0.48	1	0.488
	Female	54%	50%			
Age	18-24	2%	12%	12.20	5	0.032
	25-34	14%	17%			
	35-44	17%	18%			
	45-54	21%	18%			
	55-64	19%	15%			
	65+	27%	20%			
Education	Low	21%	37%	17.68	2	<0.001
	Middle	55%	36%			
	High	24%	27%			
Ethnicity	White or White British	90%	87%	0.79	4	0.94
	Asian or Asian British	6%	7%			
	Black or Black British	2%	3%			
	Mixed	2%	2%			
	Other	1%	1%			

expected values are significant for age and education ($p<0.05$).

Smart Energy Research Laboratory (SERL)

SERL (2022a) is a large, nationally-representative database of high-resolution electricity and gas consumption data for the Great Britain population, collected using domestic smart meters and household surveys. For 2021, data are available for a total of 12,911 households. These data include both daily and half-hourly electricity and gas readings as well as dwelling characteristics and household/occupant variables collected with a 40-question household survey. Variables include size and type of house, number of bedrooms, type of heating system and controls, and number and age of occupants. SERL makes available information on whether households have solar photovoltaics installed at an aggregated level (SERL, 2022b), but the electricity usage data in the regression model described below is net rather than gross electricity usage (2021 daily mean electricity consumption) for homes with PV.

The SERL dataset was cleaned in two stages. In the first stage, erroneous readings were excluded. These were flagged by the research team responsible for maintaining the data (Webborn et al., 2022). After removing errors, there were 9,909 households with valid readings for mean daily electricity use in 2021. Of these, 7,369 (74%) also included gas usage readings. This share is slightly lower than in the general UK population, where 78% of households use gas (BEIS, 2022). This is due to the lower number of gas meter readings collected by SERL in their sample. All zero values for gas were removed from the final dataset. There are no households with zero electricity use.

The second stage of cleaning excluded all observations where survey data was incomplete. Given the large sample size, we used list-wise deletion to remove all observations with any missing survey responses (e.g., even if only one survey question was not completed, then the entire observation was removed from the dataset). This process produced final sample sizes of 6,751 households for the electricity dataset and 5,256 for the gas dataset, which, for the modelling purposes described in the following section, were of sufficient size to justify the list-wise deletion approach.

Modelling household energy consumption using SERL survey variables

As this study aims to examine the social and behavioural determinants of energy consumption for the UK population, it was not sufficient to use the SERL database alone, as the SERL household survey did not include any questions that explored elements of Carbon Capability (as described above). Conversely, it was not possible to directly access or collect detailed energy records for the UK household survey on Carbon Capability (hereafter "CC Survey") due to privacy considerations.

For these reasons, we developed an approach to train a predictive model on the SERL data using variables that were common to the CC Survey data and then use this model on those variables to predict electricity and gas data for the CC Survey sample. This section discusses this approach.

To develop predictive models of daily mean electricity and gas usage for the SERL database, we applied a specific regression technique that is commonly used in statistical learning when there are a large number of predictor variables that could lead to cases where the model would "overfit" the data and lead to less accurate predictions. This technique is called the least absolute shrinkage and selection operator (lasso) (Tibshirani, 1996). Lasso is a member of a class of methods called "regularisation" or "penalised regression" methods that adapt ordinary least squares regression by applying a penalty term to the least squares estimator, constraining the optimization problem in such a way that resulting model coefficients are reduced or shrunk. Lasso is a specific case of penalised regression that constrains coefficients by their l_1 norm. It is given by:

$$\hat{\beta}^{lasso} = argmin_{\beta}\left(RSS(\beta) + \lambda \sum_{j=1}^{p} \left|\beta_j\right|\right) \quad (1)$$

where:

- $RSS(\beta)$ is the conventional ordinary least squares approach to minimise the residual sum-of-squares with respect to β, as in: $RSS(\beta) = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2$,

- $\lambda\sum_{j=1}^{p}\left|\beta_j\right|$is the penalty term that penalises coefficients by their absolute value or $l_1$norm.

The lasso is unique as a regularisation method due to penalising coefficients by their absolute value, which delivers a sparse solution, shrinking many coefficients' values to zero and providing the dual roles of improving model prediction accuracy while also improving its interpretability.

The primary objective in this part of the analysis is to develop an accurate predictive model that can be trained on the SERL data using variables that are also shared in the CC Survey data set. We use lasso given the somewhat large number of predictive variables in the data ($p=53$). We develop separate models for electricity and gas for the SERL samples described above. Mean daily electricity and gas consumption in the SERL data presents a log-normal distribution (see Figures 4 and 5 in the Appendix), so to reduce the heteroskedasticity of regression errors, we log-transform mean daily electricity consumption and mean daily gas consumption prior to applying lasso.

To train and test our lasso model on the SERL data, we use a conventional 80% training/20% test split of the data (separately for each of the electricity and gas data sets). We use the R package *glmnet* to tune the lasso model's hyperparameter λ, which determines the amount of regularisation that is applied. We use k-fold cross-validation, setting $k=10$ folds, and run lasso across a range of λ values. Cross-validation partitions the training data into k subsets, and in each iteration the prediction error (in this case, mean-squared error) is calculated after training the data on $k-1$ subsets and predicting on the held-out set, repeating this procedure until each of the subsets is held out. Prediction errors are averaged across the model runs. We plot in Fig. 1 how cross-validated mean-squared error varies as a function of λ to select the lasso model with the lowest prediction error that we subsequently use on the test data set.

As is typical in penalised regression models, as the value of λ increases, the coefficients are penalised to a greater extent, and mean-squared error typically increases as the number of non-zero coefficients decreases. In other words, increasing the penalty term (by increasing λ) leads to a reduction in the number of non-zero coefficients with, as you progress to very large values for λ, a concomitant increase in model mean-squared error. The tuning procedure described above facilitates identification of the model that produces the minimum mean-squared error (or, in cases where a more parsimonious model is desired, the model that produces a mean-squared error within one standard deviation of the minimum).

We find that models with 27 and 15 predictors for the electricity and gas models, respectively, deliver the models with the minimum predictive error while reducing the number of non-zero predictors considerably (52% for electricity and 73% for gas).

After tuning the lasso models for both electricity and gas usage, we then apply those models to our held-out test set data to compute their mean-squared error, which will serve as a proxy for their performance on out-of-sample or "unseen" data (in this case, the CC Survey data — see Section 3.3). Table 2 shows statistics for both the electricity and gas model's performance on the training and test data. We use root mean-squared error (RMSE) as the accuracy metric given that it is reported in the units of the dependent variable. In the case of both models, however, the dependent variables have been log-transformed, so we present the test set RMSE in terms of both $log(kWh)$ and converted to kWh.

We find that our training and test RMSE values are similar, indicating that the lasso procedure has avoided overfitting the training data and leads to fairly accurate predictions on held-out or unseen data. Given that the model error is presented in units of the dependent variable, we can assess the error in the context of the distribution of electricity and gas consumption for our samples. We find that our selected lasso models yield a test-set RMSE of 1.59 kWh and 1.81 kWh for electricity and gas, respectively, which represent an error that is 14.3% of the mean daily electricity consumption $\left(\underline{x} = 11.1kWh\right)$ and 5.2% of the mean daily gas consumption $\left(\underline{x} = 35kWh\right)$ for the sample of observations in the SERL data.

These low model errors (in relative terms) provide confidence in model performance for the next part of our analysis, which uses the trained model on a separate data set — the CC survey data — in order to predict electricity and gas usage for the purposes of assessing the relationship between CC survey variables related to occupant attitudes/behaviours and energy consumption.

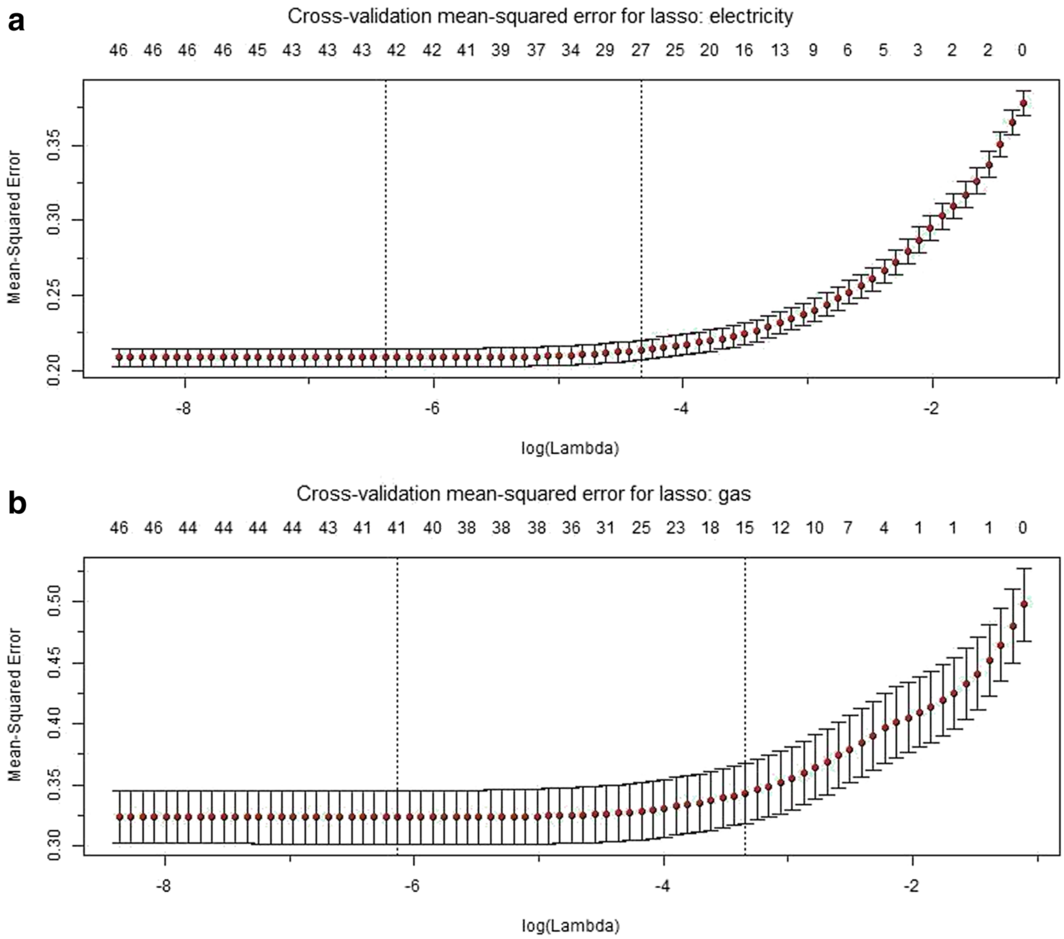

Fig. 1 Lasso regression hyperparameter tuning and model selection for electricity (**a**) and gas (**b**) models, using the SERL dataset. The y-axis shows cross-validation mean-squared error, the bottom x-axis shows the value of log(λ), and the top x-axis shows the number of non-zero coefficients included in the model. Dashed vertical lines denote the model (and value of λ) that produces the minimum mean-squared error (left dashed line) and a mean-squared error within one standard deviation of the minimum (right dashed line)

Table 2 Model performance for electricity and gas lasso models

Model	Training RMSE	Test RMSE	Test R-squared
Mean daily electricity consumption (kWh, log-transformed)	0.464 (units = $log(kWh)$)	0.466 (units = $log(kWh)$) 1.59 (units =kWh)	0.474
Mean daily gas consumption (kWh)	0.591 (units = $log(kWh)$)	0.592 (units = $log(kWh)$) 1.81 (units = kWh)	0.259

Our analysis at this stage is less concerned with interpreting the resulting regression models than it is with making accurate predictions, but given that lasso models also benefit from improved interpretability given their ability to deliver sparse solutions, we can investigate model coefficients as a further check

Table 3 Independent variables and coefficients for predicting energy consumption using SERL's survey (includes variables that are common to both SERL and CC datasets)

Variable name	SERL data lasso model coefficients	
	$b_{electricity}$	b_{gas}
Intercept	2.017	2.02
Heating type: other electric e.g. heat pump	0.427	-0.174
Heating type: electric storage heaters	0.405	0
Heating type: electric radiators	0.235	0
Total adults over 16	0.226	0.052
Number of bedrooms	0.161	0.271
Heating type: other central heating	0.155	0
Total children	0.114	0
Heating controls: smart control	0.056	0
Thermostat temperature settings	0.03	0.037
Adjusting heating: for pets	0.029	0
Tenancy: own home	0.019	0.051
Adjusting heating: for cold weather	0.008	0
Total over 65	0	0.089
Heating type: biomass boiler	0	-0.537
Adjusting heating: for working from home	0	0.014
Self-assessed financial status	-0.005	-0.016
Accommodation type: Semi-detached	-0.006	0
Wear more clothing habit	-0.009	0
Year of home construction	-0.009	-0.04
Gender of respondent	-0.022	0
Adjusting heating: for visitors	-0.024	0
Central heating has temperature setting controls	-0.028	0
Switch off lights habit	-0.041	0
Heating controls: only on/off	-0.041	0
Effort to reduce energy use	-0.048	-0.04
Accommodation type: Terraced	-0.087	-0.057
Accommodation type: block of flats	-0.109	-0.252
Heating type: oil	-0.113	0
Heating type: district heating	-0.24	-0.031
Heating type: gas boiler	-0.433	0.525

on the models' behaviour (in this case, exploring whether model coefficient signs and magnitudes are, for the most part, intuitive).

Table 3 presents the lasso-selected coefficients in each model (the non-zero coefficients in the final electricity and gas models, respectively). As shown in the table, several variables related to heating type have relatively large coefficients in the electricity model, as do variables related to number of occupants, number of bedrooms, and several variables related to thermostat settings. In the gas model, the presence of electric heating in the home is, expectedly, negatively associated with gas consumption; other predictors of large relative magnitude include total number of adults and bedrooms as well as several other demographic variables. These results add further confidence to the model training and prediction procedures described in this section, as the variables selected have mathematical relationships with electricity and gas consumption that have been demonstrated in many previous studies of household energy consumption. We return to a more thorough discussion of variables and their coefficients in the paper's results section.

Applying the model to predict energy consumption for Carbon Capability survey participants

The lasso regression models that were trained on the SERL data as described in Section 3.2 are subsequently used to predict electricity and gas consumption for the respondents to the CC survey ($n = 1{,}001$), for which electricity and gas data are not available, which facilitates analyses of the relationships between social and behavioural variables included in the CC survey that were not included in the SERL survey. The application of SERL data-trained models to CC Survey data is feasible given that both data sets share variables (which are associated with identical survey questions between the two surveys), including the lasso-selected variables shown in Section 3.2.

Both survey samples were collected using representative sampling techniques from the same general population in the UK. Still, given that the surveys were administered separately over different time periods, as well as due to survey-specific sampling nuances, we examine descriptive statistics for key variables that are shared between the data sets in order to determine how the two samples may differ. Table 4 presents comparisons of descriptive statistics for select variables between the two survey data sets as well as with the UK population (based on recent national statistics). Gas usage is predicted only for those CC Survey respondents who indicated that they have a gas boiler in their home or any respondents who indicated they are part of a district heating network.

As shown in Table 4, there are some notable but relatively minor differences in key dwelling and household variables between the SERL and CC Survey data sets, as well as between both samples and

Table 4 Comparisons of descriptive statistics for select overlapping variables between SERL, CC Survey, and UK population data sets. Sources: (ONS, 2021*; BEIS, 2022†; DLUHC, 2022**; Ofgem, 2024††)

	SERL			CC Survey			UK Population
	n	Sample Mean	SD	*n*	Sample Mean	SD	Mean
Number of bedrooms	6751	2.95	0.89	1001	2.69	0.86	2.7*
Number of occupants		2.37	1.19		2.36	1.20	2.4*
Heating fuel type: gas		74%	0.44		83%	0.37	78%†
Tenancy: own home		83%	0.37		68%	0.47	65%**
Accommodation type: block of flats		16%	0.37		10%	0.31	23%**
Mean daily electricity consumption, 2021		11.13	8.30		8.93	4.46	7.40††
Mean daily gas consumption, 2021	5256	35.03	21.68	829	25.03	11.78	31.51††

the UK population. In particular, the SERL data include larger homes (using “Number of bedrooms” as a proxy) than the CC survey (the latter of which is very well matched to the UK population average of 2.7), more homeowners (again, the CC Survey’s average of 68% of the sample closely matches the 65% in the broader UK population), and lower frequency of homes with gas heating. For this variable, the SERL data under-represent and the CC Survey data over-represent homes with gas heating compared to the UK population (SERL = 74%, CC Survey = 83%, and UK population = 78%). Both survey samples under-represent dwellings with the type “block of flats”, with the CC survey showing a lower sample mean (10%) than the SERL data (16%) and the UK population (23%).

These differences in underlying dwelling and household-level variables are important in the context of comparing the measured mean daily electricity and gas consumption between the samples and compared to the UK population. In general, the predicted mean daily electricity consumption for the CC Survey sample is an average of 8.93 kWh, which is around 20% lower than the measured mean daily electricity consumption in the SERL data (11.13 kWh), but about 18% higher than the UK population average in 2021 (Ofgem, 2023). The difference between the SERL and CC survey data means is likely explained by the aforementioned higher frequency of larger, owner-occupied homes in the SERL data as compared to the CC survey (and UK population). It’s possible that the slightly larger frequency of homeowners in the CC survey than the UK population can explain the slightly higher electricity consumption, as home ownership is a variable that has been identified as a positive correlate of electricity consumption, but there are likely many other factors that affect electricity consumption and account for the (slight differences) beyond what is shown in Table 4 (Huebner et al., 2015).

Regarding gas consumption, the table shows larger differences between measured mean daily gas consumption in the SERL data and predicted gas consumption in the CC Survey data. In the case of the SERL data, mean daily gas consumption is higher than the UK population average, whereas it is lower in the CC Survey data. While it is unlikely to explain the entire observed difference between the SERL and CC Survey samples, a review of shared variables’ descriptive statistics show that SERL households tend to have homes with slightly higher thermostat set points and a higher percentage of households with occupants who work from home (which would, in both cases, lead to higher gas consumption). Further, the gas model has a fairly large negative coefficient for the variable “Accommodation type: block of flats”, and the higher frequency of this accommodation type in the CC Survey data as compared to the SERL data is likely leading to the lower predicted mean daily gas consumption in the CC Survey data.

Sample differences and differences in mean daily energy consumption values (both measured and predicted) are to be expected due to underlying differences in the survey samples. This does not invalidate the results of the predictions for the CC Survey data, however; in fact, the explanations given above should give confidence that the predictions are taking into account the differences between the survey samples’ dwelling and household/occupant characteristics — in other words, we should expect differences in predicted and measured electricity and gas consumption between the two. We should also expect some differences between both surveys' energy consumption values and those of the broader UK population. While both surveys were conducted using sophisticated sampling techniques (with the support of independent survey administrators in both cases), we would still expect some differences due to certain factors affecting survey response (including the sociodemographic differences noted in Table 1).

Given our aims of assessing relationships between social and behavioural factors and energy consumption for the CC survey, specifically, it is not necessary to have perfect representation (though the representativeness of the sample provides some important caveats for interpretation of the results and generalisability of our findings) but rather to have predicted consumption values that, where different from the SERL data and/or UK population, can be reasonably explained by differences in the underlying survey sample characteristics.

Finally, comparing sample means is a somewhat crude way to analyse differences, so we produce histograms of the distribution of mean daily electricity and gas consumption for CC survey data sets and present these in Fig. 2 using similar histogram parameters (note: distributions are not available for the UK population). A review of these shows that our predicted consumption estimates and the measured values from the SERL data (see Figures 4 and 5) are fairly well matched across the full distributions. One important difference is the presence of some tail-end

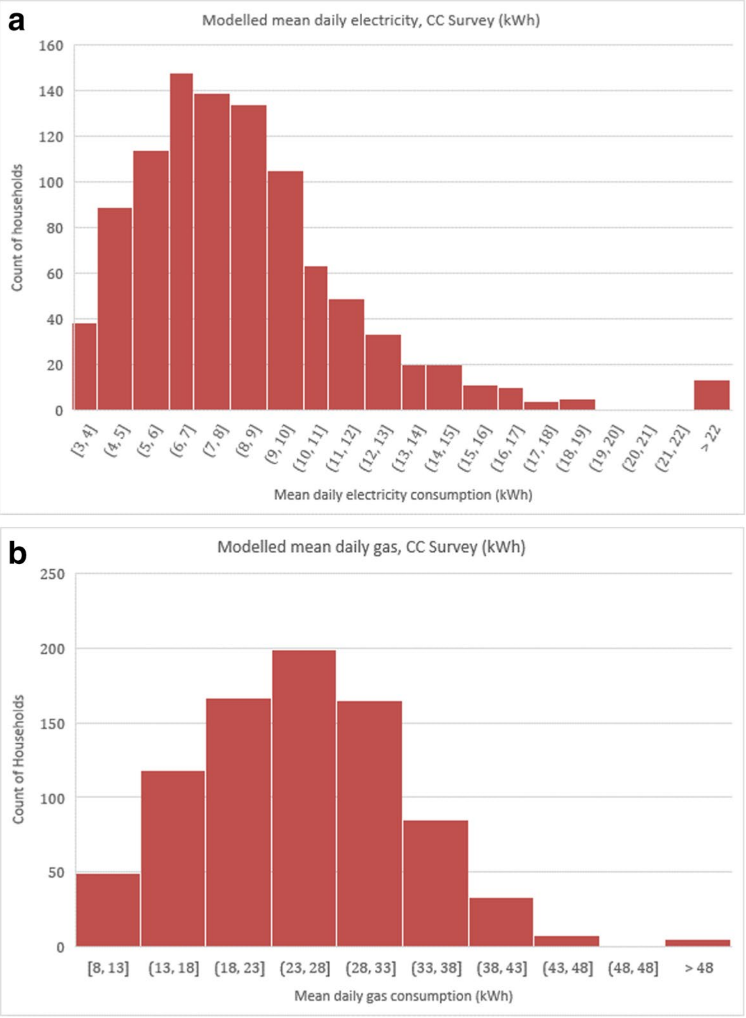

Fig. 2 Histograms of predicted mean daily electricity (**a**) and gas (**b**) consumption for CC survey data

usage households for both electricity and gas in the SERL data (though more notable and extreme in the gas data). This difference is also likely to explain the slightly higher mean gas consumption in the SERL data than predicted in the CC survey. While the presence of outliers in the SERL dataset is noteworthy, we are reassured by the extensive and transparent error checking and data cleaning processes followed by the SERL team (Webborn et al., 2021).

As such, the following section describes the final part of our methodology wherein we apply a second set of regression models to the CC survey to examine associations between dwelling, social, and behavioural variables and predicted electricity and gas consumption.

Assessing relationships between CC survey variables and energy usage

In this section, we describe our analysis to model mean daily electricity and gas consumption for the CC survey sample as a function of dwelling and household characteristics in addition to the many Carbon Capability-related variables that explore occupant attitudes, values, and behaviours in the context of energy consumption and climate change.

To do this, we repeat the methods introduced in Section 3.2, this time applying lasso regression primarily as a means to identify key variables that predict higher or lower electricity and gas consumption. The aim in this part of the analysis is to identify those variables that, in particular, are related to Carbon Capability, but because we need to control for the variables that we already know to be predictive of consumption (i.e., those variables that were selected when training the lasso models on the SERL data), we include those selected variables in addition to all of the other variables included in the CC survey data (namely, those that pertain to domains of Carbon Capability).

As was the case with the SERL dataset, our predicted electricity consumption data for the CC survey present a positive skew and thus are log-transformed prior to running lasso regression. Note that this step was not performed on the gas data given the relatively normal distribution present. All other steps described in Section 3.2 are repeated (e.g., splitting data into training and test sets, *k*-fold cross-validation with 10 folds, hyperparameter tuning of λ across a range of values to identify and select final models, etc). Regarding this last step, our approach differs from that described in Section 3.2 as we select the model with a cross-validation error that is within one standard deviation of the model with the minimum error. This decision is guided by our desire to have a sparser model that aids interpretability given that our primary aim at this stage of the analysis is to interpret resulting model coefficients rather than predict on new data (Friedman et al., 2010).

Finally, after running lasso separately on electricity and gas consumption, we take the additional step of performing ordinary least-squares (OLS) regression with all of the non-zero coefficients in each model. Doing so permits a more straightforward interpretation of the relationships between predictors and the response variable, as the coefficients in the lasso models are penalised and as such are not straightforward to interpret.

For the OLS regression, we compute both standardised and non-standardised model coefficients (Bring, 1994). We do this for several reasons: 1) to compare the magnitude of coefficients on a common scale for predictor variables that have different units, and 2) to provide a more intuitive indication of the coefficients by framing them in terms of the units of the dependent variable (e.g., "percent change in mean daily electricity consumption"). We use standardised coefficients for the former comparison and non-standardised coefficients for the latter.

Non-standardised coefficients represent the change in the dependent variable for a one-unit change in the independent variable, providing a straightforward interpretation of the impact of each predictor on the outcome. Standardised coefficients express the change in the dependent variable in terms of standard deviations, enabling a comparison of the relative importance of predictors regardless of their measurement units. Standardised coefficients are beneficial when comparing the strength of relationships across variables with different scales (i.e., when the variables involved are measured in different units, which many of the variables included in this analysis are — e.g., "Age of home" in years vs. "Household income" in dollars).

Before performing the OLS regression, to check whether the lasso regression methods applied in this part of the analysis appropriately handled the issue of multicollinearity between predictors, we inspect the variance-inflation factors (VIFs), which signal whether regression coefficients are inflated due to correlation between predictor variables. A typical rule of thumb suggests that issues of multicollinearity may be present if any variables show VIFs greater than 10 (Roberts & Thatcher, 2009). The VIFs for the predictors in the electricity OLS model range from 1.2–6.4 with a mean of 2.4 while those for the gas OLS model range from 1.2–6.9 with a mean of 2.5. In both cases, these factors suggest that multicollinearity is not likely to be unduly influencing the magnitude of coefficients in the OLS regression models.

Software

We conduct all of our analyses using RStudio (version 2023.12.0; Rdc, 2010). We use the package *caret* (Kuhn, 2008) for all model training/testing procedures, as well as final model evaluation, and we use *glmnet* (Taylor & Tibshirani, 2015) and its built-in cross-validation capabilities specifically for lasso regression, including hyperparameter tuning and model fitting. We also use *dplyr* (Wickham et al., 2023) for data manipulation and cleaning.

Results and discussion

In this section, we focus on assessing model results for the models described in Section 3.4, which explore relationships between dwelling characteristics, as well as social and behavioural factors, for the CC survey population. We primarily review the strength and direction of associations between selected variables and energy consumption in relative terms, but we also comment on what the model coefficients indicate in real terms (i.e., how to interpret the size of effect in the models in terms of increase or decrease in daily mean electricity or gas consumption).

Lasso regression model results for CC survey

As described in Section 3.4, we apply lasso regression in models of household mean daily electricity ($n = 1{,}001$) and gas ($n = 829$) consumption for our CC survey sample to select influential variables from a set of 229 predictors in the CC survey data set. Our cross-validation procedure yields a range of models shown in Fig. 3 for electricity (a) and gas (b), respectively. These figures show cross-validated model error (vertical axis) for various values of the penalty term λ (shown along the bottom horizontal axis) and resulting models with a discrete number of non-zero coefficients (shown along the top horizontal axis). We select the model with a mean-squared error within one standard deviation of the minimum (the right-hand vertical dashed line on the plot) for each of the electricity and gas cross-validation procedures, separately.

As described in Section 3.4, selecting this model has the benefit of producing a sparser model with fewer non-zero coefficients, which will aid in identifying the most influential variables from our large predictor set (without trading off a large degree of model accuracy). We select models with 40 and 21 non-zero model coefficients for the electricity and gas models, respectively. These include many of the dwelling and occupant characteristics that are shared with the SERL data (and were selected via the first round of lasso regression described in Section 3.2), but they also include some of the social, attitude, and behavioural variables that were the primary focus of the CC survey. The following sections examine the relative influence of both types of variables for predicting mean daily electricity and gas usage.

Physical, socio-psychological, and behavioural determinants of energy consumption

We present all lasso-selected variables for the electricity and gas models along with their OLS coefficients (both standardised and non-standardised) in Tables 5 and 6, respectively. In both tables, we separate these variables based on whether they are specific to the CC survey or are shared across both the CC Survey and SERL data sets, We rank variables by their standardised regression coefficient (β), and we also present non-standardised coefficients in the tables (b) with 5/95% confidence intervals.

As mentioned in Section 3.4, interpretation of regression coefficients varies between the standardised and non-standardised cases. Standardised coefficients can be compared in relative terms given that they represent changes in standard deviations of the dependent variable (mean daily electricity or gas consumption in $log(kWh)$), which is why Tables 5 and 6 rank variables by their standardised coefficient. Non-standardised coefficients, however, are

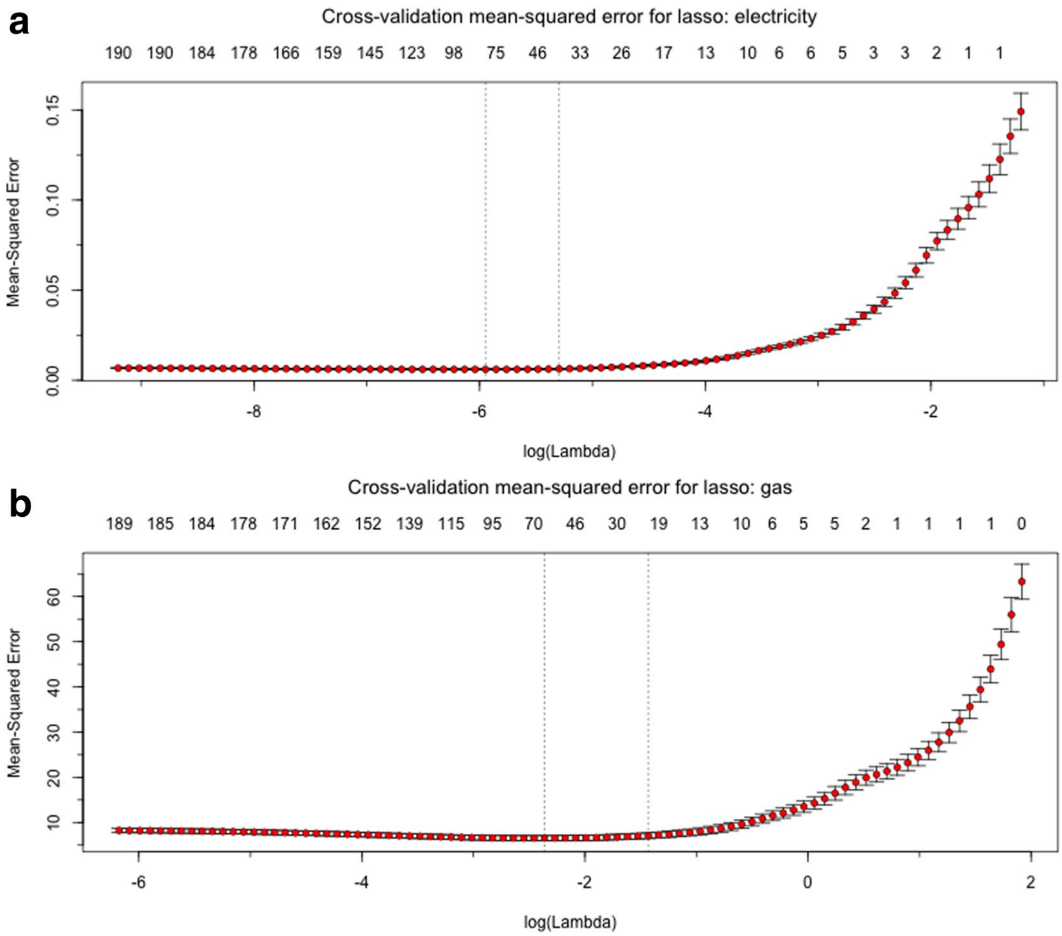

Fig. 3 Lasso regression hyperparameter tuning and model selection for electricity (a) and gas (b) models, using CC Survey data. As in Fig. 3, the y-axis shows cross-validation mean-squared error, the bottom x-axis shows the value of log(λ), and the top x-axis shows the number of non-zero coefficients included in the model. Dashed vertical lines denote the model (and value of λ) that produces the minimum mean-squared error (left dashed line) and a mean-squared error within one standard deviation of the minimum (right dashed line)

measured in the units of the dependent variable. Interpretation in the case of the gas model (Table 6) is straightforward — a one-unit increase in an independent variable is expected to increase or decrease mean daily gas consumption in kWh by the variable's coefficient. But for the mean daily electricity model, because we performed a log-transformation on the dependent variable prior to regression, the coefficients are measured in terms of $log(kWh)$. It is typical in these cases to take the exponent of the coefficients since exponentiation is the inverse of the logarithm. Doing so produces coefficients that represent changes in the ratio of the expected geometric means of the dependent variable. In other words, they represent a multiplicative factor for every one-unit increase in the independent variable — e.g., a coefficient of 1.2 represents a 1.2x increase in the dependent variable (or, more simply, a 20% increase).

The top portion of Tables 5 and 6 show lasso-selected variables that are unique to the CC survey and relate to occupant social, attitude, and behavioural factors. The bottom portions show lasso-selected variables that are shared across the CC and SERL surveys. Comparing the magnitude of standardised coefficients between both sets makes clear that the variables relating to occupant attitudes and behaviours have much lower effect sizes than many of

Table 5 Lasso-selected variables and their OLS standardised and non-standardised coefficients and 5/95% confidence intervals for the mean daily electricity consumption model

Variable name	$\beta_{electricity}$ (standardised)	$b_{electricity}$ (non-standardised, exponentiated) (5/95% CI)
Intercept		5.124 (4.763, 5.513)
Selected variables that are unique to the CC survey data		
The energy efficiency of my home is lower than I would like	0.024	1.006*** (1.003, 1.010)
Disagrees that I have done all I can to reduce footprint	0.012	1.003 (0.999, 1.008)
Pro-consumerism attitude	0.010	1.003 (0.999, 1.008)
Monthly spending on health and beauty products	0.009	1.005 (0.997, 1.013)
Monthly spending on clothes and footwear	0.008	1.005 (0.997, 1.014)
Household income	0.008	1.001 (0.999, 1.000)
Percentage of disposable income spent on clothes, cosmetics, electronics, or sporting equipment	0.006	1.000 (0.999, 1.000)
Close contacts do not buy things unnecessarily	-0.002	1.000 (0.996, 1.003)
Frequency of buying products with less packaging	-0.003	0.999 (0.996, 1.002)
Thinks carbon footprint is lower than average	-0.004	0.999 (0.993, 1.003)
Big Five Inventory (BFI) Conscientiousness scale score	-0.008	0.997 (0.991, 1.003)
Willingness to reduce car travel	-0.008	0.998 (0.995, 1.001)
Socially distrusting indicator	-0.011	0.991 (0.982, 1.001)
Others in home are unwilling to save energy	-0.012	0.997 (0.994, 1.000)
Responded to April 2022 energy price increases by conserving energy	-0.013	0.990 (0.980, 1.000)
Portrait Value Questionnaire (PVQ) Power scale score	-0.014	0.994 (0.988, 1.000)
I always think about my environmental impact	-0.014	0.996 (0.994, 1.003)
Agrees that it is unacceptable to heat an empty home	-0.015	0.996* (0.993, 0.999)
Freedom of purchasing behaviour	-0.018	0.994** (0.990, 0.998)
Has smart speaker	-0.018	0.986** (0.977, 0.996)
Frequency of reduce, reuse, recycle behaviour	-0.020	0.993** (0.988, 0.998)
Use of energy restricted by cost	-0.027	0.994 (0.990, 0.997)
Home is colder than my friend's or family member's	-0.029	0.993*** (0.990, 0.997)
Parent of child under 15	-0.163	0.862*** (0.850, 0.875)
Variables that are shared between the CC survey and SERL data		
Number of occupants	0.669	1.235*** (1.229, 1.242)
Number of bedrooms	0.364	1.175*** (1.167, 1.183)
Heating type: electric radiators	0.156	1.244*** (1.221, 1.267)
Heating type: other electric heat pump	0.128	1.494*** (1.437, 1.553)
Thermostat temperature setting on cold days	0.105	1.034** (1.030, 1.038)
Gender of respondent (Male = 1)	0.032	1.025*** (1.015, 1.035)
Adjust heating: for home working	0.013	1.016* (1.001, 1.030)
Heating type: other central heating	0.007	1.040 (0.972, 1.112)
Adjust heating: for children	-0.012	0.986 (0.971, 1.002)
Adjust heating: for visitors	-0.016	0.986* (0.975, 0.997)
Central heating has temperature setting controls	-0.018	0.985** (0.975, 0.996)
Year of home construction	-0.033	0.992*** (0.989, 0.995)
Heating type: oil	-0.066	0.882*** (0.858, 0.906)
Accommodation type: Terraced	-0.100	0.909*** (0.898, 0.920)
Accommodation type: block of flats	-0.120	0.884*** (0.871, 0.898)
Heating type: gas boiler	-0.417	0.655*** (0.644, 0.665)

*$p<0.05$, **$p<0.01$, ***$p<0.001$.

Table 6 Lasso-selected variables and their OLS standardised and non-standardised coefficients and 5/95% confidence intervals for the mean daily gas consumption model

Variable name	β_{gas} (standardised)	β_{gas} (non-standardised) (5/95% CI)
Intercept		3.388 (1.410, 5.366)
Selected variables that are unique to the CC survey data		
Financially living comfortably or doing alright	0.028	0.445* (0.054, 0.835)
Household income	0.024	0.089 (-0.010, 0.189)
Votes for Green Party	0.014	0.537 (-0.283, 1.357)
Willingness to reduce car travel	-0.016	-0.078 (-0.187, 0.032)
Home is colder than my friend's or family member's	-0.019	-0.091 (-0.203, 0.022)
Use of energy restricted by cost	-0.020	-0.093 (-0.209, 0.023)
Socially distrusting indicator	-0.021	-0.335 (-0.684, 0.013)
Would give money to a local project	-0.024	-0.387* (-0.743, -0.031)
Freedom of purchasing behaviour	-0.027	-0.189* (-0.340, -0.038)
Parent of child under 15	-0.142	-2.630*** (-3.131, -2.128)
Variables that are shared between the CC survey and SERL data		
Number of bedrooms	0.685	6.563*** (6.305, 6.821)
Number of occupants	0.211	1.366*** (1.176, 1.556)
Age of respondent	0.164	0.879*** (0.727, 1.031)
Thermostat temperature setting on cold days	0.163	1.082*** (0.934, 1.229)
Adjust heating: for home working	0.041	0.981*** (0.454, 1.507)
Not working, student, retired, or not seeking employment	0.033	0.583** (0.144, 1.021)
Living rent-free	0.031	1.574** (0.446, 2.703)
Does not know if central heating has temperature setting controls	-0.039	-2.365*** (-3.665, -1.064)
Accommodation type: Terraced	-0.112	-2.138*** (-2.563, -1.714)
Accommodation type: Block of flats	-0.185	-4.396*** (-4.973, -3.819)
Year of home construction	-0.205	-0.992*** (-1.098, -0.887)

*p<0.05, **p<0.01, ***p<0.001

the dwelling or socio-demographic variables selected. This is not a surprising result given that the latter types of factors have consistently shown the strongest statistical associations with household energy consumption in the literature (McLoughlin et al., 2012; Wahlström & Hårsman, 2015; Huebner et al., 2015; Satre-Meloy, 2019). It is likely the case that the presence of the dwelling variables and their large effect sizes in the model is dominating the effect size of the attitude and behaviour variables, but it is important to control for the effects of dwelling size, number of occupants, heating type, and other key variables in these models in order to quantify the average effect of a given variable (in this case, our focus is on the attitude and behaviour variables) when holding all other variables in the models constant. Thus, excluding dwelling variables from the models would likely inflate or otherwise corrupt model coefficients and lead to misinterpretation of results. Still, it is a clear finding in our analysis that attitude and behavioural factors appear to play a far less important role in predicting mean daily electricity and gas consumption than do more structural, dwelling-related variables. These latter variables show greater statistical significance in the models, as well.

In UK households on the gas grid, gas is typically used for heating, hot water, and some cooking appliances. In contrast, the energy services provided by electricity are relatively more diverse, including lighting, cooking, entertainment, and, in homes without gas, heating and hot water supply. As such, we see a larger number and greater diversity of variables selected in the electricity model than in the gas model. The gas model includes 21 variables, of which 12 overlap with the model from the first part of our analysis (trained on the SERL data).

For both electricity and gas, dwelling size (measured by the variable "Number of bedrooms") and number of occupants show large, positive associations with electricity consumption. Type of home shows a strong, negative association with electricity consumption (i.e., terraced homes and flats/tenements are predicted to have, on average, lower energy consumption than detached homes). However, tenancy type (e.g. home-ownership or renting) has less influence on energy consumption. "Living rent-free" was selected with a small effect size ($\beta = 0.03$), corroborating weak and mixed results in the literature (Ndiaye & Gabriel, 2011; Kavousian et al., 2013).

In EU households, space heating and water heating represent over 64% and 14.5% of total energy use, respectively (Eurostat, 2023). It is therefore unsurprising that heating type is a strong predictor of electricity and gas use in our analysis. Homes with electric heating (either resistive technologies or heat pumps) use more electricity, and ownership of a gas or oil boiler used for heating and hot water correspondingly predicts lower electricity use. Thermostat settings are also correlated with energy use: for every increase by 1°C, our findings indicate that electricity consumption increases on average by 3.4% and gas by 8.2%. The importance of building energy efficiency is also an expected finding. In the UK, where the housing stock is typically older and less efficient than elsewhere in Europe (IEA, 2022), housing age is strongly correlated with energy efficiency (ONS, 2023b). In our models, newer homes use less gas and electricity.

The CC survey includes a large number of attitudinal and behavioural variables, as well as additional sociodemographic factors. 25 of these are selected by the electricity model, and nine by the gas model, although the standardised coefficients and statistical significance levels tend to be smaller than for those which relate to building and household characteristics. This corroborates previous studies which find substantial diversity in relation to energy behaviours in the residential sector (Bartiaux & Gram-Hanssen, 2005; Vringer et al., 2007; Satre-Meloy, 2019). Variables that show the strongest positive associations primarily relate to whether occupants could do more to conserve energy, as well as attitudes toward consumerism and purchasing behaviour. The variable showing the largest negative coefficient is parenting. When controlling for the number of occupants, our models indicate that being a parent of a child under 15 is associated with 14% less electricity consumption, and 10% less daily gas use. This indicates that for households with an equivalent number of occupants, the relative contribution that children make towards energy consumption is lower than for adults—an intuitive result but not one with much practical significance.

Our evidence indicates that householders with more frugal attitudes use less electricity. Those who self-report having colder homes compared with those whose friends and family consume less, and so do those who say they frequently reduce, reuse, and recycle items. Similarly, householders who regularly think about their environmental impact are found to use less electricity on average. Those identified as being 'Conscientious' using the Big Five Inventory scale (Costa Mastrascusa et al., 2023), those with personal values relating to power (Bouman et al., 2018), and those who say they have "freedom of purchasing behaviour" are likely to use less electricity and gas. These findings support the observation that frugal attitudes, combined with perceived self-efficacy, support energy conservation.

Our modelling indicates that sensitivity to price is a predictor of energy use. Those who agreed that their energy use was "restricted" by its cost used around 1% less electricity and gas than those who did not agree. Additionally, our CC survey was conducted in April 2022, shortly after a substantial increase in the energy price cap was implemented across the UK. Those who said they had responded by conserving energy are estimated in our model to use 1% less electricity than those who did not. Income is also selected as a positive predictor in both models, and electricity use is correlated with monthly spending on health products and clothes/footwear.

Lastly, a surprising finding is that people who vote for the UK Green Party are predicted to consume 0.5kWh of gas per day *more* than those who vote for other parties. Other surprising findings include that those more distrusting of others consume less energy; and households which have smart speakers also use less electricity.

Conclusions

Summary of findings

This study addresses a key research gap relating to the social, psychological and behavioural determinants of energy consumption in households.

Developing a novel methodology for combining two large datasets from UK households, we have identified the key predictors of energy use and quantified their effects using regularisation techniques well suited to analysis of large numbers of independent variables. Using lasso regression based on a large sample of smart meter data ($n=6{,}751$ for electricity; $n=5{,}256$ for gas), we constructed a predictive model with relatively low prediction error — 14.3% of the mean daily electricity consumption $\left(\underline{x} = 11.1kWh\right)$ and 5.2% of the mean daily gas consumption $\left(\underline{x} = 35kWh\right)$.

We believe this is the first study to use regularisation to predict energy consumption values for a large-scale social and behavioural survey using an existing, representative smart-meter data set. Using the model trained on smart meter data, we then conducted a further lasso regression, using 200 independent variables from our CC Survey.

Despite using the same set of independent variables in lasso regression models for both electricity and gas, our electricity model selects a larger number of predictors (40) than the gas model (21). This supports the observation that there is a greater range and variety of factors that explain electricity consumption than gas consumption. As mentioned previously, this is likely due to the fact that the energy services provided by gas in UK households is limited to heating, water heating, and cooking whereas the services provided by electricity are far more numerous and varied.

Our results show that household electricity and gas use in the UK is best explained by a mix of physical dwelling characteristics and socio-demographic variables. Size of home and number of occupants are significantly associated with increased energy use, and the type of heating system (using electricity or gas) in a northern European context strongly influences average daily consumption.

Concurring with findings from studies of psychological and social predictors, we also find that certain behavioural and attitudinal measures are selected by the lasso model as predictors of domestic energy consumption. These include variables relating to frugality and conscientiousness such as indicated in the Big Five Inventory or those who say they responded to energy price rises by conserving energy (Broberg & Kažukauskas, 2021; Lesic et al., 2018). By contrast, income and spending correlate with greater electricity and gas usage (Frederiks et al., 2015).

We also find correlation between pro-environmental attitudes and behaviours not directly related to energy, and lower household energy use (Schultz et al., 2007; Horne & Kennedy, 2017). These include recycling, buying products with less packaging, willingness to reduce car travel, and perceptions about social norms relating to heating behaviours. One surprising finding is that individuals who tend to be less trusting of others use less energy (gas and electricity) at home, and those who value power — according to the Theory of Basic Human Values (Bouman et al., 2018) — use relatively less electricity.

Limitations

This study has limitations with respect to its design, methods, and underlying data. As discussed in section 3.1, the CC survey was designed to be wide-ranging, necessitating two waves. This led to slight under- and over- representation of some demographics in comparison to the UK population as a whole. The resulting sample over-represents older people and women, and it under-represents those with lower educational qualifications.

Our approach to predict energy consumption for CC Survey respondents using models trained on SERL data was facilitated by replicating the survey questions used by SERL verbatim. However, the SERL survey that was conducted alongside the collection of smart meter data (and thus the set of variables on which we could develop predictive models) is limited in extent. It excludes predictors known to be important for energy use, such as household income. A greater number of overlapping questions may have increased the accuracy of the predictive model that was applied to the CC survey. Nonetheless, model predictive errors are relatively low in terms of the units of the dependent variable (14.3% and 5.2% of mean daily electricity and gas usage for the sample, respectively), suggesting reasonable predictive accuracy in this context.

Regarding our selection of modelling techniques, there are some limitations to the use of lasso regression. Namely, lasso regression is not typically appropriate for use in statistical inference because it is a selective method that already “searches” the data set for the strongest associations with the response variable, which means that any follow-on statistical inference test (e.g., for determining variables' significance level) must have a higher threshold (Taylor & Tibshirani, 2015). Several studies have explored significance tests for lasso regression (Lee et al., 2016; Lockhart et al., 2014), but we do not take the step of applying these methods in this paper, instead using conventional significance tests to present p-values and confidence intervals alongside model coefficients in our results. The effect is that p-values may be slightly lower and confidence intervals slightly wider than would be the case if we used lasso-specific significance tests.

Finally, our treatment of missing data is limited given that we use a relatively coarse method (list-wise deletion) to remove observations with missing data, which reduces the resulting sample sizes from the SERL data on which we train our predictive models. Approaches to impute missing data abound in the statistics literature, and it is possible that our use of one of these approaches could have improved model predictive accuracy, but given the large starting size of our data sets and also the computational burdens of multiple imputation (as well as the inspection of model training and prediction error, which confirmed that these were relatively small), we determined that such treatment of missing data was not warranted.

Implications and future research

This paper has demonstrated the value of $l_1$regularisation (lasso regression) as a method for improving the interpretability of multiple regression models that can identify predictors of domestic energy consumption without sacrificing model predictive accuracy. Moreover, we have demonstrated the value of such an approach for generating energy consumption estimates (the metric of interest) for a national survey sample for which no energy-related data was collected. This study makes an important contribution to overcoming a significant obstacle in studies of the psychological, social, and behavioural influences on energy consumption: that survey instruments cannot be used to gather reliable data on household energy consumption (Attari et al., 2010; Marghetis et al., 2019). This has implications for social scientists focused on energy and the environment, including in psychology, sociology, and geography. The UK case has wider relevance in Europe and beyond given its broad mix of housing types and tenancy and diverse population. For countries with advanced smart-meter programmes, there is scope to reproduce the SERL data portal, and our study demonstrates the potential to leverage this unique dataset to generate broader social scientific insights.

This study contributes to the growing literature on carbon capability (Hampton & Whitmarsh, 2024; Wei et al., 2016; Whitmarsh et al., 2011), which highlights the multiple influences on an individuals' climate impacts, including attitudes, social norms, infrastructures, and systems of provision. The carbon capability framework was used to design the extensive survey that underpins this study and led to the inclusion of variables relating to 'citizenship' (e.g., voting preferences and policy support), as well as the 'influence' of social networks and norms (Hampton & Whitmarsh, 2023). The models reveal that several variables related to social influence have a relationship with household energy use, whereas relatively few of the citizenship behaviours are selected (e.g., voting green is included, but policy support is not). Building on this study, there is scope for future research to quantify the other main categories of household-level emissions, including transport, food consumption, and the purchase of goods and services.

Our study adds to the large base of empirical evidence that identifies dwelling characteristics (age, size), occupancy, and type of heating system as the primary determinants of household energy consumption in the UK. Generally in our model, psychological, social, and behavioural variables have smaller magnitudes of effect size than those relating to demographic and building characteristics, as found in other research modelling predictors of household energy consumption (Huebner et al., 2016; McKenna et al., 2022; Satre-Meloy, 2019). Still, the cumulative effect of multiple related factors (such as those related to householder frugality and conscientiousness) could be substantial, highlighting the potential for behaviour to drive down consumption and associated emissions (Dietz et al., 2009).

Further, it would be an error to infer that small effect sizes means that the factors which relate to the choices and behaviours of householders are insignificant in explaining variance in domestic energy consumption. Our models of electricity and gas, like most similar studies, have R-squared values of less than 0.5, meaning their explanatory power is limited. The remaining variability observed in household consumption across the population is likely to *also* be attributable to diverse choices and behaviours and influenced by myriad social and psychological factors that are difficult to capture with survey methods (Sonderegger, 1978; Frederiks et al. 2016; Burgess & Whitehead, 2020). This supports the need for in-depth qualitative research that uncovers the variable and unpredictable nature of household energy consumption. This in turn can help to improve large scale studies using quantitative techniques.

These insights can also inform policy design and implementation. This includes promoting frugality and conscientiousness amongst householders via information campaigns, education, and targeted and personalised feedback. Interventions which emphasise pro-environmental benefits and injunctive norms are likely to be effective. Our findings also highlight the need to target interventions toward high-income, high-consumption households. These groups, with greater capacity to adopt low-carbon technologies such as heat pumps and electric vehicles, are prime candidates for the promotion of renewable heating systems and efficiency measures.

Acknowledgements This work was funded by the US Department of Energy Office of Energy Efficiency and Renewable Energy Building Technologies Office under Lawrence Berkeley National Laboratory contract no. DE-AC02-05CH11231. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, or The Regents of the University of California.

Author contribution Study conception and design, material preparation, and data collection were performed by Sam Hampton. All authors contributed to data analysis, drafting the manuscript, and revising previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The data sets generated and analysed in this study are available from the corresponding author upon reasonable request.

Declarations

Competing interests The authors declare no competing interests.

Appendix

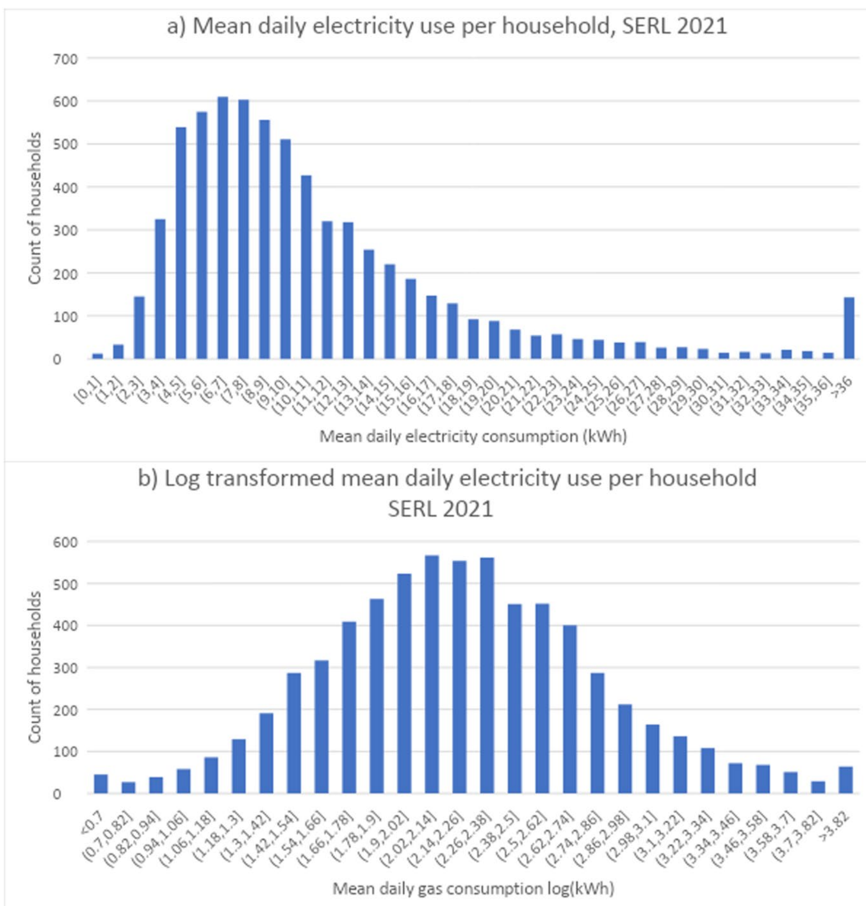

Fig. 4 Distribution of SERL data mean daily electricity consumption for 2021 (n=6,751). **a** untransformed distribution. **b** log-transformed distribution

Additional Methods and Results

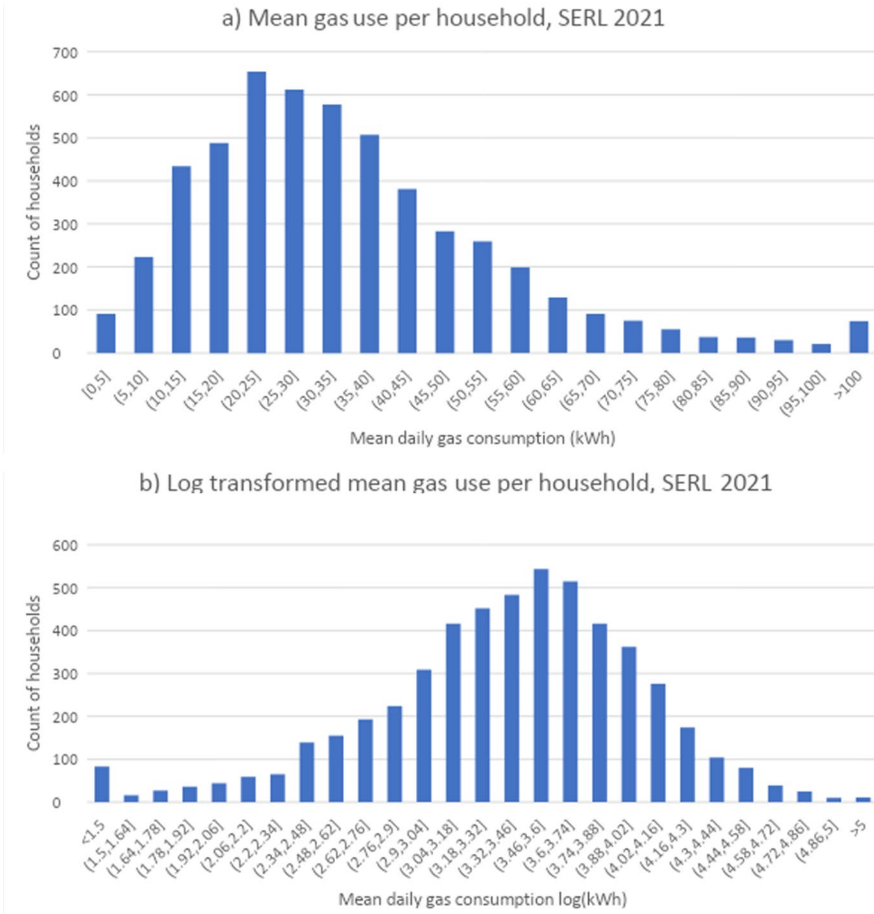

Fig. 5 Distribution of SERL data mean daily gas consumption for 2021 (n=5,256). **a** untransformed distribution. **b** log-transformed distribution

References

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2007). The effect of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents. *Journal of Environmental Psychology, 27*(4), 265–276. https://doi.org/10.1016/j.jenvp.2007.08.002

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Attari, S. Z., DeKay, M. L., Davidson, C. I., & Bruine de Bruin, W. (2010). Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Sciences, 107*, 16054–16059. https://doi.org/10.1073/pnas.1001509107

Bartiaux, F., & Gram-Hanssen, K. (2005). Socio-political factors influencing household electricity consumption: A comparison between Denmark and Belgium (pp. 1313–1325). Presented at the Proceedings of the ECEEE.

Becker, L. J., Seligman, C., Fazio, R. H., & Darley, J. M. (1981). Relating Attitudes to Residential Energy Use. *Environment and Behavior, 13*(5), 590–609. https://doi.org/10.1177/0013916581135004

BEIS. (2020). International comparisons of household energy efficiency. https://assets.publishing.service.gov.uk/media/5e7b58fcd3bf7f133f849328/International_comparisons_of_household_energy_efficiency_2020.pdf

BEIS. (2022). *BEIS Public Attitudes Tracker: Heat and Energy in the Home*. Department for Business, Energy and Industrial Strategy.

BEIS. (2023). *Smart meters: Quarterly report to end December 2023*. Department for Business, Energy & Industrial Strategy https://www.gov.uk/government/collections/smart-meters-statistics

Bouman, T., Steg, L., & Kiers, H. A. L. (2018). Measuring Values in Environmental Research: A Test of an Environmental Portrait Value Questionnaire. *Frontiers in Psychology, 9*.

Brandon, G., & Lewis, A. (1999). Reducing household energy consumption: A qualitative and quantitative field study. *Journal of Environmental Psychology, 19*(1), 75–85. https://doi.org/10.1006/jevp.1998.0105

Bring, J. (1994). How to Standardize Regression Coefficients. *The American Statistician, 48*(3), 209–213. https://doi.org/10.2307/2684719

Broberg, T., & Kažukauskas, A. (2021). Information policies and biased cost perceptions - The case of Swedish residential energy consumption. *Energy Policy, 149*, 112095. https://doi.org/10.1016/j.enpol.2020.112095

Burgess, M., & Whitehead, M. (2020). Just Transitions, Poverty and Energy Consumption: Personal Carbon Accounts and Households in Poverty. *Energies, 13*(22), 5953. https://doi.org/10.3390/en13225953

Cole, R. J., Robinson, J., Brown, Z., & O'shea, M. (2008). Recontextualizing the notion of comfort. *Building Research and Information, 36*(4), 323–336. https://doi.org/10.1080/09613210802076328

Corner, A., & Clarke, J. (2016). *Talking Climate: From Research to Practice in Public Engagement*. Palgrave.

Costa Mastrascusa, R., de Oliveira Fenili Antunes, M. L., de Albuquerque, N. S., Virissimo, S. L., Foletto Moura, M., Vieira Marques Motta, B., de Lara Machado, W., Moret-Tatay, C., & Quarti Irigaray, T. (2023). Evaluating the complete (44-item), short (20-item) and ultrashort (10-item) versions of the Big Five Inventory (BFI) in the Brazilian population. *Scientific Reports, 13*, 7372. https://doi.org/10.1038/s41598-023-34504-1

Curtin, N., Stewart, A. J., & Duncan, L. E. (2010). What Makes the Political Personal? Openness, Personal Political Salience, and Activism. *Journal of Personality, 78*, 943–968. https://doi.org/10.1111/j.1467-6494.2010.00638.x

Daoud, J. I. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series, 949*, 012009. https://doi.org/10.1088/1742-6596/949/1/012009

Darby, S. (2006). *The Effectiveness of Feedback on Energy Consumption*. ECI.

de Dear, R. J., & Brager, G. S. (1997). Developing an Adaptive Model of Thermal Comfort and Preference - Final Report on RP-884. *ASHRAE Transactions, 104*(1).

Dietz, T., Gardner, G. T., Gilligan, J., Stern, P. C., & Vandenbergh, M. P. (2009). Household Actions Can Provide a Behavioral Wedge to Rapidly Reduce US Carbon Emissions. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 18452–18456. https://doi.org/10.2307/25593033

DLUHC. (2022). *English Housing Survey: Headline Report 2021-22*. Ministry of Levelling Up, Housing and Communities.

Dubois, G., Sovacool, B., Aall, C., Nilsson, M., Barbier, C., Herrmann, A., Bruyère, S., Andersson, C., Skold, B., Nadaud, F., Dorner, F., Moberg, K. R., Ceron, J. P., Fischer, H., Amelung, D., Baltruszewicz, M., Fischer, J., Benevise, F., Louis, V. R., & Sauerborn, R. (2019). It starts at home? Climate policies targeting household consumption and behavioral decisions are key to low-carbon futures. *Energy Research and Social Science, 52*, 144–158. https://doi.org/10.1016/j.erss.2019.02.001

Duncan, L. E., & Stewart, A. J. (2007). Personal Political Salience: The Role of Personality in Collective Identity and Action. *Political Psychology, 28*(2), 143–164. https://doi.org/10.1111/j.1467-9221.2007.00560.x

DWP. (2023). *Family Resources Survey: Financial year 2021/22*. UK Department for Work and Pensions Retrieved from https://www.gov.uk/government/collections/family-resources-survey%2D%2D2

Eurostat (2023). Energy use in households up 6% in 2021.

Fanger, O. (1970). *Thermal Comfort, Analysis and Applications in Environmental Engineering*. McGraw Hill.

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in Regression Analysis: The Problem Revisited. *The Review of Economics and Statistics, 49*(1), 92–107. https://doi.org/10.2307/1937887

Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews, 23*, 272–288. https://doi.org/10.1016/j.rser.2013.03.004

Frederiks, E. R., Stenner, K., & Hobman, E. V. (2015). The Socio-Demographic and Psychological Predictors of Residential Energy Consumption: A Comprehensive Review. *Energies, 8*, 573–609. https://doi.org/10.3390/en8010573

Frederiks, E. R., Stenner, K., Hobman, E. V., & Fischle, M. (2016). Evaluating energy behavior change programs using randomized controlled trials: Best practice guidelines for policymakers. *Energy Research & Social Science, 22*, 147–164. https://doi.org/10.1016/j.erss.2016.08.020

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software, 33*, 1–22. https://doi.org/10.18637/jss.v033.i01

Fumo, N., & Rafe Biswas, M. A. (2015). Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews, 47*, 332–343. https://doi.org/10.1016/j.rser.2015.03.035

Gatersleben, B., Steg, L., & Vlek, C. (2002). Measurement and Determinants of Environmentally Significant Consumer Behavior. *Environment and Behavior, 34*(3), 335–362. https://doi.org/10.1177/0013916502034003004

Gram-Hanssen, K. (2010). Residential heat comfort practices: understanding users. *Building Research and Information, 38*(2), 175–186. https://doi.org/10.1080/09613210903541527

Gram-Hanssen, K. (2013). Efficient technologies or user behaviour, which is the more important when reducing households' energy consumption? *Energy Efficiency, 6*(3), 447–457. https://doi.org/10.1007/s12053-012-9184-4

Hampton, S., & Whitmarsh, L. (2023). Choices for climate action: A review of the multiple roles individuals play. *One Earth, 6*(9), 1157–1172. https://doi.org/10.1016/j.oneear.2023.08.006

Hampton, S., & Whitmarsh, L. (2024). Carbon capability revisited: Theoretical developments and empirical evidence. *Global Environmental Change, 87*, 102895. https://doi.org/10.1016/j.gloenvcha.2024.102895

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity (1st edition)*. Routledge.

Horne, C., & Kennedy, E. H. (2017). The power of social norms for reducing and shifting electricity use. *Energy Policy, 107*, 43–52. https://doi.org/10.1016/j.enpol.2017.04.029

Hsu, D. (2015). Identifying key variables and interactions in statistical models of building energy consumption using regularization. *Energy, 83*, 144–155. https://doi.org/10.1016/j.energy.2015.02.008

Huebner, G., Shipworth, D., Hamilton, I., Chalabi, Z., & Oreszczyn, T. (2016). Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes. *Applied Energy, 177*, 692–702. https://doi.org/10.1016/j.apenergy.2016.04.075

Huebner, G. M., Cooper, J., & Jones, K. (2013). Domestic energy consumption—What role do comfort, habit, and knowledge about the heating system play? *Energy and Buildings, 66*, 626–636.

Huebner, G. M., Hamilton, I., Chalabi, Z., Shipworth, D., & Oreszczyn, T. (2015). Explaining domestic energy consumption – The comparative contribution of building factors, socio-demographics, behaviours and attitudes. *Applied Energy, 159*, 589–600. https://doi.org/10.1016/j.apenergy.2015.09.028

IEA (2022). Energy Efficiency 2022. https://www.iea.org/reports/energy-efficiency-2022

Ivanova, D., Barrett, J., Wiedenhofer, D., Macura, B., Callaghan, M., & Creutzig, F. (2020). Quantifying the potential for climate change mitigation of consumption options. *Environmental Research Letters, 15*, 093001. https://doi.org/10.1088/1748-9326/ab8589

Jones, R. V., Fuertes, A., & Lomas, K. J. (2015). The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. *Renewable and Sustainable Energy Reviews, 43*, 901–917. https://doi.org/10.1016/j.rser.2014.11.084

Kavousian, A., Rajagopal, R., & Fischer, M. (2013). Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. *Energy, 55*, 184–194. https://doi.org/10.1016/j.energy.2013.03.086

Kleinschafer, J., & Morrison, M. (2014). Household norms and their role in reducing household electricity consumption. *International Journal of Consumer Studies, 38*(1), 75–81. https://doi.org/10.1111/ijcs.12066

Köster, E. P. (2009). Diversity in the determinants of food choice: A psychological perspective. *Food Quality and Preference, 20*(2), 70–82. https://doi.org/10.1016/j.foodqual.2007.11.002

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software, 28*(5). https://doi.org/10.18637/jss.v028.i05

Larson, N., & Story, M. (2009). A Review of Environmental Influences on Food Choices. *Annals of Behavioral Medicine, 38*(suppl_1), s56–s73. https://doi.org/10.1007/s12160-009-9120-9

Lee, J. D., Sun, D. L., Sun, Y., & Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics, 44*(3), 907–927. https://doi.org/10.1214/15-AOS1371

Lesic, V., de Bruin, W. B., Davis, M. C., Krishnamurti, T., & Azevedo, I. M. L. (2018). Consumers' perceptions of energy use and energy savings: A literature review. *Environmental Research Letters, 13*(3), 033004. https://doi.org/10.1088/1748-9326/aaab92

Lever, J., Krzywinski, M., & Altman, N. (2016). Model selection and overfitting. *Nature Methods, 13*(9), 703–704. https://doi.org/10.1038/nmeth.3968

Lockhart, R., Taylor, J., Tibshirani, R. J., & Tibshirani, R. (2014). A significance test for the lasso. *The Annals of Statistics, 42*(2), 413–468. https://doi.org/10.1214/13-AOS1175

Marghetis, T., Attari, S. Z., & Landy, D. (2019). Simple interventions can correct misperceptions of home energy use. *Nature Energy, 4*(10), 874–881. https://doi.org/10.1038/s41560-019-0467-2

McKenna, E., Few, J., Webborn, E., Anderson, B., Elam, S., Shipworth, D., et al. (2022). Explaining daily energy demand in British housing using linked smart meter and socio-technical data in a bottom-up statistical model. *Energy and Buildings, 258*, 111845. https://doi.org/10.1016/j.enbuild.2022.111845

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

MHCLG. (2023). *English Housing Survey: 2021 to 2022 headline report*. Ministry of Housing, Communities & Local Government Retrieved from https://www.gov.uk/government/collections/english-housing-survey

Midden, C. J., & Ritsema, B. S. (1983). The meaning of normative processes for energy conservation. *Journal of Economic Psychology, 4*(1–2), 37–55.

Mukai, T., Nishio, K., Komatsu, H., & Sasaki, M. (2022). What effect does feedback have on energy conservation? Comparing previous household usage, neighbourhood usage, and social norms in Japan. *Energy Research & Social Science, 86*, 102430. https://doi.org/10.1016/j.erss.2021.102430

Ndiaye, D., & Gabriel, K. (2011). Principal component analysis of the electricity consumption in residential dwellings. *Energy and Buildings, 43*(2), 446–453. https://doi.org/10.1016/j.enbuild.2010.10.008

Ofgem (2023). Typical Domestic Consumption Values.

Ofgem. (2024). *Average gas and electricity usage [WWW Document]*. Ofgem https://www.ofgem.gov.uk/infor

mation-consumers/energy-advice-households/average-gas-and-electricity-use-explained (accessed 1.12.24).
ONS (2021). Census 2021: People, population and community.
ONS. (2023a). *UK Environmental Accounts: 2023, UK Environmental Accounts*. Office for National Statistics.
ONS. (2023b). *Energy efficiency of housing in England and Wales: 2023*. Office for National Statistics https://www.ons.gov.uk/peoplepopulationandcommunity/housing/articles/energyefficiencyofhousinginenglandandwales/2023. Accessed 30 January 2024
Rdc, T. 2010. R: A language and environment for statistical computing. (No Title). https://cir.nii.ac.jp/crid/1370294721063650048. Accessed 30 January 2024
Roberts, N., & Thatcher, J. (2009). Conceptualizing and testing formative constructs: Tutorial and annotated example. *ACM SIGMIS Database, 40*(3), 9–39.
Roser-Renouf, C., Maibach, E. W., Leiserowitz, A., & Zhao, X. (2014). The genesis of climate change activism: from key beliefs to political action. *Climatic Change, 125*(2), 163–178. https://doi.org/10.1007/s10584-014-1173-5
Samuelson, C. D., & Biek, M. (1991). Attitudes toward energy conservation: A confirmatory factor analysis 1. *Journal of Applied Social Psychology, 21*(7), 549–568.
Satre-Meloy, A. (2019). Investigating structural and occupant drivers of annual residential electricity consumption using regularization in regression models. *Energy, 174*, 148–168. https://doi.org/10.1016/j.energy.2019.01.157
Satre-Meloy, A., Diakonova, M., & Grünewald, P. (2020). Daily life and demand: an analysis of intra-day variations in residential electricity consumption with time-use data. *Energy Efficiency, 13*(3), 433–458. https://doi.org/10.1007/s12053-019-09791-1
Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science, 18*(5), 429–434. https://doi.org/10.1111/j.1467-9280.2007.01917.x
Seligman, C., Kriss, M., Darley, J. M., Fazio, R. H., Becker, L. J., & Pryor, J. B. (1979). Predicting Summer Energy Consumption from Homeowners' Attitudes. *Journal of Applied Social Psychology, 9*(1), 70–90. https://doi.org/10.1111/j.1559-1816.1979.tb00795.x
SERL. (2022a). *Smart Energy Research Lab: Statistical Data, (2019-2021): Safeguarded Access*. https://doi.org/10.5255/UKDA-SN-8963-1
SERL (2022b). Energy Use in GB Domestic Buildings 2021 - Aggregated Statistics.
Shmueli, G. (2010). To Explain or to Predict? *Statistical Science, 25*(3), 289–310. https://doi.org/10.1214/10-STS330
Sonderegger, R. C. (1978). Movers and stayers: The resident's contribution to variation across houses in energy consumption for space heating. *Energy and Buildings, 1*(3), 313–324. https://doi.org/10.1016/0378-7788(78)90011-7
Taylor, J., & Tibshirani, R. J. (2015). Statistical learning and selective inference. *Proceedings of the National Academy of Sciences, 112*(25), 7629–7634. https://doi.org/10.1073/pnas.1507583112
Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B: Methodological, 58*, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
Van Raaij, W. F., & Verhallen, T. M. M. (1983). A behavioral model of residential energy use. *Journal of Economic Psychology, 3*(1), 39–63. https://doi.org/10.1016/0167-4870(83)90057-0
Vining, J., Ebreo, A., Bechtel, R., & Churchman, A. (2002). Emerging theoretical and methodological perspectives on conservation behavior. *Handbook of Environmental Psychology, 2*, 541–558.
Vringer, K., Aalbers, T., & Blok, K. (2007). Household energy requirement and value patterns. *Energy Policy, 35*(1), 553–566. https://doi.org/10.1016/j.enpol.2005.12.025
Wahlström, M. H., & Hårsman, B. (2015). Residential energy consumption and conservation. *Energy and Buildings, 102*, 58–66. https://doi.org/10.1016/j.enbuild.2015.05.008
Webb, D., Soutar, G. N., Gagné, M., Mazzarol, T., & Boeing, A. (2022). Saving energy at home: Exploring the role of behavior regulation and habit. *International Journal of Consumer Studies, 46*(2), 621–635. https://doi.org/10.1111/ijcs.12716
Webborn, E., Few, J., McKenna, E., Elam, S., Pullinger, M., Anderson, B., Shipworth, D., & Oreszczyn, T. (2021). The SERL Observatory Dataset: Longitudinal Smart Meter Electricity and Gas Data, Survey, EPC and Climate Data for over 13,000 Households in Great Britain. *Energies, 14*, 6934. https://doi.org/10.3390/en14216934
Webborn, E., Few, J., & McKenna, E. (2022). Smart meter consumption data: Technical documentation (UK Data Archive Study Number 8666 - Smart Energy Research Lab Observatory Data: Secure Access). London: University College London. https://doc.ukdataservice.ac.uk/doc/8666/mrdoc/pdf/serl_smart_meter_documentation_edition05.pdf
Wei, J., Chen, H., Cui, X., & Long, R. (2016). Carbon capability of urban residents and its structure: Evidence from a survey of Jiangsu Province in China. *Applied Energy, 173*, 635–649. https://doi.org/10.1016/j.apenergy.2016.04.068
Whitmarsh, L., Seyfang, G., & O'Neill, S. (2011). Public engagement with carbon and climate change: To what extent is the public 'carbon capable'? *Global Environmental Change, 21*, 56–65. https://doi.org/10.1016/j.gloenvcha.2010.07.011
Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., Software, P., & PBC. (2023). dplyr: A Grammar of Data Manipulation. https://cloud.r-project.org/web/packages/dplyr/index.html. Accessed 30 January 2024
Wynes, S., & Nicholas, K. A. (2017). The climate mitigation gap: education and government recommendations miss the most effective individual actions. *Environmental Research Letters, 12*, 074024. https://doi.org/10.1088/1748-9326/aa7541
Zibarras, L. D., & Coan, P. (2015). HRM practices used to promote pro-environmental behavior: a UK survey. *International Journal of Human Resource Management, 26*, 2121–2142. https://doi.org/10.1080/09585192.2014.972429

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.