This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

Explaining daily energy demand in British housing using linked smart meter and socio-technical data in a bottom-up statistical model

Eoghan McKenna [a], Jessica Few [a], Ellen Webborn [a], Ben Anderson [b], Simon Elam [a], David Shipworth [a], Adam Cooper [c], Martin Pullinger [a], Tadj Oreszczyn [a]

[a] UCL Energy Institute, 14 Upper Woburn Place, University College London, WC1H 0NN, UK
[b] Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, UK
[c] Department of Science, Technology, Engineering and Public Policy, University College London, Shropshire House, Capper Street, London WC1E 6JA, UK

Abstract

This paper investigates factors associated with variation in daily total (electricity and gas) energy consumption in domestic buildings using linked pre-COVID-19 smart meter, weather, building thermal characteristics, and socio-technical survey data covering appliance ownership, demographics, behaviours, and attitudes for two nested sub-samples of 1418 and 682 British households selected from the Smart Energy Research Laboratory (SERL) Observatory panel.

Linear mixed effects modelling resulted in adjusted R^2 between 63% and 80% depending on sample size and combinations of contextual data used. Increased daily energy consumption was significantly associated (*p*-value<0.05, VIF<5) with: households living in buildings with more rooms and bedrooms, that are older, more detached, have air-conditioning, and experience colder (more heating degree days) or less sunny weather; households with more adult occupants, more children, older adult occupants, higher heating temperature setpoints, and that do not try to save energy.

The results demonstrate the value of smart meter data linked with contextual data for improving understanding of energy demand in British housing. Accredited UK researchers are invited to apply to access the data, which has recently been updated to include over 13,000 households from across Great Britain. This paper provides guidance on appropriate methods to use when analysing the data.

Key words: building; energy; heating; gas; electricity; demand; consumption; household; residential; domestic; smart meter; daily; longitudinal; regression; mixed effects; random effects; survey; energy performance certificate; weather; temperature; solar radiation; building physics; sociodemographic; occupant; behaviour; attitudes.

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

1 Introduction

Increasingly governments are pledging to reduce greenhouse gas emissions to net zero by 2050 [1]. The building sector requires rapid decarbonisation of energy supply and wide-spread reduction in energy demand through improvements in energy efficiency, changes in behaviour and avoided energy use [1,2]. A critical starting point to achieve this is an effective characterisation of energy demand in buildings i.e. 'what norms, values, preferences and structural factors shape energy demand?' [3]. Cooper has emphasised the need for better integration of research approaches across social and physical science research for energy policy impact [4] and provides a conceptual framework for reasoning how to integrate data validly from different physical and social sources to enable socio-technical research [5]. This integration is required to better explain patterns in demand, identify the factors which are associated with greatest impact on demand, and to better inform effective policy instruments targeted at improving energy efficiency or changing occupant behaviour [4,6]. Moreover, effective characterisation can enable improved predictions of demand. This could reduce demand in buildings if used to identify (and potentially reward) changes in demand (e.g. due to an intervention, or in response to a tariff or energy efficiency installation). Prediction can also be used to diagnose problems such as malfunctioning heating systems, poor quality build, or energy waste in the form of heating or lighting in unoccupied buildings.

A greater understanding of energy demand in buildings has been impeded by limited data about energy demand and its influencing factors [7–9]. The Smart Energy Research Lab (SERL) is a five-year UK research council funded project which aims to address this by bringing together, for the first time, half-hourly resolution household-level electricity and gas demand data with detailed socio-technical and weather data for a representative sample of over 13,000 households in Great Britain (GB) (the 'SERL Observatory'). In this respect the data captures a much wider array of energy demand co-variates *and* more detailed energy use data than has previously been reported in the literature.

The first aim of this paper is therefore to evaluate the SERL Observatory as a data resource to improve current characterisations of household-level energy demand. Linear regression is commonly used in the literature to characterise household demand given multi-variate demand-side datasets. This is usually done in two ways; first by assessing the overall *explanatory power* of an appropriately validated statistical model applied to the data, usually by assessing the model's errors (residuals) and associated statistics such as the R^2/adjusted R^2, or coefficient of determination. Low errors imply that the data includes appropriate variables, and the model captures appropriate relationships between them, such that variation in the variable of interest can be explained given the model and data. The model can be applied to other data and tested for prediction or forecasting purposes. Second, studies scrutinise the results of the model to identify specific variables which are statistically significantly associated with variation in the variable of interest *and* which have a substantively interesting or useful effect. Such variables can be interpreted as important factors related to household-level demand, leading to a more detailed understanding of residential demand, and can inform policies aimed at targeting such key factors and reducing demand in future. This leads to our first two research questions:

> *1: What is the overall explanatory power of SERL Observatory data with respect to variation in household-level daily residential energy consumption and does this improve on studies reported in the literature?*

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

> *2: Which variables observed in SERL Observatory data are most strongly associated with household-level daily residential energy consumption, are these associations statistically significant, and do these confirm and extend results reported in the literature?*

The SERL Observatory links:

- Energy consumption data from smart meters (at daily and half-hourly resolution) with the following three contextual datasets (described in more detail later);
- Basic data (non-building specific publicly available area data): dwelling region, local area Index of Multiple Deprivation[1] (IMD) for 2019, and local area hourly weather variables;
- SERL survey: occupant-reported household-specific sociodemographic characteristics and energy saving behaviour, and some building-specific physical characteristics;
- EPC (Energy Performance Certificate) data: building-specific physical and thermal characteristics plus a modelled normative fuel cost.

These four datasets have different levels of availability: all households in the SERL Observatory have daily and half-hourly energy consumption data from smart meters, and the basic data (see above), around 80% have complete SERL survey data, and approximately half have EPC data as only about half of British properties have an EPC. Researchers using SERL data are therefore presented with a choice: to increase sample size but reduce contextual data, or decrease sample size and increase contextual data. Determining the usefulness of the datasets separately and together is therefore important for the overall objective of characterising demand. This leads to our final research question:

> *3: What is the additional explanatory power of the EPC and SERL survey contextual data beyond that of the basic data?*

We answer this question by investigating how the explanatory power of the model changes with different levels of contextual data and sample size, and testing whether the differences are statistically significant.

Our analysis uses the SERL Observatory Edition 2 dataset [10]. This contains data from almost 5000 households and energy demand data from August 2018 to October 2020. Data collection is ongoing and subsequent editions will be updated with this newly collected data. The first coronavirus lockdown in GB started on 23rd March 2020, meaning Edition 2 includes data from before the onset of the coronavirus pandemic. It is important to understand the impact of the coronavirus pandemic on residential energy consumption in buildings and what constitutes the post-pandemic 'new normal', and the SERL Observatory is a well-suited data resource to do this and currently supports several research projects investigating the effects of the pandemic. This paper aims to provide a foundation to this forthcoming research by seeking to understand and characterise residential energy consumption *pre-coronavirus*.

Given the contextual data availability requirements from waves 1 and 2 only, and the focus on the pre-lockdown period, the resulting samples of households analysed here are relatively small subsamples (N=1418 and 682) compared to the number of households that will be available in later editions of the SERL Observatory (Edition 3 increases the sample size to >13,000). These results should therefore be seen as an initial analysis and should be interpreted with caution. In particular,

[1] Government statistical estimate of relative deprivation in small areas – see https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019

results should not be viewed as representative of the future full-size SERL Observatory sample, nor indeed the GB population and so should not be generalised.

2 Literature review

A substantial body of existing literature investigates the factors which shape household energy use, and the first section below describes this literature. The remaining sections describe relevant literature on quantitative approaches to modelling building energy demand. The literature has informed the development of the SERL Observatory dataset and the analytical methods used in this study.

2.1 Factors influencing household energy demand

Household energy demand can be viewed as the outcome of occupants making use of the energy-using appliances and equipment in their home, largely through everyday activities such as cleaning, food preparation, leisure and keeping warm (or cool). At the population level and over periods of years, structural drivers can be seen as having a major effect on average energy use for these activities, with policy-driven incremental energy efficiency improvements in technology often being offset by an increasing intensity of appliance use [11,12]. The development and diffusion of new technologies and more radical changes in social norms and expectations can also lead to more significant changes in the average energy intensity of such practices [13]. Variation in energy use at shorter timescales is a matter of variation between households, the currently available technologies they have, and how they make use of them. Multiple studies have attempted to identify which aspects of occupants, of their activities, and of the technologies they use and buildings they occupy, are most important for explaining inter-household variation in final energy demand. Huebner et al [8] found building characteristics, particularly size, type and energy performance rating (as provided by EPCs), dominate in explaining between-household variation in energy use in a sample of contemporary English households, with household size (number of occupants) also important, as well as the length of the heating season and reported beliefs about climate change.

Other quantitative studies unpack overall energy use. Gram-Hanssen [12] separately investigates energy used for heating and energy used for appliances and lighting, in Danish households. Drawing on multiple data sources and analytical approaches, she concludes that user behaviour (including appliance ownership) is a more substantial factor shaping energy used for appliances and lighting than is appliance energy efficiency, noting, for example, that energy use in physically similar houses can vary by a factor of 5. Energy use for heating meanwhile is found to be roughly equally explained by building characteristics, including size and age, and by user behavioural factors, whilst sociodemographic characteristics (age, income and education) explain very little, indicating that they only weakly correlate with a person's heating behaviours.

Many other studies focus on a single fuel type rather than end use. Jones et al. [14] provided a literature review of nearly 40 empirical studies of household electricity use, identifying 62 factors that potentially affect it, with 20 "found to unambiguously have a significant positive effect on electricity use" (defined by the authors as the number of papers confirming a positive effect being more than three higher than the number of papers finding a negative or non-significant effect). These 'unambiguous' variables were classed by the authors into socio-economic factors (more occupants, presence of teenagers, higher income and higher disposable income), characteristics of the dwelling (older dwellings, and higher number of rooms or number of bedrooms, or larger total floor area; presence of an electric space heating system, air-conditioning and/or an electric water heating system) and appliance-related factors (higher number of appliances, ownership of: desktop computers, televisions, electric ovens, refrigerators, dishwashers, tumble dryers; greater use of:

washing machines, tumble dryers). The categorical variables 'age of household reference person' and 'level of detachment of the building' also significantly affected electricity use. Further quantitative studies aim to specifically consider the influence of occupant behaviour, by combining time use data and electricity use data. Satre-Meloy et al. [15] find that variation between occupants in when and how electricity-using activities are performed does have a statistically significant effect on energy use, at least over the course of the day, finding from their own data and a review of previous studies that quotidian activities related to chores, food consumption and preparation, and leisure are particularly high energy intensity, and sleep and rest low intensity. Grunewald and Diakonova [16] extend this analysis to consider gendered differences in activity patterns and their corresponding electricity use, showing that while women report more household chores, their associated consumption of electricity is lower than for men in many cases. Differences and similarities between reported activities and associated electricity use for GB and German households is investigated in[17], finding that the need for flexibility and willingness to provide it differs significantly between two seemingly similar regions.

Regarding heating use, a review by Wei et al [18] of 41 papers found 27 factors identified in them as affecting space heating behaviour in residential buildings, concluding the following eight factors 'unambiguously' influenced it (using the same definition of unambiguous as above): "outdoor climate, dwelling type, room type, house insulation, type of temperature control, occupant age, time of day and occupancy".

Overall, the literature provides evidence that inter-household variation in energy use is related to building and appliance characteristics, occupant sociodemographics, behaviours, and contextual factors around climate, indoor conditions and time. Although existing studies provide some insight into which factors within these broad classes are 'unambiguously' important, Wei et al [18] note that the literature does not definitively rule out the influence of any factor that has been studied. Huebner et al [8] highlight that limitations in measurement methods, particularly for measuring behaviours, and collinearity and interaction effects between variables, can lead to factors appearing to have non-significant effects or being excluded from models, while Jones et al. [14] note that the often incomplete contextual information about sample characteristics (such as the fuel type used in the dwellings for space heating and cooling and water heating, or if there was mechanical ventilation) could explain some of the conflicting results found between studies regarding the influence of certain factors. We note further that a narrow framing of importance or influence of variables on demand in terms of statistical significance is often problematic considering it is not a measure of the size of the effect. In sum, there is value in continued research to investigate the effects of a wide range of variables within these broad classes of building, appliance, occupant and contextual factors.

2.2 Characterising energy demand in buildings

Characterising building energy demand is an active field of research employing a wide range of methods, depending on the data available and research objectives. Swan and Ugursal [19] provide a taxonomy of residential energy demand modelling approaches, grouping them into two broad categories of ‘top-down’ (a ‘macro’ approach where the housing stock as a whole is usually the unit of analysis) and ‘bottom-up’ (a ‘micro’ approach where the basic unit of analysis is usually individual dwellings), with the latter further sub-categorised into ‘statistical’ and ‘engineering’ methods. As this paper aims to characterise individual households, we adopt a bottom-up approach. Statistical regression is a common bottom-up approach that, while requiring large, detailed datasets, offer simple implementation and relatively easy interpretability. Jones et al. [14] provide a recent systematic review of studies using regression methods to explain electricity demand in residential buildings. Satre-Meloy et al. [20] provide a complementary and updated summary of the literature. Rather than duplicating these works, we draw broad observations relevant to the present work from the literature. The focus is on studies that used statistical approaches to characterise energy demand in residential buildings. Detailed reviews of alternative 'bottom-up' approaches (e.g. engineering, artificial neural networks) can be found in [21–24].

2.3 Explanation versus prediction

Multiple linear regression using ordinary least squares (OLS) is a technique commonly used in studies seeking to characterise energy demand in buildings using linked contextual data [20,25–28]. However, OLS relies on an assumption of independent observations which reduces its appropriateness for longitudinal data, in which there are repeated observations of individual cases [29]. Anderson et al. [28] used a linear mixed effects model (a type of multi-level regression model which accounts for grouped data) to address this in their study of daily electricity consumption using daily aggregates of sub-half-hourly household level energy use, similar to the current study.

Recent advances in statistical learning [30] have resulted in the emergence of new techniques in this field of research. There has been increased interest in techniques such as tree-based methods, support vector machine, and artificial neural networks [23,24]. These can be considered more 'flexible' than OLS because they allow non-linear relationships between variables. However, increased flexibility can come at a cost, with greater risk of over-fitting, increased model variance error, and potentially less interpretability [23,31]. These techniques tend to be more suitable for prediction, rather than inference which is the primary interest of this work.

As discussed in section 2.1 energy demand in buildings can be characterised by its large number of potentially influential factors. Studies seeking to characterise demand can therefore be faced with 'dense' models i.e. with many explanatory variables. Adding more variables to a model can spuriously increase its overall explanatory power. Nonetheless, assuming the increase is non-spurious, denser models can be accompanied by reductions in model interpretability, reductions in the reliability of estimates for individual variable coefficients (e.g. due to multi-collinearity), and reductions in sample sizes due to missing data. Increased model complexity can also result in over-fitting and a decrease in the model's predictive power [31].

Numerous techniques have been developed to deal with these issues including imputation to fill missing data, resampling methods such as cross-validation to test models, regularisation techniques to reduce the complexity of the model by selecting a subset of the total number of variables to include in the final model [32] and some of these have been applied in the field. For example, Kavousian et al. [33] use forward stepwise variable selection, while Huebner et al. [8] and Satre-Meloy et al. [20] use regularisation methods (or 'shrinkage' or 'penalised' regression), and the latter uses cross-validation and imputation. These techniques can be useful for increasing sample sizes, improving interpretation and, depending on the nature of the underlying data, can also improve model prediction. Imputation and cross-validation are used in the present paper, while regularisation techniques are not.

2.4 Heating demand and gas meter data

In GB natural gas is widely used for space and water heating and cooking e.g. 86% of dwellings in England supplied by the gas grid [34]. Moreover, heating demand is strongly weather dependent, and so it is crucial to understanding how total domestic energy demand changes over time. Therefore, observing both gas and electricity demand is necessary to achieve a data-driven characterisation of *total* residential energy consumption in GB dwellings *including heating* (note that cooling is currently very uncommon in UK homes) where gas and electricity are the only fuel sources.

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

In the case of the 14% of English dwellings that are not connected to the gas network, data on oil, LPG and solid fuel use would also be required but is not collected by SERL. We note, however, the relative difficulty of accessing gas demand data compared to electricity data (smart electricity meters are more widespread than smart gas meters [35] and it is easier to retro-fit sensor equipment to measure electricity demand data than it is for gas demand). This is reflected in the literature, which predominantly focuses on analysis of electricity demand compared to gas demand, even in countries where gas demand is present [15,25,27,36–38]. One of this paper's key contributions is that we analyse electricity and gas data where used by the household (as discussed below) thereby focussing on *total* domestic energy consumption including space and water heating for these households.

2.5 Temporal resolution of demand data

As noted above, the majority of studies in the literature focus on data of relatively low temporal resolution i.e. monthly, seasonal or annual summaries [14]. Studies focussing on daily or higher-resolution data are comparatively rare, presumably because of the relative difficulty of accessing large, high-resolution energy meter data sets which also have the necessary linked contextual data to investigate factors explaining variation in high-resolution energy demand [9]. Table 1 summarises key characteristics of studies from the literature chosen for their relevance (focus on household-level residential energy consumption, use of regression, household-level contextual data, annual or higher time resolution) and shows how much of the variation in demand (the 'coefficient of determination' or R^2) was explained by their models. We note that interpreting and comparing R^2 values between different studies using different models and different data should be treated with caution as R^2 values should not generally be compared across different data due to the fact that the same model can have highly variable R^2 values on different data [39]. Model error (e.g. mean squared error) is generally a better measure to compare however we note that often it is not reported in the literature.

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

Table 1 – summary of characteristics and key results of previous relevant studies.

Study	Data source	Country or area	Sample size (N)	Resolution of demand data	Observes heating / cooling?	Contextual data	Coefficient of determination (R^2/Adjusted R^2)
[26]	Korea Energy Economics Institute survey	Korea	2436	Annual	Yes	Building physical characteristics, socio-demographics, appliance usage	R^2: 0.009 - 0.017
[8]	Energy Follow-Up Survey	England	924	Annual	Yes	Building physical characteristics, socio-demographics, heating behaviour, attitudes and other behaviours	Adjusted R^2: 0.44
[20]	City of Palo Alto Utilities survey	Palo Alto, California	1008	Annual	Yes	Building physical characteristics, socio-demographics, energy literacy, attitudes	R^2: 0.373-0.398
[37]	Smart Grid Smart City	New South Wales	3446	Annual	No (no gas meter data)	Building physical characteristics, socio-demographics, appliances	Adjusted R^2: 0.55
[33]	Convenience sample	Silicon Valley, California	952	Averaged over a period of 238 days	Yes	Weather, building physical characteristics, appliances, and behaviour	Adjusted R^2: 0.43-0.68

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

[27]	Irish Commission for Energy Regulation's (CER) Smart Metering Electricity Customer Behaviour Trials	Ireland	~4200	Averaged over a period of 6 months	No (no gas meter data)	Building physical characteristics, socio-demographics, appliances	R^2: 0.32
[25]	Convenience sample	Japan	740	Monthly averaged demand	No (no gas meter data)	Weather, building and heating system information, household and appliance ownership and usage	Adjusted R^2: 0.18-0.60
[40]	Smart Grid Smart City	New South Wales	3446	Daily peak demand	No (no gas meter data)	Building physical characteristics, socio-demographics, appliances	Adjusted R^2: 0.29
[28]	Irish Commission for Energy Regulation's (CER) Smart Metering Electricity Customer Behaviour Trials	Ireland	3488	Daily	No (no gas meter data)	Income, employment status, presence of children, number of residents	Marginal R^2: 0.20. Conditional R^2: 0.81.
[15]	Convenience sample	UK	173	Daily	No (no gas meter data)	Building physical characteristics, socio-demographics, appliances, activity	Adjusted R^2: 0.44

[41]	National Energy Efficiency Data (NEED)	England and Wales	11.3M	Annual	Yes	Property characteristics Energy efficiency measures installed Household characteristics Local area characteristics	R^2: 0.38

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

McLoughlin et al. [27] analysed half-hourly electricity smart meter data and linked survey data for a representative sample of approximately 4200 Irish households involved in a time-of-use tariff trial. Multiple linear regression was used to estimate the influence of survey data (which covered dwelling and occupant characteristics) on the dwelling-level variability of total electricity demand, maximum demand, load factor, and the time of maximum demand all averaged over six months.

Anderson et al. [28] analysed the same dataset and investigated the extent to which the data from the survey can explain variability in load profile 'indicators' such as 97.5% percentile load, lunchtime load, morning maximum, etc. Dependent variables were sampled for midweek days (Tuesday-Thursday) over a four-week period resulting in 12 observations for each variable per participant. A mixed effects framework was used including a random effects coefficient to quantify how much each household deviated from the average.

Kavousian et al. [33] examined structural and behavioural determinants of residential energy consumption for a convenience sample of 1628 Californian households (952 used in final analysis). Participants were all employees of a single Silicon Valley technology company. As such the sample was biased towards higher income, higher education, and higher interest in energy efficiency households. 10-minute resolution electricity data were collected over 238 days and survey data were collected covering weather, location, building physical characteristics, appliances, and occupant information. While high-resolution energy data were collected, and daily electricity consumption was used as the dependent variable in the regression analysis, this was averaged over the collection period (238 days).

Iwafune and Yagita [25] analysed high-resolution (30-60 min) energy data for 740 Japanese households collected over a period of one year (Dec 2013-Nov 2014), and performed a regression on monthly-averaged daily electricity consumption using data on weather, building and heating system information, household and appliance ownership and usage. The study used a convenience sample determined by the presence of specific home energy management systems. Separate regressions were conducted for the different seasons of the year. Unlike Anderson et al. [28] the authors of the study did not include a household-specific effect but instead used a time-specific effect for each month.

Satre-Meloy et al. [15] analysed high-resolution (1 second) electricity data measured over a period of 28 hours on a convenience sample of 173 GB households. Electricity data were averaged over the collection period and within day sub-periods. Satre-Meloy et al. used 'de-minned' electricity demand in their regression in addition to average demand. De-minning subtracts each household's minimum demand from its average demand to remove its baseload and is particularly appropriate for studies aiming to characterise intra-day variations in demand that are affected by occupant activities.

Fan et al. [40] conducted a statistical analysis of drivers of peak demand by analysing half-hourly electricity consumption data collected over one year (2013) for 9900 households from the greater Sydney region linked with survey data for 3500 of these households covering housing type, demographics, appliance ownership, occupant living habits, and socioeconomic status. The study estimated individual peak demand over 12 selected peak demand periods in a year. A General Linear Model was used with 5-fold cross-validation. A mixed effects framework for analysing panel data was not used, unlike in Anderson et al. [28] or Iwafune and Yagita [25].

The review indicates that a linear mixed effect framework with random effects is appropriate when analysing panel data (cross-sectional plus time series data) and so will be used here for the analysis

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

of daily household-level total energy consumption as we have repeated (daily) observations at the household level alongside cross-sectional socio-technical and contextual data. Analysis of panel data without using mixed effects would not be correct as the structure of the model would not account for the grouped nature of the data [42], effectively assuming that every observation is independent, even where they are from the same household.

Finally, we note the high variability of R^2 for the studies above and, without performing a systematic analysis, and notwithstanding the previous warning about comparing R^2 across studies, make the general observations that higher R^2 appears to be associated with studies with smaller sample sizes, lower data resolution, more contextual data, and that do not include heating or cooling.

3 Method

This section describes the datasets, data preparation and analysis methods used to address the research questions.

3.1 Datasets

This paper uses Edition 2 of the SERL observatory which contains data from almost 5,000 households who have consented for SERL to collect their smart meter data and to link to other datasets, including Energy Performance Certificate (EPC), Index of Multiple Deprivation 2019 (IMD) quintile and weather data, as well as an (optional) survey completed at sign up. The first participants were recruited during wave 1 in Autumn 2019 and the second tranche were recruited in wave 2 in August 2020 which broadened the sample to include the North of England and Scotland as described in [43–45]. The data used in this paper is drawn from the ~5,000 participants recruited in these two waves.

3.1.1 SERL smart meter data

Half-hourly and daily[2] electricity and gas readings are collected via the DCC gateway[3] [46,47] for all participants with an accessible gas (76%) and/or electricity (100%) smart meter. Historic data is stored on the smart meters, and this is collected for up to 12 months prior to recruitment date. The observations run from 19th August 2018 – 29th February 2020. The latest date of meter data was chosen to be sufficiently in advance of the start of the first COVID-19 lockdown period in GB (23rd March 2020) for the observations to be reasonably unaffected by the pandemic. The data documentation describes the quality of the SERL smart meter data in detail [10] and data quality processes were applied before conducting the analysis for this research, as described below.

3.1.2 SERL survey

The SERL survey consists of 40 questions covering physical dwelling characteristics, household and respondent sociodemographic characteristics, energy use and heating behaviour, environmental attitudes, and appliance ownership. A copy of the survey is available in the documentation [10]. The aims of the survey were to collect contextual data to enable the production of nationally representative estimates, allow the creation and comparison of matched samples, and to help explain the variability of energy demand in the sample based on variables representing factors which existing research indicates are likely to influence household energy consumption (see literature review above), while also being reliably self-completed by a member of the public in about 10 minutes. Questions were designed in consultation with SERL consortium partners and Ipsos MORI

---

[2] Note that only SMETS2 meters record daily readings, but all record half-hourly.

[3] The Smart Data Communications Company (DCC) is the central communications infrastructure for the GB smart meter network.

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

and, where possible, were harmonised with national surveys such as the English Housing Survey, the 2011 Census and Understanding Society. Survey data are available for 4,753 (Edition 2) participants.

3.1.3 Energy performance certificate (EPC)

Energy performance certificates (EPCs) are EU-mandated ratings of domestic building energy performance which aim to rate a building's energy performance to enable comparisons of buildings energy use independent of occupant behaviour [48]. An EPC assessment is required by law when properties are sold or let in England and Wales. Address-level EPC data is publicly available [49], along with a description of variables (which include descriptions and energy efficiency estimates for building components, heating and lighting), and approximately half of the participants have an EPC[4]. At present, EPCs are not available for SERL participants in Scotland. While many dwelling-characteristic variables are available, it should be noted that there are measurement uncertainties associated with EPCs [50] e.g. due to surveyor error, or inaccuracies due to age of EPC and not reflecting subsequent retro-fitted measures. We note that limiting analysis to those households with an EPC could be a source of bias, as buildings which have not been sold or let since EPCs were introduced (in 2008) will not appear in the sample.

3.1.4 Weather data (ECMWF ERA5 reanalysis)

Weather data linked to the SERL observatory households is sourced from the Copernicus ERA5 reanalysis of the ECMWF (European Centre for Medium-Range Weather Forecasts) global weather data [51]. This combines observations and modelled data to produce a global, complete, and consistent dataset. The data are available hourly on a grid with spatial resolution of approximately 28 sq. km. The SERL observatory provides over 20 variables relating to temperature, wind, irradiance, precipitation, and humidity conditions for participant grid cells. The ERA5 website gives full details of the data, and details of the data available through the SERL observatory are provided in [10]. The present analysis made use of two weather variables: air temperature 2m above the surface (°C) and global horizontal irradiance reaching the surface (MJ/m^2 per day).

3.2 Data preparation

To avoid the influence of coronavirus lockdowns, this analysis used smart meter data from 19th August 2018 to 29th February 2020. The number of households with smart meter data increases over this period due to the second recruitment wave in 2020 and the lack of historical data for some households (up to 12 months before sign-up depending on move in date and meter installation date[5]).

The following criteria were applied to the initial sample of approximately 5000 households, such that households were excluded if any of the following applied:

- More than five questions with missing data for the SERL survey data (those with five or less had this missing data imputed, see below).
- Gas and electricity data did not record most of the energy used in the home (any of the following):
    - Solar thermal hot water heating or solar PV reported in the survey or EPC data, or electricity export readings in the smart meter data (indicates presence of solar PV). This will bias the sample away from buildings that tend to have solar PV e.g. more recent, larger, more likely to be detached, as well as households that are more likely to have retro-fitted energy efficiency technologies [52].

[4] SERL retains the most recent version.

[5] Second-generation GB smart meters ("SMETS2") can retain up to 13 months historic half-hourly consumption data.

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

    - Any form of central heating other than gas or electric (for example an oil boiler) reported in the survey. A consequence of this will be to bias the sample away from rural households where oil is more commonly used.
    - Gas heating reported in the survey or EPC but no gas smart meter data available.
    - Electric vehicle reported in the survey, since we are only concerned in this paper with energy use *within* the home. This will bias the sample away from the wealthy, middle-aged, male, well-educated, and affluent [53].
    - Buildings of multiple occupancy (not 'self-contained' in the survey) because the smart meter data relates to occupants not considered in the survey.
- Insufficient valid smart meter data available (see missing data below)
- Ages of adult occupants not self-reported in the survey as this precludes the calculation of the average age of adult occupants.

The above criteria produced a first sample of participants used in the following analysis. A second, smaller sample was also produced which had the additional criteria of requiring EPC data. We refer to the former as the 'larger' or 'Basic plus SERL survey' sample, and the latter as the 'smaller' or 'All data' sample. To be clear, the smaller sample is a subset of the larger sample.

These two samples allow the analysis of the impact of increasing contextual data availability across all contextual datasets for the 'all data' sample, as well as analysis of the impact of increasing sample size by comparing the results for equal levels of contextual data across the smaller and larger samples.

Table 2 shows the number of households excluded at each stage. The 65% drop due to insufficient data can be attributed to participants having smart meters installed close to SERL recruitment and therefore not having smart meter data for the analysis period. This will not be an issue in future editions of the SERL data but we expect the level of survey non-response and EPC absence to remain roughly constant. The exclusion rates shown in Table 2 are therefore the worst case we anticipate.

Table 2. Sample size after the application of successive exclusion criteria

Exclusion criteria	**Households excluded in each step**	**Sample size remaining**	Used in analysis?
Initial sample of households with smart meter data		4716	No
Excluding dwellings with insufficient data and where not all energy use in the home was recorded by smart meters	3063 (65%)	1653	No
Excluding dwellings without sufficient SERL survey data	235 (14%)	1418	Yes ('Basic plus SERL survey' or 'larger' sample)
Excluding dwellings without EPC data	736 (52%)	682	Yes ('All data' or 'smaller' sample)

Both the SERL survey and EPC data contained categorical variables for which small categories of less than 10 were merged to avoid statistical disclosure. Where possible categories were merged with the most similar category, otherwise with the next smallest category.

Daily summaries were derived from hourly weather data for use in the regression models. To account for increased space heating in cold weather, hourly temperature data for each grid cell was transformed into heating degree days (*hdd*) using the method described by [54]. We used a UK standard base temperature of 15.5°C. The daily sum of the hourly solar radiation reaching a horizonal plane at the surface of Earth was also included in the models. This acts as a proxy for solar gains and day length. Future work will explore the use of different base temperatures and whether more sophisticated methods, perhaps making use of the hourly resolution of the weather data, could improve the models. The models contained continuous and categorical variables. The continuous variables were centred on the population mean to remove structural multicollinearity [55]. Similarly, the categorical variables were 'one-hot' encoded i.e. dummy encoded with the largest category used as the reference to reduce multicollinearity [56].

3.3 Imputation of missing data

The SERL Observatory is affected by missing data. It is important to address missing data where possible as excluding observations due to missing data can lead to biased results [57]. The following sections describe our approaches to dealing with missing smart meter data and SERL survey data.

3.3.1 Smart meter data

The smart meter documentation provided by Elam et al. [10] describes the conditions used to flag a read as valid (below a high threshold and in the correct units). In addition, we required valid read times (midnight for daily, on the hour/half-hour otherwise). Due to higher quality of half-hourly data overall, the sum of half-hourly readings was used where valid, otherwise daily reads were used.

The following approach was taken to determine a missing data 'threshold': the proportion of missing data for each participant that is considered acceptable. Participants with more missing data than the threshold are rejected from the analysis:

- Specify estimate of interest to the analysis and estimate its mean and standard deviation (σ): here we use daily energy consumption per participant averaged over a period of a month
- Specify a confidence interval: Here we use a default confidence interval of 95%
- Specify margin of error: Here we use a default 10% margin of error
- Perform a standard sample size calculation, and by extension required threshold for missing data, using the following formula: Calculation: $n = (Z^* \sigma/E)^2$, where n is the sample size, Z is the value from the standard normal distribution for the chosen confidence interval (95%) and E is the desired margin of error (10%).

The equation shows that higher sample sizes (i.e. less missing data) are required for smaller margins of error, or higher confidence intervals, or for variables with greater standard deviations.

A key issue with this equation, for the purposes of dealing with missing smart meter data, is that it applies to situations where the sample size is assumed to be small compared to the population e.g. a survey of the general population. For missing smart meter data, the population is small compared to the sample size. For example, the population might be a year or a month's data for a single household, and the proportion of missing data might be relatively high for example 10%-20%.

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

We are therefore operating in a 'small population context', where the sample represents >5% of the population. For such situations, the 'finite population correction factor' $fpc = \sqrt{(N-n)/(N-1)}$ can be used to adjust the calculations, where N is the population size (e.g. N=365 for a year), and n is the sample size (e.g. for 10% missing data in a year n = 365-36 = 329).

This allows the calculation of an adjusted sample size (n_a), that takes into account the finite population, $n_a = (n*N)/(n + (N-1))$.

We can then calculate the level of missing data (*missing*), this corresponds to $missing = (N - n_a)/N$.

For example, assuming the mean and standard deviation of daily gas consumption over a period of a month for a household was 21.99 kWh and 10.61 kWh respectively produces a missing data tolerance of up to 24.9% missing data in a month of daily gas consumption data for a household, and confidence that 95 times out of 100 the resulting estimate will be within 10% of the true value. Varying the margin of error (or indeed confidence interval) will result in different required thresholds, as illustrated in Figure 1, and the required threshold will vary depending on the variable to be estimated. Variables which do not vary much will tolerate greater levels of missing data than those which vary more.

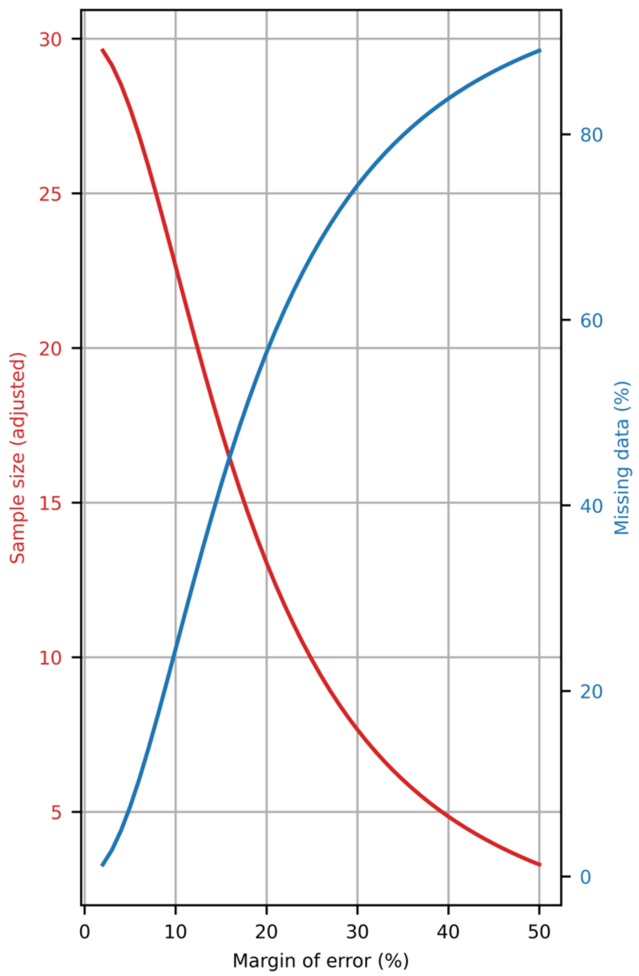


Figure 1. Effect of varying margin of error on required sample size (left) or missing data (right) for estimating average monthly daily gas demand per participant.

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

The above approach implicitly assumes that data is missing at random, and that the missing data has the same statistical characteristics as the data that is not missing. It means that we assume that the data that is not missing is representative of the full population.

From an imputation perspective, this is equivalent to filling the gaps created by the missing data with the estimated variable calculated from the non-missing data. Using the example given above, it is equivalent to filling the missing data with the mean daily consumption for each month and household and using the resulting data to calculate the mean for the month.

There are a variety of ways of imputing missing smart meter data found in the literature with varying complexity. The benefit of the chosen approach is that it is parsimonious by only (implicitly) imputing observations to the extent that they are required to produce the results, while not altering the underlying data. The disadvantage is that it assumes data is missing at random which, while a defensible starting assumption, might not be true in practice.

The missing data threshold was calculated separately for monthly average daily gas and electricity demand using data from 19 August 2018 to 29 February 2020. The missing data thresholds were 94% for electricity data and 72% for gas data.

3.3.2 SERL survey

Approximately 20% of participants in the SERL Edition 2 dataset have at least some survey data missing, see Table 3. Although simple list-wise deletion of cases with missing data is a common approach, Austin [58] note that this approach is potentially problematic. For example, if data is not missing completely at random (MCAR) then this may introduce bias in parameter estimations, and if data are MCAR then the reduction in sample size will reduce precision and increase confidence intervals in parameter estimation. Kang [57] discusses several approaches to imputing missing data and their limitations, concluding that multiple imputation is often an appropriate technique. We imputed missing values if the number of missing survey answers was 4 or fewer. We chose this as a threshold as it is a relatively low value so that imputed values are associated with reduced uncertainty, that the missing data might reasonably be the result of respondent oversight (rather than drop-out), and which was associated with most of the surveys with missing data.We note however that this is a relatively arbitrary choice of threshold and future research may wish to investigate the effect of varying threshold on uncertainty in imputed values. Table 3 shows that this imputation increased the survey sample by 12% and only 7.4% were rejected because they had too little survey data.

Multiple imputation involves imputing multiple values for the missing data based on the available data to generate multiple 'complete data sets' which are then used in the subsequent analysis [59]. The spread in the imputed values reflect the uncertainty over the missing values, then repeating the analysis with each of the imputed datasets indicates how much uncertainty in the results is due to the uncertainty in the imputed values. For this work we used a Multiple Imputation by Chained Equations (MICE) algorithm, see for example Austin [58] for details. We repeated the imputation 5 times, to give 5 versions of the survey data with no missing values (having filtered out those surveys which originally had more than 4 missing values). In the following analysis, we present the results for one of the imputed datasets. We repeated the analysis for the remaining 4 versions of the imputation and there were no notable differences in the results, suggesting the imputed values did cause significant uncertainty in the results.

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

Number of missing values in survey	Number (%) of survey responses
0	3713 (80.6%)
1 - 4	554 (12.0%)
More than 4	342 (7.4%)

Table 3. Number of survey responses with different amounts of missing data.

3.4 Sample representativeness

The SERL Observatory sample was designed to be representative of households in GB with a DCC-enrolled smart meter (see [47] and [60] for further details), but response bias, the exclusion of the final recruitment wave and the application of the above exclusion criteria will result in biased final analytic samples. Our results should not be taken as generalisable to the SERL Observatory as a whole, nor to the broader GB population. Future work will explore the use of larger and weighted samples to enable results that are more generalisable. Table 4 compares the regional distribution of the samples compared to the population of England and Wales, showing that in particular they under-represent the North of England and Scotland (due to delayed smart meter rollout) and areas with greater deprivation (low IMD quintiles).

Description	Response	Sample	Frequency (N)	Frequency (%)	Population Percentage[6]
Region	EAST MIDLANDS	Larger (N=1418)	141	9.90%	7.3%
Region	EAST MIDLANDS	Smaller (N=682)	69	10.10%	7.3%
Region	EAST OF ENGLAND	Larger (N=1418)	155	10.90%	9.3%
Region	EAST OF ENGLAND	Smaller (N=682)	81	11.90%	9.3%
Region	GREATER LONDON	Larger (N=1418)	211	14.90%	13.3%
Region	GREATER LONDON	Smaller (N=682)	118	17.30%	13.3%
Region	NORTH WEST	Larger (N=1418)	120	8.50%	11.5%
Region	NORTH WEST	Smaller (N=682)	51	7.50%	11.5%
Region	SCOTLAND	Larger (N=1418)	71	5.00%	9.1%
Region	SCOTLAND	Smaller (N=682)	0	0%	9.1%
Region	SOUTH EAST	Larger (N=1418)	271	19.10%	13.7%
Region	SOUTH EAST	Smaller (N=682)	133	19.50%	13.7%
Region	SOUTH WEST	Larger (N=1418)	157	11.10%	9.1%

[6] Calculated from ONS AddressBase.

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

Region	SOUTH WEST	Smaller (N=682)	85	12.50%	9.1%
Region	WALES	Larger (N=1418)	83	5.90%	5.5%
Region	WALES	Smaller (N=682)	45	6.60%	5.5%
Region	WEST MIDLANDS	Larger (N=1418)	136	9.60%	8.7%
Region	WEST MIDLANDS	Smaller (N=682)	64	9.40%	8.7%
Region	YORKSHIRE AND NORTH EAST	Larger (N=1418)	73	5.10%	12.6%
Region	YORKSHIRE AND NORTH EAST	Smaller (N=682)	36	5.30%	12.6%
IMD quintile	1	Larger (N=1418)	207	14.60%	20.5%
IMD quintile	1	Smaller (N=682)	101	14.80%	20.5%
IMD quintile	2	Larger (N=1418)	272	19.20%	21.0%
IMD quintile	2	Smaller (N=682)	125	18.30%	21.0%
IMD quintile	3	Larger (N=1418)	299	21.10%	20.6%
IMD quintile	3	Smaller (N=682)	153	22.40%	20.6%
IMD quintile	4	Larger (N=1418)	315	22.20%	19.7%
IMD quintile	4	Smaller (N=682)	152	22.30%	19.7%
IMD quintile	5	Larger (N=1418)	325	22.90%	18.2%
IMD quintile	5	Smaller (N=682)	151	22.10%	18.2%

Table 4. Regional representation of the dwellings in the samples used for analysis.

Table 5 and Table 6 compare some key characteristics of the sample with the population in England using data from the English Housing Survey 2018-2019 [61]. The samples under-represent flats and rental tenures, and this is worse for the larger sample. The smaller sample is comparable to the national average in terms of size of dwelling and household and building energy efficiency rating (SAP), but the larger sample has smaller household size (the other measures cannot be calculated as the larger sample does not have EPC data which is used for their calculation).

Characteristic	**Category**	**SERL Sample**	**SERL Sample number**	**SERL Sample Proportion**	**EHS Population Proportion (England)**[7]
Built form	Detached	Larger (N=1418)	415	29.3%	26.1%
Built form	Detached	Smaller (N=682)	178	26.1%	26.1%

[7] Population proportions are for England for 2018-2019 and are taken from [61].

Built form	Semi-detached	Larger (N=1418)	437	30.8%	25.4%
Built form	Semi-detached	Smaller (N=682)	204	29.9%	25.4%
Built form	Terraced	Larger (N=1418)	392	27.6%	28.4%
Built form	Terraced	Smaller (N=682)	196	28.7%	28.4%
Built form	Purpose built flat	Larger (N=1418)	144	10.2%	16.5%
Built form	Purpose built flat	Smaller (N=682)	87	12.8%	16.5%
Built form	Converted house or commercial building	Larger (N=1418)	30	2.1%	3.6%
Built form	Converted house or commercial building	Smaller (N=682)	17	2.5%	3.6%
Tenure	Own / part-own	Larger (N=1418)	1222	86.2%	63.3%
Tenure	Own / part-own	Smaller (N=682)	561	82.30%	63.3%
Tenure	Private rental	Larger (N=1418)	90	6.3%	19.9%
Tenure	Private rental	Smaller (N=682)	72	10.60%	19.9%
Tenure	Social rental / rent free	Larger (N=1418)	106	7.5%	16.8%
Tenure	Social rental / rent free	Smaller (N=682)	49	7.20%	16.8%

Table 5. Key characteristics of the dwellings in the sample used for analysis.

Characteristic	**Category**	**SERL Sample**	**SERL Sample mean**	**EHS Population mean (England)[8]**
Household size	Number of persons per household	Larger (N=1418)	2.29	2.39
Household size	Number of persons per household	Smaller (N=682)	2.40	2.39
Building energy efficiency	SAP rating	Larger (N=1418)	n/a	63.2

[8] Population proportions are for England for 2018-2019 and are taken from [61].

Building energy efficiency	SAP rating	Smaller (N=682)	62.2	63.2
Size of dwelling	Floor area (m^2)	Larger (N=1418)	n/a	94
Size of dwelling	Floor area (m^2)	Smaller (N=682)	97.5	94

Table 6. Further key characteristics of the dwellings in the sample used for analysis.

Table 7 shows key statistics regarding the energy consumption of the dwellings in the samples, and the degree days during the period of analysis. For comparison, in 2019 the mean UK daily domestic consumption was 31.56 kWh/day for gas (for households connected to the gas grid) and 10.22 kWh/day for electricity [62], mean gas use for the samples is higher and lower for electricity. The higher gas use is consistent with larger properties (less flats) and wealthier occupants. Note our sample is drawn from Great Britain not UK (i.e. no dwellings from Northern Ireland). Also note that (as to be expected) the distributions of energy variables are highly skewed, hence the large relative standard deviations.

	Sample	**Mean (SD)**	**1st quartile**	**Median**	**3rd quartile**
Total daily household energy consumption (kWh)	Larger (N=1418)	49.25 (42.97)	15.7	38.7	71.53
Total daily household energy consumption (kWh)	Smaller (N=682)	46.92 (42.19)	14.47	35.75	68.17
Daily gas consumption (kWh)	Larger (N=1418)	41.09 (39.89)	8.65	31.97	62.06
Daily gas consumption (kWh)	Smaller (N=682)	39.26 (39.18)	7.73	29.67	59.55
Daily electricity consumption (kWh)	Larger (N=1418)	9.7 (8.62)	4.84	7.69	11.9
Daily electricity consumption (kWh)	Smaller (N=682)	9.41 (7.91)	4.62	7.53	11.7
Daily mean external temperature (°C)	Larger (N=1418)	10.32	6.62	9.43	14.05
Daily mean external temperature (°C)	Smaller (N=682)	10.44	6.71	9.55	14.2
Heating degree days (°C per day)	Larger (N=1418)	5.73	1.98	6.08	8.88
Heating degree days (°C per day)	Smaller (N=682)	5.63	1.87	5.95	8.79

Table 7. Energy consumption and temperature statistics for the sample used for analysis.

3.5 Statistical analysis

3.5.1 Analytic design

As noted in the method section, linear mixed effects models are an appropriate method for longitudinal panel data, as this allows the structure of the data (repeated observations for the same dwelling) to be explicitly accounted for in the model [63]. Coupled with this, linear mixed models are

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

relatively straightforward to interpret and for these reasons this method was selected for this work. As one of the objectives of this paper is to assess the explanatory power of the SERL contextual datasets separately and in combination, variable subset selection is not implemented. We acknowledge that the inclusion of large numbers of variables without theoretically or model-driven selection runs the risk of potentially spurious effects and increases the temptation to 'fish' for 'p-values'. However, the need to assess the associations between the measured variables, and thus inform both future analysis and future data collection, led us to cautiously proceed with a larger than normal set of explanatory variables.

To investigate how the explanatory power of the model and results for individual coefficients change given different levels of contextual data and sample size, linear mixed models of daily total (gas + electricity) energy consumption per participant were fitted using four levels of contextual data (where applicable) on two samples of different sizes (the smaller being a subsample of the larger). The first level of contextual data is 'basic data' consisting of: region, IMD quintile, day of the week, bank holiday indicator, heating degree days and solar radiation. These are widely available, area-based variables that are easily linked to smart meter data and available for all participants. Additional models were developed with further levels of added contextual data: from the SERL survey and EPC data separately and then in combination. This results in the following 'levels' of contextual data which, due to missing data, result in smaller sample sizes:

1. Basic data only
2. Basic plus SERL survey data
3. Basic plus EPC data
4. All data

The first two can be applied to both samples, while the last two can only be applied to the smaller sample (as not every participant in the larger sample has EPC data, while all do in the smaller sample).

3.5.2 Statistical model

We note that log-transforming of the dependent variable is sometimes performed in previous studies to address heteroscedasticity [20] or symmetry in residuals [8]. We do not log-transform as the residuals of the model are normally distributed (an example plot of the residuals for one model is provided in Supplementary Data) and not strongly affected by heteroscedasticity, and log-transforming has the adverse effect of producing residuals which are not normally distributed.

To take advantage of the longitudinal (repeated measures) nature of the dataset we applied a random effects (RE) approach, similar to that used by Anderson et al. [28]. We use both random intercepts and random slopes applied to the heating degree day (*hdd*) variable; this allows each individual dwelling to deviate from the group mean intercept and gradient. The thermal performance of each dwelling will strongly affect the gradient of the *hdd* variable and the random slope component allows this to deviate from the mean for each participant. Every variable is included by itself as well as interacting with the *hdd* slope variable. Following Snijders and Bosker [42], the random slope model with all contextual data therefore takes the form:

$$Y_{ti} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0}\, x_{pti} + \sum_{q=1}^{Q} (\gamma_{0q}\, z_{qi} + \gamma_{1q} z_{qi} hdd_{ti}) + U_{0i} + U_{1i} hdd_{ti} + R_{ti}$$

Equation 1

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

$Y_{t,i}$ is the energy consumption of dwelling *i* at time period *t*. The first part of the equation with γ coefficients is the *fixed part* (because the coefficients are fixed i.e. non-stochastic), while the remainder is the *random part* of the model, comprising 'level two' residuals (i.e. participant-level) random intercept U_{0i} and random *hdd* slope U_{1i} for each participant *i*, and 'level one' residual (i.e. measurement-level) error R_{ti}. It is assumed that level one and two residuals have mean 0, and that the pair of level two residuals, and the level one residual, are independent and identically distributed.

The fixed part includes the intercept for the average participant γ_{00}; regression coefficients γ_{p0} associated with *P* measurement-level variables x_{pti} i.e. those that change for each participant *i* and each time step *t* (heating degree days and solar radiation); and regression coefficients associated with *Q* participant-level variables z_{qi}, consisting of all other variables, all of which are also interacted with the heating degree day variable *hdd*. γ_{1q} is the *hdd* slope for the average participant. A full list of the variables from each dataset used in the regressions is given in Supplementary Data.

EPC variables related to cost, carbon dioxide emissions and environmental efficiency were excluded. Text descriptions of building elements such as type of external wall were also excluded as a categorisation of the element's thermal performance was included instead[9]. All SERL survey variables were included except for those relating to the respondent ('About you' section)[10] as the unit of analysis of interest here is the household, not the respondent. An interaction term between solar radiation and heating degree days was included in all models since solar gains can provide space heating during the heating season.

3.5.3 Statistical tests

5-fold cross-validation [64] is used to compute training and testing statistics of root mean squared error (RMSE) and R^2. As our model is multi-level, with some variables relating to between-household variation (e.g. SERL survey and EPC data) and others relating to within-household variation (e.g. weather data), we construct two levels of cross-validation: a 'level 1' within-household cross-validation where daily consumption observations are held out for the test 'fold' but each fold includes data from each participant. And a 'level 2' between-household cross-validation where households (and all their contextual data) are held out for the test fold. The former tests for within-sample prediction errors, and is relevant where counterfactual demand for a sample is to be predicted and compared with actual consumption. Applications include estimating the impact of the coronavirus pandemic or energy efficiency interventions on a sample of households' energy consumption. While the latter between-house cross-validation tests for out-of-sample errors, and is relevant for predicting energy consumption of other households that are not included in the sample used to train the model. Applications include estimating energy statistics, or providing energy efficiency advice to households.

Further, note that as our samples consist of grouped measurements that are unbalanced (i.e. unequal number of measurements per household), we ensured that each level 1 cross-validation fold was approximately equally unbalanced. In other words, each fold contains approximately the same number of daily consumption observations for each household.

Our model contains random effects for each participant. We have not included random effects when computing training or testing statistics and the errors are therefore associated with the fixed effects i.e the various levels and combinations of contextual data. Because we omit random effects, we do

[9] The exception was that the descriptive variable for secondary heating was included as the variable describing its energy efficiency had no data.

[10] Note the 'managing financially' question is included.

not calculate the marginal, conditional and residual R^2 developed by Nakagawa and Sheilzeth [65] for linear mixed effects models. We instead compute the conventional R^2 as in [66].

Additionally, we calculate an adjusted R^2 as we note that the number of independent variables is moderately large compared to the number of households in our sample (the parameter/household ratio is 356/682 = 0.52 for the All Data model and smaller sample). We note that calculating adjusted R^2 for is not well documented in statistical textbooks on multi-level mixed effects models. We have therefore used number of groups (participants) as the number of observations in the adjustment factor, though we note this is a conservative estimate of this statistic, given each group can itself contain several hundred observations.

All of this results in a relatively large number of combinations of sample, method, model, and cross-validation level. We provide a summary of all the combinations tested in Table 8.

	Basic Data only model	**Basic plus SERL Survey model**	**Basic plus EPC Data model**	**All Data model**
Linear mixed effects	Yes	Yes	Yes	Yes
Level 1 cross-validation	Yes	Yes	Yes	Yes
Level 2 cross-validation	Yes	Yes	Yes	Yes
Smaller sample (N=682)	Yes	Yes	Yes	Yes
Larger sample (N=1418)	Yes	Yes	No	No

Table 8. Summary of the different combinations of sample, model, contextual data and cross-validation included in analysis.

We use *p*-values to evaluate statistical significance of independent regression variables and take $p<0.05$ as our statistical significance threshold noting that statistical significance does not necessarily imply substantive significance. We note that we do not use variable selection and we include many explanatory variables in our model which means that, given the large number of covariates used in some of the models, it is highly probable that some of the variables found to be significant will indeed be spurious results.

Models with larger numbers of independent variables can spuriously appear to fit data better than a smaller nested model. Therefore, as a further test of statistical significance between the models, we perform an F-test to compare models with their smaller, 'nested', or 'restricted' counterparts using the same underlying data, where applicable. In this case, the F-test gives the probability that the simpler (nested) model provides a better fit to the data, or rather that any improvements in fit associated with the larger unrestricted model are spurious. For example, the 'All data model' is compared with the 'Basic plus SERL survey' model and 'Basic plus EPC data' model. While both the 'Basic plus SERL survey' and 'Basic plus EPC data' models are compared with the 'Basic data only' model. For each, F-statistics are calculated as in [39].

We also note the population for which the sample is intended to be representative of is not the SERL Observatory nor GB population. It is a biased sample affected by sample design, recruitment strategy, non-response bias, and exclusion criteria (see above for descriptive statistics of the sample that indicate biases).

3.5.4 Multicollinearity

Many explanatory variables were included in this analysis and there is multi-collinearity in the original regressors (prior to transformation). An obvious example is that both the SERL survey and the EPC data include categorical variables relating to the age of the building. However, many other variables are also affected by collinearity, a common issue in similar energy demand research [8].

The effect of multicollinearity is to reduce the accuracy of the estimates of the regression coefficients [30] and thereby reducing the probability of correctly detecting a significant coefficient, and to make the coefficient of collinear variables sensitive to changes in the input data. This makes interpretation of the model's results challenging, though multicollinearity does not affect the model's goodness of fit (R^2) or (within-sample) predictive accuracy.

Multicollinearity is commonly assessed using the variance inflation factor (VIF). This showed high levels of multi-collinearity for our initial data samples prior to transformation (e.g. 98 variables with VIF > 10). Correcting multicollinearity for categorical variables can involve removing, combining, or transforming variables. Removing variables is unwelcome as we want to assess the explanatory power of the dataset in total. Nonetheless an EPC variable which indicates whether a flat was top storey or not (*flatStorey*) was also excluded as its inclusion caused the model to fail to converge for one of the training sets. We believe this is because this variable is highly collinear with other variables in particular *propertyType*, which indicates if a property is a flat, and *roofEnergyEfficiency* which is 'n/a' for flats which are not top storey. To reduce multicollinearity, we therefore use combination and transformation. Continuous variables were centred (the population mean was subtracted); categorical variables dummy encoded so that the reference category was the largest, and some categorical variables were combined (total number of rooms; central heating fuel type). Centred variables are denoted with the suffix *_c*. The exception to this is that heating degree day *hdd* was not centred for the basic data model. While the resulting VIF for this model are high, this was to retain one model where the intercept and slope parameters were more intuitive i.e. where the intercept is when *hdd* equals zero.

3.5.5 Software

All analysis was performed within the UCL Data Safe Haven using Python (version 3.8), spyder (version 4.1.4), pandas (version 1.0.5) [67,68], and statsmodels (version 0.11.1) [69] for the linear mixed effects regressions. The code used to perform the analysis of SERL data presented in this paper will be made available on GitHub[11].

4 Results and discussion

In the following subsections we present results which illustrate the main findings from the above analysis, full results from all models are provided in the supplementary data.

4.1 Cross-validation test statistics

As described above, we performed two types of cross-validation: 'level 1' in which all households are represented in all cross-validation folds but approximately 1/5 of daily readings from each household are held-out, and 'level 2' in which approximately 1/5 of households are held out in each fold. This section considers cross-validation using the smaller sample (N=682) only, while the following section compares this to cross-validation for the larger sample (N=1418).

Results of the level 1 cross validation are summarised in Figure 2 (full results for the cross-validation are provided in the supplementary data). The Figure shows that training and testing RMSEs and R^2

[11] https://github.com/smartEnergyResearchLab/

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

are very similar for all models with level 1 cross-validation. The lack of discrepancy between training and testing errors indicates the models are not over-fitting, and we can be confident that the errors in our training data are good estimates of the expected error in predicted values. This suggests that the models are suitable for predicting energy consumption when historic consumption data from the dwellings is available to train the models, and that models that use more of the available contextual data are more accurate than those that use less.

We noted above (and discuss in further detail in section 4.3.1) that several of the models are adversely affected by high levels of multicollinearity, however the results confirm that while multicollinearity may be a concern when models are used for inference, it does not affect the validity of using these models for within-sample prediction, as demonstrated here by the small differences between training and testing errors.

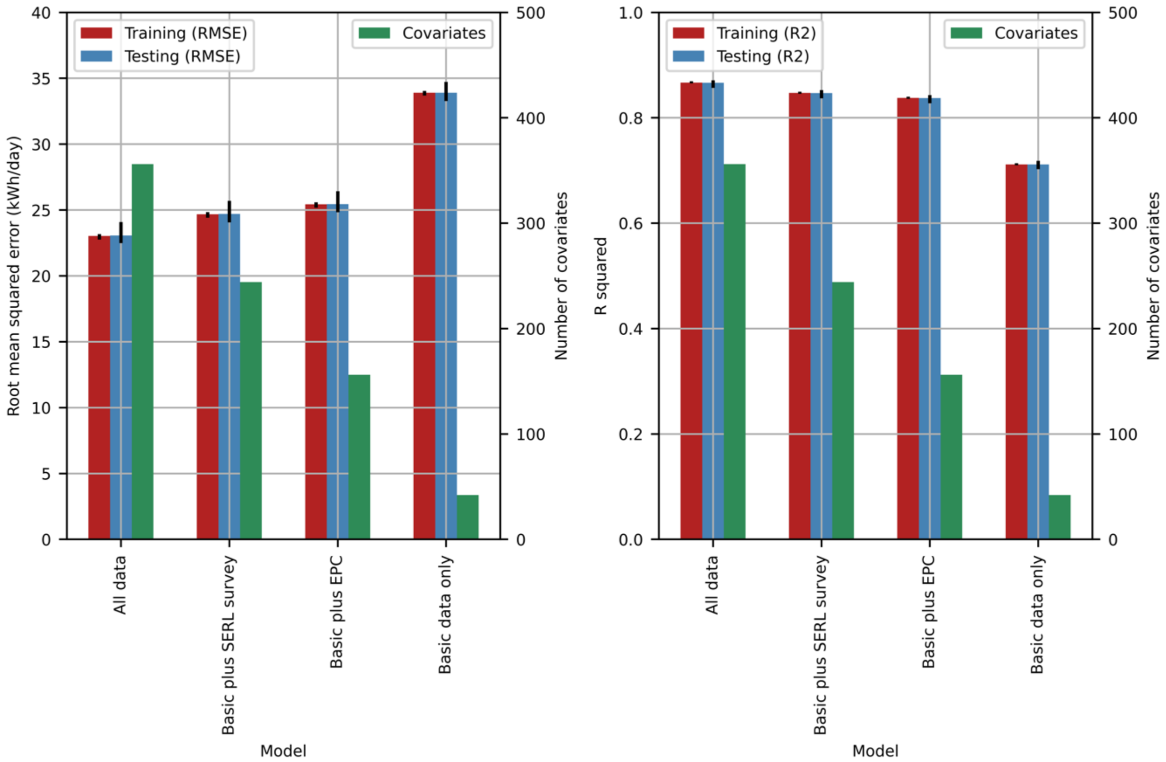


Figure 2. Mean training and testing errors and R^2 for level 1 within-group cross-validation prediction using smaller sample (N=682). Error bars indicate the range of values from the cross-validation folds.

Figure 2 shows that there are diminishing returns (in terms of reducing RMSE and increasing R^2) when increasing the number of variables in the model. Although the All Data model has the highest R^2 and lowest RMSE, its improvement over the survey only or EPC only models is modest. Indeed, Table 9 shows that the All Data model fails the F-test (p=1) compared to the more restricted models. This indicates the following:

1. The increase in R^2 and decrease in RMSE for the All Data model is not significant. Although it gives lowest RMSEs, the improvement over the more restricted models is not enough that it could not be down to chance, given the substantial increase in explanatory variables used in the model.
2. That the All Data model is not correctly specified and that it is unlikely that the model gives generalisable relationships between the explanatory variables and daily energy

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

consumption. Note that this does not mean that the restricted models are themselves correctly specified, just that compared to them the All Data model is worse.

Despite this, the All Data model is the most accurate for within group prediction. This indicates that this is an application which is relatively robust to overfitting and to which relatively complex models are well-suited. It indicates that at least some of the difference between within-sample modelled and predicted energy consumption is due to model bias error. These models are simple linear models which are unlikely to fully reflect the complex reality of domestic energy consumption. Future work to reduce the error associated with predicting energy consumption from dwellings with historic data should benefit from exploring more complex non-linear models, such as artificial neural networks, which we note have proven to be highly popular for this purpose but which, to the authors' knowledge, have rarely if ever been applied to data comparable to that of the SERL Observatory.

Both the Basic plus EPC data and Basic plus SERL Survey models have very low p-values (<0.01) when compared against the more restricted Basic Data Only model and so pass the F-test, indicating that their improvement in accuracy is not likely to be due to chance. Again, we note that this this does not mean that these models are necessarily correctly specified.

Cross-validation	Sample	Unrestricted model	Restricted model	F-statistic (mean)	p-value (mean)
Level 1	Smaller (N=682)	All data	Basic plus SERL survey	0.43	p=1
Level 2	Smaller (N=682)	All data	Basic plus SERL survey	0.27	p=1
Level 1	Smaller (N=682)	All data	Basic plus EPC data	0.36	p=1
Level 2	Smaller (N=682)	All data	Basic plus EPC data	0.24	p=1
Level 1	Smaller (N=682)	Basic plus EPC data	Basic data only	3.58	p<0.001
Level 2	Smaller (N=682)	Basic plus EPC data	Basic data only	2.71	p<0.001
Level 1	Smaller (N=682)	Basic plus SERL survey	Basic data only	1.92	p<0.001
Level 2	Smaller (N=682)	Basic plus SERL survey	Basic data only	1.38	p<0.01
Level 1	Larger (N=1418)	Basic plus SERL survey	Basic data only	3.78	p<0.001
Level 2	Larger (N=1418)	Basic plus SERL survey	Basic data only	2.92	p<0.001

Table 9. Summary of F-tests evaluated using cross-validation training errors. Level 1 cross-validation refers to within-group predictions, level 2 refers to between-group predictions.

Turning to the level 2 between-group cross-validation results, the training and testing RMSE and R^2 values are notably different from each other, as shown by Figure 3. Training RMSEs are similar in magnitude and decrease as the number of explanatory variables increase. Contrary to the level 1 results however the testing RMSEs are larger than training RMSEs. This indicates that the models are

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

less accurate at between-group prediction than within-group prediction. The level 2 testing errors however show a different trend to level 1: they are higher than training errors, and the discrepancy increases with the number of covariates. The Basic Data only model shows the highest testing error, while the others have similar testing errors, with the Basic plus EPC Data slightly outperforming the others. The discrepancy between training and testing error indicates that the models are over-fitting, and that this gets worse as the number of variables is increased. The F-test results for level 2 cross-validation (Table 9) are similar to those for level 1, with the All Data model failing the test and Basic plus SERL survey and Basic plus EPC data models passing.

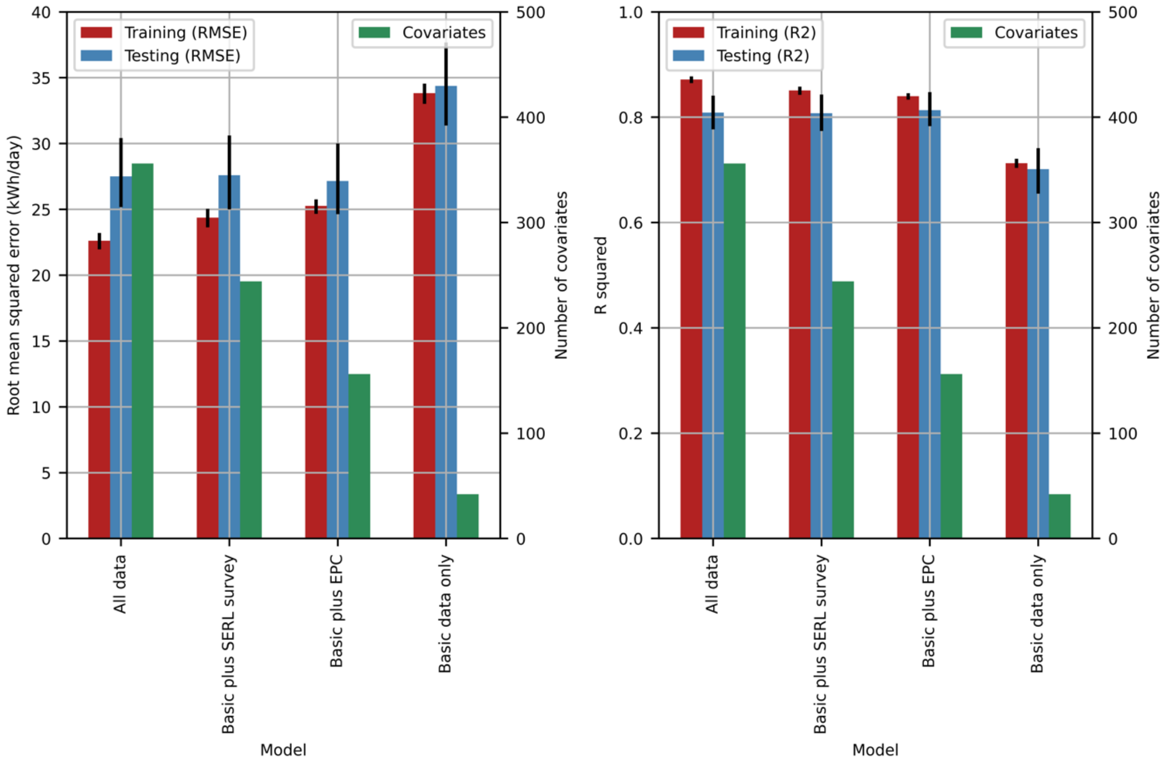


Figure 3. Mean training and testing errors and R^2 for level 2 between-group cross-validation predictions using smaller sample (N=682). Error bars indicate the range of values from the cross-validation folds.

These results suggest that for between-group prediction (i.e. for prediction of energy consumption where there is no historic data), models with a selected, intermediary number of explanatory variables between 'basic' and 'all data' will likely perform best. In this work we take the admittedly rudimentary approach of selecting variables according to their presence in different combinations of contextual data available to be linked in the SERL Observatory. The results show that for this type of prediction, this type of variable selection, and this sample, the model using Basic plus EPC data marginally outperforms the others, while having less of a discrepancy between training and testing errors, indicating it is less affected by overfitting. That being said, the Basic plus SERL Survey model has the advantages that almost all participants have survey data, whereas only approximately 50% have EPC data, and it is less affected by multicollinearity than the Basic plus EPC data model.

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

4.2 Sample size effects

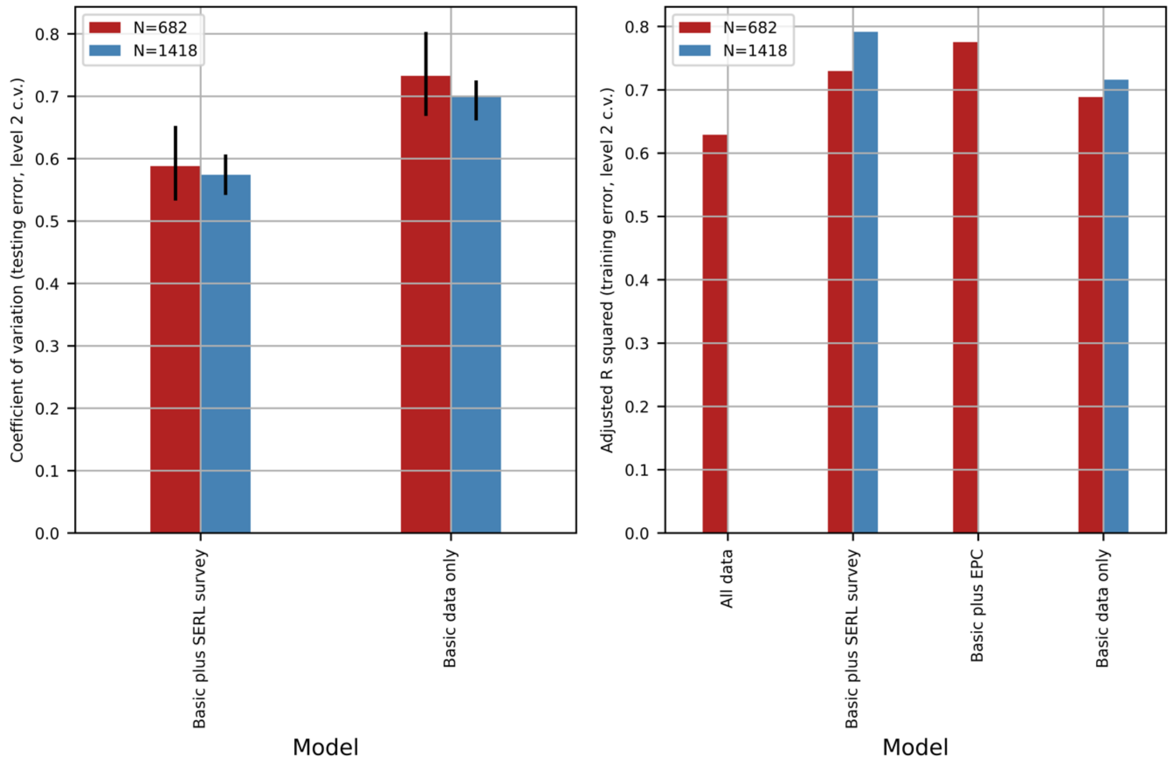


Figure 4. Comparison of level 2 between-group cross-validation testing errors (coefficient of variation, left) and adjusted R^2 calculated from training errors (right) for different models and sample sizes. Error bars indicate range across the cross-validation folds.

Figure 2 and Figure 3 present unadjusted R^2 values which can be affected by inflation if models include many independent variables. Indeed, the results show R^2 values increasing as with number of covariates in the models used here. When R^2 values are adjusted, as shown in Figure 4 for training errors, the models with more variables is penalised. Indeed, the All Data model performs worse overall, even worse than the Basic Data only model, and the Basic plus EPC data model performing best. This is because the number sample size (N=682) is not large relative to the number of variables used in the models.

When the sample size is almost doubled to the larger sample (N=1418) for the Basic plus SERL Survey and Basic Data only models, these models perform better. Figure 4 compares the CVRMSE and adjusted R^2 for these models trained and tested with the different sample sizes. While the testing RMSE slightly increased for the larger samples for the Basic Data only and Basic plus SERL Survey models and both cross-validation types, the mean of the daily energy consumption is larger for the larger sample (as shown in Table 7), so we compare the (testing) CVRMSE instead. The CVRMSE decreases and adjusted R^2 increases, indicating that the models performed better overall with the larger sample. We also see that the range of CVRMSE from each cross-validation fold decreases with increasing sample size, suggesting that the prediction error can be more accurately characterised. While the results indicate that increasing sample size has the benefit of improving accuracy the changes resulting from doubling the sample size are nonetheless modest.

4.3 Assessment of individual variables

4.3.1 Variance inflation factors

Model	Sample	No. covariates (not including intercept)	VIF (mean)	VIF >10	VIF >5

All Data	Smaller (N=682)	355	4.33	29	65
Basic plus EPC data	Smaller (N=682)	155	3.91	13	20
Basic plus SERL Survey	Smaller (N=682)	243	2.74	1	19
Basic plus SERL Survey	Larger (N=1418)	249	2.49	5	19
Basic Data only	Larger (N=1418)	43	5.70	1	28

Table 10. Summary of variance inflation factors for selected models.

Variance inflation factors were considerably reduced by the measures described in 3.5.4, though they remained high (both on average, and for individual variables) for some models (see Table 10). The coefficients of variables affected by collinearity (commonly taken to be where VIF is greater than 5 or 10) should be interpreted with extreme caution as they may be unstable (if a different data set is used) and have inflated p-values (i.e. reduced significance) due to systematic bias in the underlying standard errors. Note that the Basic Data only model has unusually high VIF considering its relatively small number of variables; this is because this model includes *hdd*, not *centred hdd*, and this indicates the importance of centring to reduce VIF.

The All Data model has the most variables with VIF above 5 or 10 and the second highest mean VIF after the Basic Data only model. The models using Basic Data only and Basic plus SERL Survey data had lowest mean VIF and, despite including relatively large numbers of covariates, had the fewest number of individual variables with VIF >5.

The following section therefore only presents results relating to individual variables for the Basic plus SERL Survey model applied to the two sample sizes. We note however that despite the relatively low VIF, these results for these models should be viewed with scepticism given the lack of variable selection, large number of variables and thus high probability of spurious results, and evidence of over-fitting. The results for the other models are not reported here but provided in the supplementary data because of the high VIFs and caution with which they should be interpreted. VIFs for all the models and their individual variables are reported in the Supplementary Data.

4.3.2 Basic and SERL survey model

We now return to our second research question and consider what individual variables observed in the SERL Observatory data explain household-level daily energy consumption. The following reports on the effect size and statistical significance of individual variables for the model including Basic plus SERL Survey data (i.e. not including EPC data), and compares the results for the two sample sizes. For clarity of presentation, variables are not included if they have VIF > 5 or p-value >= 0.05. Figure 5 and Figure 6 display the size and 95% confidence interval estimates of the coefficients with the ten largest positive and negative effects on intercept and *hdd* slope for the larger sample and smaller samples respectively. We acknowledge that filtering on *p*-value runs counter to the prevailing advice **not** to use *p*-values as binary thresholds but the presentation of confidence intervals with point estimates provides some mitigation. In addition, full regression results for all models are included in the Supplementary Data to enable detailed examination.

First, note that the confidence interval of many of the coefficient estimates is large, with lower edges approaching zero. This indicates the large uncertainty regarding the estimates and the high probability that the significance of estimates may be spurious, particularly given the large number of variables included in this model.

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

Overall, if we consider the size of the coefficients for the covariates for the larger sample, the variables associated with a larger intercept include presence of an air-conditioning unit (ACU) and the three oldest dwelling age bands (before 1900, 1900-1929 and 1930 to 1949), a warmer thermostat set point, and a more detached dwelling.

Answering 'not applicable' to whether changes to the heating or energy supply are being considered in the next 12 months, having electricity supply hot water for taps and considering installing a gas boiler are all also associated with an increase in intercept, however these three variables have large confidence intervals which are close to including zero.

Variables associated with a decreased intercept include having electric rather than gas heating, solar radiation, the two newest dwelling age bands (1991-2002 and 2003 onwards), dwellings in the East Midlands, Wales and North West, less opening of windows on warm days, making a great deal of effort to limit energy use, and having a cooler thermostat set point.

In terms of the effect on the *hdd* slope we see that, similar to the intercept, the three oldest age bands, detached dwellings, warmer heating set-points and presence of an ACU have an increased slope. Number of rooms, number of bedrooms and IMD 5 (least deprived) are also associated with an increased slope. Having no timing heating control is also associated with an increased slope, but this has confidence intervals very close to enclosing zero.

For the smaller sample broadly similar trends are observed, with older and more detached dwellings associated with larger intercepts and steeper slopes and vice versa for newer buildings and flats and terraces, solar radiation and making a great deal of effort to reduce energy consumption are again associated with decreased intercept and shallower slope. Variable use of standalone heaters are associated with decreased intercept and slope for the smaller sample although this was not significant in the larger sample.

Less switching off lights, always opening windows on cold days, less putting on clothes in cold weather and having electric and gas heating are associated with increased intercepts for the smaller sample, but none are significant in the larger sample.

Thermostat set-point question being 'not applicable' is associated with a reduced *hdd* slope for the smaller sample, but this has confidence intervals very close to including zero and is not significant for the larger sample.

Sundays, number of adult occupants, and IMD quintiles 4 and 5 (least deprived) are associated with increased slopes, but these have confidence intervals very close to including zero and, with the exception of IMD quintile 5, are not significant for the larger sample.

Comparing to previous studies, our results agree with a number of existing findings regarding the association between building physical characteristics and energy consumption [14,41]: buildings that have more rooms, more bedrooms, are more detached, are older, and that experience colder or less sunny weather are associated with increased energy consumption.

Similarly, we find presence of air-conditioning increases demand, though unlike previous studies [14], we do not find any significant association between presence of any other appliances and demand. A possible explanation is that while previous studies have tended to focus on *electricity* consumption only, which is more likely to be affected by appliances, we are analysing *total* energy consumption, which is dominated by space and water heating, and as such unlikely to be significantly affected by electrical appliances such as laptops, dishwashers etc.

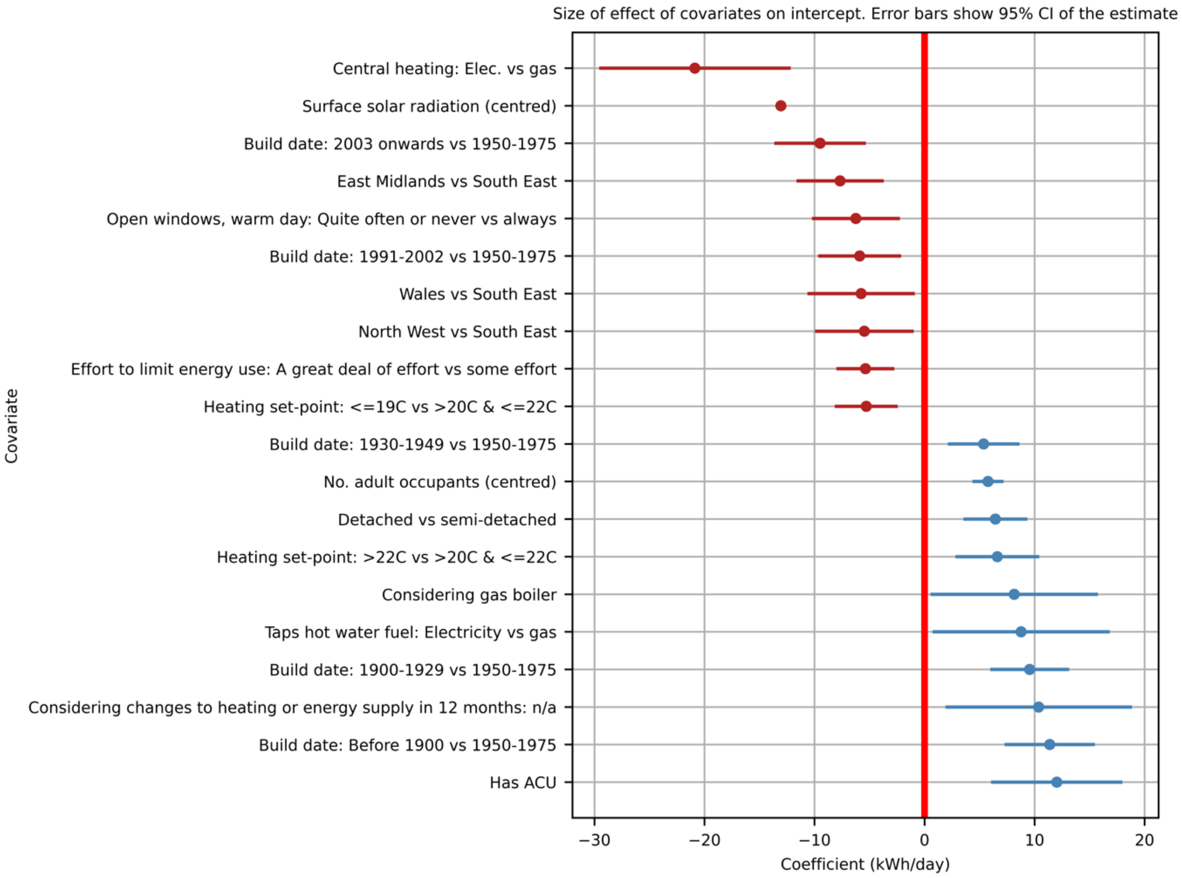
This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

Our results confirm a number of existing findings regarding the effect of sociodemographic characteristics on energy consumption [14,41]: households with more adult occupants, more children, and with older adult occupants, are associated with increased energy consumption. While the latter two are not shown in Figure 5 or Figure 6 both have significant coefficients and VIFs < 5 for both samples.

Previous studies report mixed results for the effect of tenure, and education on energy consumption. We find no significant effect associated with tenure, or education (a higher proportion of adults with qualifications) when multiple confounding factors are controlled.

Behavioural factors can include energy conservation behaviour in the form of 'purchasing' activities or 'habitual' actions and are less well studied than the previous categories of factors [20]. Nonetheless some previous studies report an association between habitual energy saving behaviours and reduced consumption [20,70,71]. We found that households that set lower heating temperature set-points consumed less than those that set higher set-points. Households who made 'a great deal of effort' to limit or reduce their energy consumption were associated with lower consumption than those who made 'some effort'.

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).



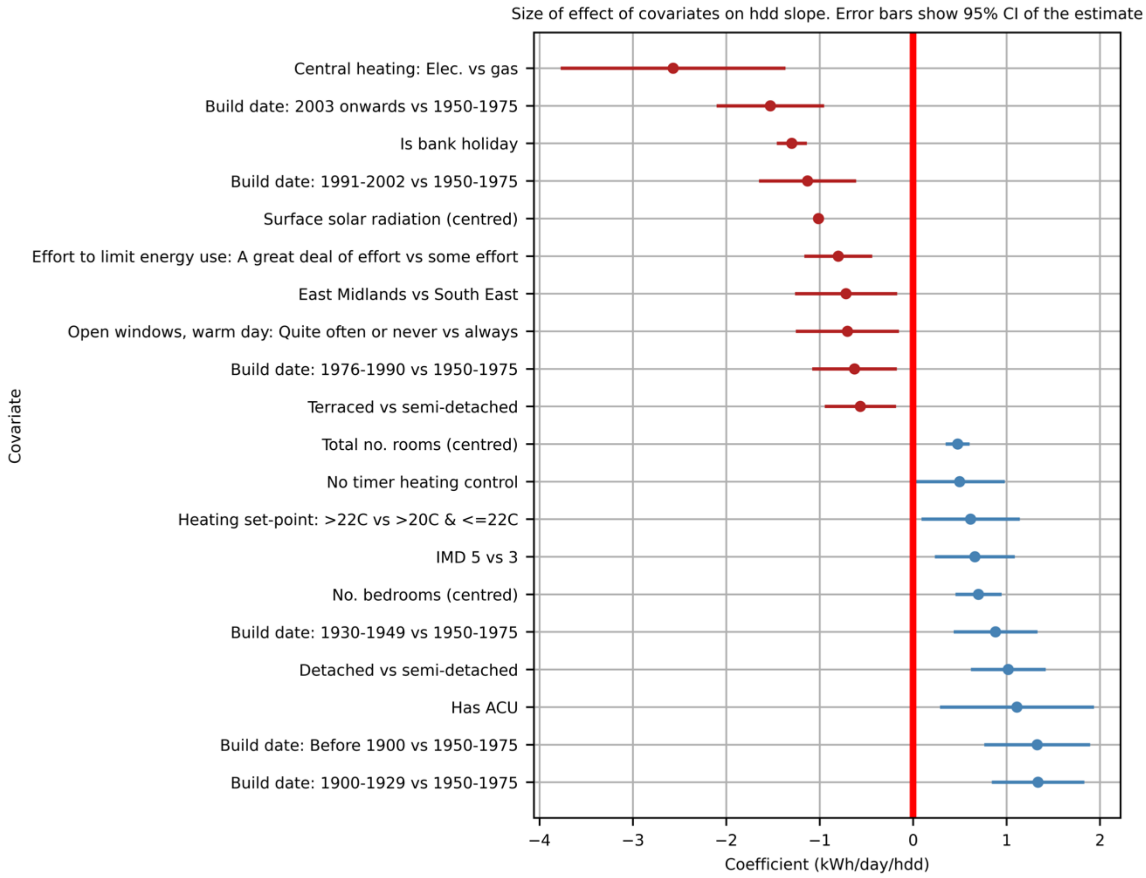


Figure 5. Size of coefficients for statistically significant (p<0.05) and low VIF (<5) variables for the Basic plus SERL survey model and larger sample (N=1418), showing the variables with the ten largest positive and negative effects on intercept (upper) and hdd slope (lower). Those with negative effect are shown in red, those with positive effect shown in blue. Error bars show 95% confidence interval of the estimate.

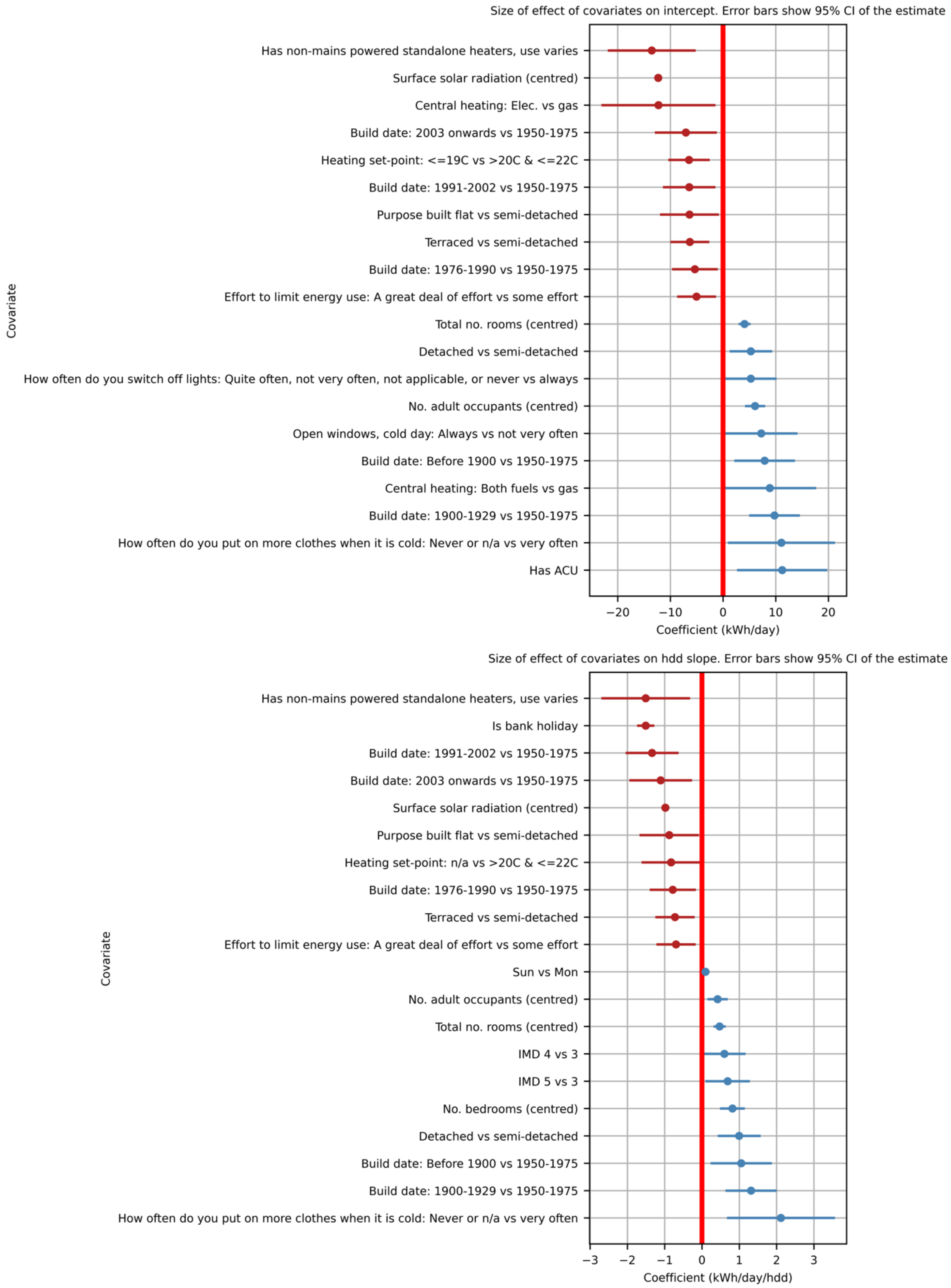


Figure 6. Size of coefficients for statistically significant (p<0.05) and low VIF (<5) variables for the Basic plus SERL survey model and smaller sample (N=682), showing the variables with the ten largest positive and negative effects on intercept

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

(upper) and hdd slope (lower). Those with negative effect are shown in red, those with positive effect shown in blue. Error bars show 95% confidence interval of the estimate.

4.4 Comparison with previous studies

Returning to our first research question, '*What is the overall explanatory power of SERL Observatory data with respect to variation in household-level daily residential energy consumption and does this improve on studies reported in the literature?',* we find that when measured in terms of adjusted R^2 calculated using cross-validation testing errors, the SERL Observatory data explains between 63% and 80% of the variation in daily household total energy consumption, depending on sample size and combinations of contextual data used. For a given sample with full data availability, a model using all available data performs the best for within-group prediction while a model using Basic plus EPC data (i.e. not including SERL survey data) performs marginally better on between-group prediction. However, given the relatively small sample sizes considered here, and the resulting penalisation of adjusted R^2 values, the model using Basic plus SERL Survey (i.e. not including EPC data) performed best overall simply because SERL survey data is available for more participants than EPC data, and this allowed the sample size to be more than doubled from N=682 to 1418. Other studies in the literature (see Table 1) report adjusted R^2 of 0.29-0.44 for daily demand.

While we report errors for the models, we note that none of the cited studies report comparable errors with the exception of [15,20]. Direct comparison with these is complicated because they log-transform the dependent variable while we do not. Nonetheless, [20] reports errors that are approximately 20% smaller than the standard deviation of the dependent variable. The standard deviation of the dependent variable for our samples is 42.97 kWh/day (87% of the mean) and 42.19 kWh/day (90% of the mean) for our larger and smaller samples respectively. Our best performing models have RMSE errors ranging from 23.06 to 28.27 kWh/day, equivalent to 45% to 36% smaller than the standard deviations. Overall, therefore, our results compare favourably with those found in the literature, however there is clearly substantial scope for improving model accuracy.

4.5 SERL Observatory: a new national data resource for energy demand research

Returning to the final research question, we have shown that the EPC data and SERL survey data, when included alongside the basic data, are broadly similar in terms of explanatory power, with the EPC data marginally outperforming the SERL survey data. The SERL survey data is however much less affected by multicollinearity and has higher data availability for the SERL Observatory. Future researchers using similar techniques may wish to opt for a balance of maximising sample size and explanatory power by not requiring complete EPC data for their analytic sample. We believe these results demonstrate the value of the SERL survey as a tool for collecting useful contextual data with relatively low participant burden, and note the complementarity of the SERL survey with EPC data which is nonetheless widely available for UK dwellings.

Overall, our results demonstrate that a large amount of variation can be explained by data collected within the SERL Observatory, and have demonstrated a number of methodological approaches that should prove useful for researchers aspiring to use the data. The results largely support existing theory and add to the empirical evidence base that building physical characteristics, household sociodemographic information, and household behavioural factors all explain aspects of demand, across a wide range of contexts. Considering the complexity of the subject under investigation (daily residential energy consumption), the simplicity of the approach to data selection used here, and the relatively low burden on participants for data collection, we believe this is a promising result that demonstrates the value of the SERL Observatory dataset as a data resource for improving the understanding of energy demand in residential buildings. The final (third) wave of SERL participant recruitment was completed in March 2021 and over 8,000 further participants were recruited,

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

bringing the total participant number to over 13,000. We therefore encourage future energy demand projects involving surveys to harmonise with the SERL survey to support greater interpretation, reproducibility and cross-validation between research findings [72].

4.6 Statistical issues and limitations

Nonetheless, this work has revealed and is subject to a number of statistical issues and limitations, listed here in no particular order of importance, all of which restrict our ability to draw robust inferences from these particular results:

- While one of the primary goals of the study is inference and to improve understanding of residential energy consumption, the samples analysed here are non-random, and non-representative due to biases in data collection and sample preparation. The current results cannot therefore be generalised to the population from which the SERL sample was drawn;
- The modelling approach is limited: it employs simplified assumptions (e.g. linearity) regarding the relationship between the variables of interest and makes no attempt at variable selection beyond combining all variables available in the different combinations of linkable contextual data;
- The inclusion of such large numbers of variables in models is highly likely to result in a number of spurious inferences of statistically significant effects (Type I errors). However, since one purpose of the paper is to guide future analysis we felt it important to demonstrate this approach (and the potential problem) for rhetorical purposes. To some extent this is mitigated by the presentation of confidence intervals alongside point estimates of the effect sizes as suggested by Anderson et al [73];
- The models are affected by multicollinearity, some severely so. We have however been careful to highlight and address this problem where possible. This was an outcome of our aim here to explore the explanatory power of including the full range of data available in the SERL dataset, which itself supports the aim of highlighting the potential of the dataset for future studies that focus down on relationships between domestic energy use and specific contextual factors. We have demonstrated where future analyses using SERL data will need to guard against this issue by using smaller theoretically-informed or model-based variable sets. This supports our aim of guiding future high quality analysis of the SERL data.

4.7 Future work

This paper presents a first initial step in a larger programme of research by multiple organisations using SERL Observatory data. We have started with simple but limited analysis; for example, using a fixed degree-day base to account for variations in heating of buildings, whereas it is possible that the temperatures at which heating is turned on is much more complex and interrelated with many of the variables. We also present models employing more covariates than would be usual and which, as we have noted, display multicollinearity and instability as a result. We therefore plan more sophisticated analysis of the above data using weightings to produce population estimates, applying non-linear methods for inference and predictive models, and using variable selection methods to identify the most important individual factors.

Further, we plan to use the full 13,000 observatory release 3 data to give greater statistical power and conduct research of relevance to policy, notably: investigating the impact of coronavirus on energy demand; producing a range of residential energy statistics of relevance to a wider audience of non-academics; and investigating the use of the SERL Observatory as a counterfactual group for the evaluation of energy efficiency measures and policies. We also plan to analyse gas and electricity use separately to improve our understandings of the factors that correlate with each. There is

This is an accepted manuscript version of a published journal article in *Energy & Buildings* (DOI: https://doi.org/10.1016/j.enbuild.2022.111845).

considerable scope for research using the SERL Observatory data. The dataset is available for other UK academic researchers – we encourage such UK researchers to submit proposals to access it. More information about how to do this can be found on the SERL website (www.serl.ac.uk) and UKDS data catalogue [10].

5 Conclusions

This paper presents analysis of the SERL Observatory: a dataset of linked smart meter data and socio-technical contextual data for a representative sample of over 13,000 GB households. Here we analyse data from two nested sub-samples (N=1418 & 682) of the first two recruitment waves (initial sample N=4716) and for the pre-coronavirus period (taken to be before March 2020).

The first aim was to quantify how much of the variation in total energy consumption can be explained by different combinations of SERL Observatory variables: ‘basic’ (e.g. local weather, region, date), EPC (where available), and the SERL survey (questions relating to the dwelling and occupants). As multiple observations were available per participant, linear mixed effects models were used to regress household-level daily total energy consumption over time against successive levels of contextual data to reveal the relationship between energy use and static (constant) and temporally changing variables (basic: weather, region, IMD and date; EPC; SERL survey; all data combined).

The explanatory power of the models was quantified using adjusted R^2 and root mean squared error (RMSE). The SERL Observatory data explains between 63% and 80% of the variation in daily household total energy consumption, depending on sample size and combinations of contextual data used. For within-sample prediction (i.e. where historic observations for each household are available), the model using all available contextual data performed best, while for between-sample prediction (i.e. where historic data is not available) the model using basic plus EPC data marginally outperformed others for the smaller sample. However, the model using Basic plus SERL Survey (i.e. not including EPC data) performed best overall as it could be applied to the larger sample (N=1418 rather than 682) and so resulted in smaller penalisation of adjusted R^2 value.

The best performing models have RMSE errors ranging from 23.06 to 28.27 kWh/day, equivalent to 45% to 36% less than the standard deviations of the samples. Overall, these results compare favourably with those found in the literature, however there is clearly substantial scope for improving model accuracy, and the results indicate that non-linear models and regularisation techniques could help achieve this.

The second aim was to identify variables observed in SERL Observatory data that are strongly associated with variation in household-level residential energy consumption using a p-value<0.05, VIF<5 threshold for demonstration purposes. Given high multicollinearity particularly associated with EPC data, this was restricted to the Basic plus SERL Survey model applied to the two sample sizes. The results were broadly as expected: buildings that are older, have more rooms and bedrooms, have air-conditioning, and experience colder or less sunny weather were associated with increased energy consumption. Households with more occupants, more children, and with older adult occupants were also associated with increased energy consumption. Energy consumption in households was found to be lower in households that set lower heating temperature setpoints, and that tried to save energy.

In summary, this paper has demonstrated that the SERL Observatory dataset is a rich resource of energy data and relevant contextual data, and that the contextual data is robust as it explains energy use to a good degree in much the way that existing literature would lead us to expect. The dataset is available to UK Accredited Researchers and we encourage researchers to submit proposals to access it. This paper provides guidance on appropriate methods to use when analysing the data.

6 Acknowledgements

This work has been funded by EPSRC through grant EP/P032761/1. There are over 30 individuals across 8 organisations in the SERL Consortium (University College London, the University of Essex (UK Data Archive), University of Edinburgh, Cardiff University, Loughborough University, Leeds Beckett University, the University of Southampton and the Energy Saving Trust) who have contributed to the development of SERL and thus the content of this paper. Particular thanks go to the SERL technical team at the UK Data Archive: Darren Bell, Deirdre Lungley, Martin Randall and Jacob Joy and to SERL Consortium Manager James O'Toole at UCL.

We acknowledge support from the SERL Independent Advisory Board, Data Governance Board and Research Programme Board which played critical role in the establishment and ethical operation of SERL.

The SERL Observatory includes European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 data. Neither the European Commission nor the European Centre for Medium-Range Weather Forecasts is responsible for any use that may be made of the Copernicus information or data it contains.

7 Author contributions

Eoghan McKenna: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft.

Jessica Few: Methodology, Writing - Original Draft, Visualization, Software, Formal analysis.

Ellen Webborn: Methodology, Data Curation, Software, Investigation, Writing - Review & Editing.

Ben Anderson: Methodology, Writing - Review & Editing.

Simon Elam: Data Curation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition, Investigation.

David Shipworth: Investigation, Writing - Review & Editing, Funding acquisition.

Martin Pullinger: Investigation, Writing - Review & Editing.

Adam Cooper: Investigation, Writing - Review & Editing.

Tadj Oreszczyn: Investigation, Writing - Review & Editing, Supervision, Funding acquisition.

8 References

[1] IEA, Net Zero by 2050, Paris, 2021. https://www.iea.org/reports/net-zero-by-2050.

[2] J. Wachsmuth, V. Duscha, Achievability of the Paris targets in the EU—the role of demand-side-driven mitigation in different types of scenarios, Energy Effic. 12 (2019) 403–421. https://doi.org/10.1007/s12053-018-9670-4.

[3] F. Creutzig, J. Roy, W.F. Lamb, I.M.L. Azevedo, W. Bruine De Bruin, H. Dalkmann, O.Y. Edelenbosch, F.W. Geels, A. Grubler, C. Hepburn, E.G. Hertwich, R. Khosla, L. Mattauch, J.C. Minx, A. Ramakrishnan, N.D. Rao, J.K. Steinberger, M. Tavoni, D. Ürge-Vorsatz, E.U. Weber, Towards demand-side solutions for mitigating climate change, Nat. Clim. Chang. 8 (2018)

268–271. https://doi.org/10.1038/s41558-018-0121-1.

[4] A.C.G. Cooper, Building physics into the social: Enhancing the policy impact of energy studies and energy social science research, Energy Res. Soc. Sci. 26 (2017) 80–86. https://doi.org/10.1016/j.erss.2017.01.013.

[5] J. Love, A.C.G. Cooper, From social and technical to socio-technical: Designing integrated research on domestic energy use, Indoor Built Environ. 24 (2015) 986–998. https://doi.org/10.1177/1420326X15601722.

[6] A.C.G. Cooper, Evaluating energy efficiency policy: understanding the 'energy policy epistemology' may explain the lack of demand for randomised controlled trials, Energy Effic. 11 (2018) 997–1008. https://doi.org/10.1007/s12053-018-9618-8.

[7] A. Cooper, D. Shipworth, A. Humphrey, UK Energy Lab: A feasibility study for a longitudinal, nationally representative sociotechnical survey of energy use, London, 2014. https://www.ucl.ac.uk/steapp/sites/steapp/files/synthesis.pdf (accessed May 20, 2021).

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

[9] E. Webborn, T. Oreszczyn, Champion the energy data revolution, Nat. Energy 2019 48. 4 (2019) 624–626. https://doi.org/10.1038/s41560-019-0432-0.

[10] S. Elam, E. Webborn, E. McKenna, T. Oreszczyn, B. Anderson, Ministry of Housing Communities & Local Government, European Centre for Medium-Range Weather Forecasts, Royal Mail Group Limited, Smart Energy Research Lab Observatory Data, 2019-2020: Secure Access, (2020). https://doi.org/http://doi.org/10.5255/UKDA-SN-8666-1.

[11] E. Shove, M. Pantzar, M. Watson, The dynamics of social practice: Everyday life and how it changes, Sage, Los Angeles, Calif. ; London, 2012.

[12] K. Gram-Hanssen, Efficient Technologies or User Behaviour, Which Is the More Important When Reducing Households' Energy Consumption?, Energy Effic. 6 (2013) 447–457. https://doi.org/10.1007/s12053-012-9184-4.

[13] M. Hand, E. Shove, D. Southerton, Explaining showering: a discussion of the material, conventional, and temporal dimensions of practice, Sociol. Res. Online. 10 (2005). http://www.socresonline.org.uk/10/2/hand.html.

[14] R. V. Jones, A. Fuertes, K.J. Lomas, The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings, Renew. Sustain. Energy Rev. 43 (2015) 901–917. https://doi.org/10.1016/j.rser.2014.11.084.

[15] A. Satre-Meloy, M. Diakonova, P. Grünewald, Daily life and demand: an analysis of intra-day variations in residential electricity consumption with time-use data, Energy Effic. 13 (2020) 433–458. https://doi.org/10.1007/s12053-019-09791-1.

[16] P. Grünewald, M. Diakonova, Societal differences, activities, and performance: Examining the role of gender in electricity demand in the United Kingdom, Energy Res. Soc. Sci. 69 (2020) 101719. https://doi.org/10.1016/J.ERSS.2020.101719.

[17] M. Gleue, J. Unterberg, A. Löschel, P. Grünewald, Does demand-side flexibility reduce emissions? Exploring the social acceptability of demand management in Germany and Great Britain, Energy Res. Soc. Sci. 82 (2021) 102290. https://doi.org/10.1016/J.ERSS.2021.102290.

[18] S. Wei, R. Jones, P. de Wilde, Driving Factors for Occupant-Controlled Space Heating in Residential Buildings, Energy Build. 70 (2014) 36–44. https://doi.org/10.1016/j.enbuild.2013.11.001.

[19] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renew. Sustain. Energy Rev. 13 (2009) 1819–1835. https://doi.org/10.1016/j.rser.2008.09.033.

[20] A. Satre-Meloy, Investigating structural and occupant drivers of annual residential electricity consumption using regularization in regression models, Energy. 174 (2019) 148–168. https://doi.org/10.1016/j.energy.2019.01.157.

[21] A. Foucquier, S. Robert, F. Suard, L. Stéphan, A. Jay, State of the art in building modelling and energy performances prediction: A review, Renew. Sustain. Energy Rev. 23 (2013) 272–288. https://doi.org/10.1016/j.rser.2013.03.004.

[22] A.T. Nguyen, S. Reiter, P. Rigo, A review on simulation-based optimization methods applied to building performance analysis, Appl. Energy. 113 (2014) 1043–1058. https://doi.org/10.1016/j.apenergy.2013.08.061.

[23] K. Amasyali, N.M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renew. Sustain. Energy Rev. 81 (2018) 1192–1205. https://doi.org/10.1016/j.rser.2017.04.095.

[24] Y. Wei, X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, M. Han, X. Zhao, A review of data-driven approaches for prediction and classification of building energy consumption, Renew. Sustain. Energy Rev. 82 (2018) 1027–1047. https://doi.org/10.1016/j.rser.2017.09.108.

[25] Y. Iwafune, Y. Yagita, High-resolution determinant analysis of Japanese residential electricity consumption using home energy management system data, Energy Build. 116 (2016) 274–284. https://doi.org/10.1016/j.enbuild.2016.01.017.

[26] M.J. Kim, Understanding the determinants on household electricity consumption in Korea: OLS regression and quantile regression, Electr. J. 33 (2020) 106802. https://doi.org/10.1016/j.tej.2020.106802.

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

[28] B. Anderson, S. Lin, A. Newing, A.B. Bahaj, P. James, Electricity consumption and household characteristics: Implications for census-taking in a smart metered future, Comput. Environ. Urban Syst. 63 (2017) 58–67. https://doi.org/10.1016/j.compenvurbsys.2016.06.003.

[29] E.W. Frees, Longitudinal and Panel Data, Cambridge University Press, 2004. https://doi.org/10.1017/cbo9780511790928.

[30] G. James, D. Witten, T. Hastie, R. Tibshirani, An introduction to statistical learning, Springer, 2017. https://link.springer.com/content/pdf/10.1007/978-1-4614-7138-7.pdf (accessed February 5, 2021).

[31] G. James, D. Witten, T. Hastie, R. Tibshirani, An introduction to Statistical Learning, 2000. https://doi.org/10.1007/978-1-4614-7138-7.

[32] J. Friedman, T. Hastie, R. Tibshirani, The elements of statistical learning, 2001. http://statweb.stanford.edu/~tibs/book/preface.ps (accessed February 5, 2021).

[33] A. Kavousian, R. Rajagopal, M. Fischer, Determinants of residential electricity consumption:

Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior, Energy. 55 (2013) 184–194. https://doi.org/10.1016/j.energy.2013.03.086.

[34] MHCLG, English Housing Survey 2017 to 2018: energy, 2019. https://www.gov.uk/government/statistics/english-housing-survey-2017-to-2018-energy.

[35] ACER, Annual report on the results of monitoring the internal electricity and natural gas markets in 2017, 2018.

[36] R. V. Jones, K.J. Lomas, Determinants of high electrical energy demand in UK homes: Appliance ownership and use, Energy Build. 117 (2016) 71–82. https://doi.org/10.1016/j.enbuild.2016.02.020.

[37] H. Fan, I.F. MacGill, A.B. Sproul, Statistical analysis of driving factors of residential energy demand in the greater Sydney region, Australia, Energy Build. 105 (2015) 9–25. https://doi.org/10.1016/j.enbuild.2015.07.030.

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

[39] C. Shalizi, The Truth About Linear Regression, Carnegie Mellon University, 2019. http://www.stat.cmu.edu/~cshalizi/TALR/TALR.pdf (accessed December 5, 2021).

[40] H. Fan, I.F. MacGill, A.B. Sproul, Statistical analysis of drivers of residential peak electricity demand, Energy Build. 141 (2017) 205–217. https://doi.org/10.1016/j.enbuild.2017.02.030.

[41] BEIS, NEED Annex D: Determinants of household gas use, 2019.

[42] T. Snijders, R. Bosker, Multilevel analysis: An introduction to basic and advanced multilevel modeling, SAGE, 2012.

[43] E. Webborn, S. Elam, E. McKenna, Utilising Smart Meter Data for Research and Innovation in the UK (forthcoming), in: Proc. Eur. Counc. an Energy Effic. Econ. Summer Study, 2019.

[44] E. Webborn, E.J. McKenna, S. Elam, B. Anderson, A. Cooper, T. Oreszczyn, Increasing response rates and reducing bias: Learnings from the Smart Energy Research Lab pilot study, (n.d.). https://doi.org/10.31219/OSF.IO/F82B7.

[45] E. McKenna, E. Webborn, … P.L.-E. 2019 S., undefined 2019, Analysis of international residential solar PV self-consumption, Discovery.Ucl.Ac.Uk. (n.d.). http://discovery.ucl.ac.uk/id/eprint/10075770 (accessed October 4, 2019).

[46] E. Webborn, S. Elam, E. McKenna, T. Oreszczyn, Utilising smart meter data for research and innovation in the UK, ECEEE Summer Study Proc. (2019) 1387–1396.

[47] E. Webborn, E.J. McKenna, S. Elam, B. Anderson, A. Cooper, T. Oreszczyn, Increasing response rates and reducing bias: Learnings from the Smart Energy Research Lab pilot study, OSF Prepr. (2021). https://doi.org/10.31219/OSF.IO/F82B7.

[48] J. Crawley, E. McKenna, V. Gori, T. Oreszczyn, Creating Domestic Building Thermal Performance Ratings Using Smart Meter Data, Build. Cities. 1 (2020) 1–13. https://doi.org/10.5334/BC.7.

[49] MHCLG, Energy Performance of Buildings Data: England and Wales, (2020). https://epc.opendatacommunities.org/.

[50] J. Crawley, P. Biddulph, P.J. Northrop, J. Wingfield, T. Oreszczyn, C. Elwell, Quantifying the Measurement Error on England and Wales EPC Ratings, Energies. 12 (2019).

[51] H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, J.-N. Thépaut, ERA5 hourly data on single levels from 1979 to present., (2018). https://doi.org/10.24381/cds.adbb2d47.

[52] DECC, Energy Trends: December 2014, special feature article - Energy usage in household with solar PV installations, 2014. https://www.gov.uk/government/statistics/energy-trends-december-2014-special-feature-article-energy-usage-in-household-with-solar-pv-installations.

[53] Brook Lyndhurst, Uptake of Ultra Low Emission Vehicles in the UK, 2015. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/464763/uptake-of-ulev-uk.pdf.

[54] J. Spinoni, J. Vogt, P. Barbosa, European degree-day climatologies and trends for the period 1951-2011, Int. J. Climatol. 35 (2015) 25–36. https://doi.org/10.1002/joc.3959.

[55] D. Iacobucci, M.J. Schneider, D.L. Popovich, G.A. Bakamitsos, Mean centering helps alleviate “micro” but not “macro” multicollinearity, Behav. Res. Methods. 48 (2016) 1308–1317. https://doi.org/10.3758/S13428-015-0624-X.

[56] M. Wissmann, H. Toutenburg, Role of categorical variables in multicollinearity in the linear regression model, 2007. https://epub.ub.uni-muenchen.de/2081 (accessed August 23, 2021).

[57] H. Kang, The prevention and handling of the missing data, Korean J. Anesthesiol. 64 (2013) 402–406. https://doi.org/10.4097/kjae.2013.64.5.402.

[58] P.C. Austin, I.R. White, D.S. Lee, S. van Buuren, Missing Data in Clinical Research: A Tutorial on Multiple Imputation, Can. J. Cardiol. 37 (2021) 1322–1331. https://doi.org/10.1016/j.cjca.2020.11.010.

[59] D.B. Rubin, Multiple Imputation after 18+ Years, J. Am. Stat. Assoc. 91 (1996) 473–489. https://doi.org/10.1080/01621459.1996.10476908.

[60] E. Webborn, J. Few, E. McKenna, S. Elam, M. Pullinger, B. Anderson, D. Shipworth, T. Oreszczyn, 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 (2021) 6934. https://doi.org/10.3390/en14216934.

[61] MHCLG, English Housing Survey 2018 to 2019: headline report, 2020. https://www.gov.uk/government/statistics/english-housing-survey-2018-to-2019-headline-report (accessed May 26, 2021).

[62] BEIS, Energy consumption in the UK - GOV.UK, 2021. https://www.gov.uk/government/statistics/energy-consumption-in-the-uk (accessed May 26, 2021).

[63] J. Wooldridge, Introductory econometrics: A modern approach, 2015. https://books.google.com/books?hl=en&lr=&id=wUF4BwAAQBAJ&oi=fnd&pg=PR3&dq=wooldridge+introductory+econometrics&ots=cATyYDlngo&sig=AkalfyXzQggN67iYhrU5UKaKCH0 (accessed September 10, 2021).

[64] G. James, D. Witten, T. Hastie, R. Tibshirani, An introduction to statistical learning with applications in R, 2013. https://doi.org/10.1017/CBO9781107415324.004.

[65] S. Nakagawa, H. Schielzeth, A general and simple method for obtaining R^2 from generalized linear mixed-effects models, Methods Ecol. Evol. 4 (2013) 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

[66] J. Miles, R -Squared, Adjusted R -Squared , Encycl. Stat. Behav. Sci. (2005). https://doi.org/10.1002/0470013192.BSA526.

[67] T. pandas development Team, Pandas, (2020). https://doi.org/10.5281/zenodo.3509134.

[68] W. McKinney, Data Structures for Statistical Computing in Python, in: S. van der Walt, J. Millman (Eds.), Proc. 9th Python Sci. Conf., 2010: pp. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a.

[69] S. Seabold, J. Perktold, Statsmodels: Econometric and statistical modeling with python, in: Proc. 9th Python Sci. Conf., 2010: p. 92.

[70] H. Wallis, M. Nachreiner, E. Matthies, Adolescents and electricity consumption; Investigating sociodemographic, economic, and behavioural influences on electricity consumption in households, Energy Policy. 94 (2016) 224–234. https://doi.org/10.1016/j.enpol.2016.03.046.

[71] K. Steemers, G.Y. Yun, Household energy consumption: A study of the role of occupants, Build. Res. Inf. 37 (2009) 625–637. https://doi.org/10.1080/09613210903186661.

[72] G. Huebner, M. Fell, N. Watson, Improving energy research practices: guidance for transparency, reproducibility and quality, Build. Cities. 2 (2021) 1–20. https://doi.org/10.5334/bc.67.

[73] B. Anderson, T. Rushby, A. Bahaj, P. James, Ensuring statistics have power: Guidance for designing, reporting and acting on electricity demand reduction and behaviour change programs, Energy Res. Soc. Sci. 59 (2020) 101260. https://doi.org/10.1016/j.erss.2019.101260.