# Domestic heating behaviour and room temperatures: empirical evidence from Scottish homes

Martin Pullinger[a], Niklas Berliner[a], Nigel Goddard[a], David Shipworth[b]

[a] University of Edinburgh, School of Informatics, Informatics Forum, 10 Crichton Street,, Edinburgh, EH8 9AB, United Kingdom
[b] University College London Energy Institute, No.10 Montague Street, Holborn, London, WC1B 5BJ, United Kingdom

**Abstract**

In this paper, we describe patterns of residential heating based on data from 255 homes in and around Edinburgh, Scotland, UK, spanning August 2016 to June 2018. We describe: (i) the room temperatures achieved, (ii) the diurnal durations of heating use, and (iii) common diurnal patterns of heating behaviour. We investigate how these factors vary between weekdays and weekends, over the course of the year, by external temperature, and by room type. We compare these empirical findings with the simplifying assumptions about heating patterns found in the UK's Standard Assessment Procedure (SAP), a widely-used building energy performance model. There are areas of concurrence and others of substantial difference with these model assumptions. Indoor achieved temperatures are substantially lower than SAP assumptions. The duration and timings of heating use varies substantially between homes and along lines of season and outdoor temperature, whereas the SAP model assumes no such variation. Little variation is found along the lines of weekday vs. weekend, whereas the SAP model assumes differences, or between living space and other rooms, consistent with the SAP. The results are relevant for those interested in how SAP assumptions regarding household heating behaviours and achieved indoor temperatures concur with empirical data.

*Keywords:* Residential heating behaviours, Achieved temperatures, Heating durations, Diurnal heating patterns, Cluster analysis, Heating zoning, Seasonal change

## 1. Introduction

Efforts to decarbonise the residential heating system are gathering pace in the UK, as in much of the rest of the world, as part of achieving the goal of reaching net zero carbon emissions by 2050 for the UK as a whole, and 2045 in Scotland [1]. One crucial element in achieving this efficiently is accurately estimating the energy performance of the buildings, from the level of individual dwellings through to the entire building stock, or sections of it. This includes predicting the impacts of different interventions, such as installing double glazing and insulation, and switching heating fuel types. In the UK, the government's recommended model to make such predictions is the Standard Assessment Procedure (SAP). The SAP is a simplified version of the BRE's Domestic Energy Model (BREDEM) [2] developed for assessing the performance of buildings under a standardised set of conditions (a set 'occupancy schedule') describing when a dwelling is occupied and when associated energy-using practices, such as heating, are engaged in. Standardised conditions are adopted to permit comparison between dwellings independently of occupancy effects. These standardised conditions are also widely used in BREDEM-based building stock models used to estimate energy demand from buildings in use. In this context, the standardised conditions represent simplifying assumptions about the average occupancy schedule, and as such enable the energy use of the build stock to be estimated without the unfulfillable requirement of gathering and using full occupancy schedule data for each dwelling. However, this use of standardised conditions is problematic if they do not sufficiently capture aspects of occupants' energy-using behaviours observed empirically in the actual building stock. Model assumptions may be overly simplified or based on incorrect or out-of-date specifications of occupant behaviour, so do not accurately reflect population averages or the diversity and drivers of different behaviours [3]. Model energy use estimates are then more likely to deviate from observation, particularly at finer-grained spatial and temporal resolutions or for particular types of dwelling or occupant, where conditions may differ substantially from the full-population average. It is thus important for the development of stock models to evaluate how well the standardised conditions reflect those found in buildings in use, as part of the process of evaluating potential opportunities to improve model performance by aligning assumptions more closely with empirical reality.

To that end, the aim of this paper is to evaluate how the simplifying assumptions about heating behaviour and indoor temperatures that are found in the SAP model compare to recently published empirical data from a sample of Scottish homes.

The SAP model assumes the following patterns and outcomes of heating use in UK homes:

i **Achieved room temperatures:** During periods of active heating (i.e. central heating use), the model assumes achieved temperatures are 21°C for the living areas (generally this is the living room/lounge - see [4, p. 23] for the detailed definition), and between 18°C and 21°C elsewhere, depending on the building's Heat Loss Parameter - a function of multiple building physical characteristics [4, p. 219]. The SAP assumes these achieved indoor temperatures are standard across the heating season, invariant to external conditions such as outdoor temperature, and the same on weekdays and at weekends.

ii **Patterns and durations of active heating:** For homes with boiler central heating systems, such as those in our reference dataset, the assumption is that the whole home is actively heated between 07:00-09:00 and 16:00-23:00 (total 9 hours) for weekdays, and 07:00-23:00 (total 16 hours) for weekends [4, p. 219]. These heating patterns are assumed to be standard across the heating season, invariant to external conditions such as the outdoor temperature, and the same for each room of the home.

iii **Heating season:** This is the period when central heating is used, and is taken to span October to May [4, p. 220]. Outside this period, the model assumes there is no active heating, i.e no use of the central heating system.

The empirical data used in this paper is drawn from a recently published dataset that includes data from a sample of 255 homes from the region in and around Edinburgh, Scotland, UK, collected by our research team. The data was collected from the homes for a mean of 286 days over a period spanning two heating seasons, from August 2016 to June 2018. The homes all had radiators heated by gas-fired combi-boilers as the main heating source, and included a range of occupancy levels and building types, ages and sizes.

This paper compares and contrasts the SAP assumptions described above with the empirical reality from this sample of Scottish homes. As such, we focus on the principle patterns in the data for:

i the room temperatures achieved,

ii the diurnal durations of heating use,

iii the common patterns of diurnal heating behaviour, in terms of the periods of the day when heating is on and off.

We furthermore describe if and how these factors vary between weekdays and weekends, over the course of the year, by external temperature, and by room type.

This paper adds to the relatively small published literature on UK residential heating patterns and temperature outcomes. To our knowledge, it is the first paper to focus on Scottish homes and to draw on data covering radiator use and ambient temperature from all rooms in the dwellings. The findings complement the existing literature, in terms of indicating possibilities for future refinements to the SAP model.

The rest of this article is structured as follows: Section 2 reviews the literature on previous empirical work on heating patterns and indoor temperatures in UK homes, focusing on aspects related to the above assumptions in the SAP. Section 3 describes the methodology. Section 4 describes the results. Section 5 discusses the results and how they relate to previous work and the SAP model. Section 6 concludes, including considering future work directions.

## 2. Literature review: Domestic room temperatures and heating patterns

Here we review previous work relating to patterns of heating in UK homes and the temperature outcomes, focusing on work that draws on empirical data from homes, particularly where comparisons are made to the Standard Assessment Procedure model assumptions that we are focusing on.

**Achieved room temperatures**

Previous empirical work has investigated how homes' indoor temperatures compare to SAP assumptions. Two papers provide insight into demand temperatures. Hughes *et al* 2010 [5] use data from the Energy Follow-Up Survey (EFUS), a subsample of 2,616 English homes from the 2010/11 English Housing Survey that participated in interviews and provided meter readings; a subsample also had temperature data loggers recording at 20-minute intervals in their living rooms, main bedrooms and hallways, covering November 2010 to January 2011. Based on living room temperature gradients, the authors identified the heating season average dwelling internal demand temperature across 405 dwellings with the complete range of data to be 19.8°C (Standard Deviation, S.D., 2.14°C, median 20.02°C). Shipworth *et al* 2010 [6] meanwhile analysed survey and 45-minute temperature data from data loggers placed in bedrooms and living rooms between July 2007 and February 2008, from a stratified random sample of 358 English households with "gas or oil-fired central heating systems with radiators as their main form of heating". Based on inferred periods of active heating (when temperatures increased between time points, for data from November 2007 to February 2008), living room average maximum temperatures, which were taken as being the mean thermostat settings for the dwellings, across the sample for the heating season were identified as 21.1°C (S.D. 2.5°C, median 21.3°C); meanwhile self-reported figures from the surveys indicated a mean of 19.0°C (S.D. 3.0°C, median 20.0°C). Huebner *et al* 2013 [7] further analysed the living room temperature data for a different subsample of 248 centrally heated homes from the same dataset, covering 92 days from November 2007 to January 2008, to look at achieved heating period temperatures. They found that the temperatures seldom reached the SAP-assumed demand temperature of 21°C during the SAP heating periods, with mean temperatures of 18.3 °C for the SAP weekday morning heating period, and a somewhat warmer 19.8 °C for the weekday evening heating period, and 19.3°C for the weekend heating period, with a similarly large standard deviation of around 2.5°C in each case. Averaged across all the data, for most times of the day few homes were above 20.5°C, although from early evening the proportion rapidly increased, to stabilise at around 50% of homes being above that temperature from approximately 18:45 until midnight, typically the warmest period of the day. The study found substantively little difference in achieved temperatures between weekends and weekdays, but substantial variation between homes. A similar pattern of fluctuating average temperatures across the day was also found by Hanmer *et al* 2019 [8], drawing on data from digital heating control units for a sample of 337 UK homes for an 8-week period across an unspecified heating season. Kane *et al* 2015 [9] meanwhile found English living rooms to be generally colder during the assumed morning and evening heating periods than the SAP assumes: averaging 17.5°C and 19.0°C respectively. This was based on hourly spot temperature data from a stratified random sample of 249 homes from Leicester, UK, 93% of which were centrally heated, from 1 December 2009 to 28 February 2010. This study also included bedroom data, which found a closer agreement to the SAP-assumed 18.0°C for non-living spaces: averaging 17.1°C and 17.9°C for the morning and evening heating periods, respectively. These average figures lend support to the SAP-assumed presence, although not degree, of zoning in the temperature between rooms in English homes, although they also note that in 32% of the sample, 'the bedrooms were, in fact, warmer than the living rooms'. A study by Hulme et al 2013 [10] of the same EFUS temperature logger data used by Hughes *et al* 2010 [5] also found evidence that, across most of the heating season, living rooms were on average warmer than bedrooms and hallways, and found no statistically significant difference between weekday and weekend temperatures for homes overall or for any particular room. The same study also found evidence that achieved indoor temperatures varied over the SAP heating season, being statistically significantly lower during November to March than in October, April and May. They also found that indoor temperatures correlated with outdoor temperatures, although all the study's analyses were based on full-day mean temperatures rather than focusing just on temperatures during periods of heating, so these results could be due to the indoor temperatures dropping during non-heating periods.

Finally, in the literature review of Wei *et al* 2014 [11] of the driving factors of occupant-controlled residential space heating, the authors identify consistent findings across five relevant papers that indoor temperatures vary across the day and are correlated with room type, with living rooms being the warmest. There was little consistent evidence that temperature settings varied by day of the week, with just two reviewed papers that touched on this finding conflicting results.

Overall, the existing empirical work finds evidence of zoning between rooms, with living rooms on average being warmest, but rooms typically do not reach the setpoint temperatures assumed by the SAP, or do so only for short periods, particularly living rooms. The literature broadly concurs that there is little sign of variation in indoor temperature by day of the week, but generally highlights a large degree of variation between homes. The previously published work that we identified provides no clear evidence about if or how achieved temperatures during active heating periods vary by time of year or external temperature.

## Duration and patterns of active heating over the day

Various empirical studies have investigated actual heating patterns (or behaviours). The Shipworth *et al* 2010 study described above [6] estimated from the room temperature data that central heating hours per day during the heating season were a mean of 8.2 (S.D. 1.5, median 8.2) on weekdays and a mean of 8.4 (S.D. 1.5, median 8.4) at weekends. The participants' self-reported heating hours were somewhat higher, at a mean of 9.8 (S.D. 5.4, median 8.0) on weekdays, and a mean of 9.8 (S.D. 5.2, median 8.5) on weekends. Hughes et al (2016) [5] meanwhile, using the EFUS temperature data and a manual data inspection method rather than an automated rule-based method for identifying heating periods, estimated heating-season heating periods to be a mean of 9.8 hours per day (SD 4.3, median 8.8) on weekdays, and 10.4 hours per day (SD 4.3, median 9.7) for weekends.

Looking at the timing of heating over the course of the day, Hanmer *et al* 2019 [8] argued that a variety of standard 'thermal routines' would be expected, as each household's particular routine is shaped in part by wider societal diurnal rhythms around work, sleep, food preparation, etc. The central heating settings data that they analysed included user-programmed periods of 'in', 'out' and 'asleep'; their analysis focused on the 'in' periods, which indicated when heating systems were on. Peaks in programmed start times for heating occurred at 07:00 and 16:00, although with large variations, particularly in the evening (Interquartile Range of 150 minutes), and a median off-time at the end of the day of 22:00. At the morning peak in on-times, around 65% of homes had the heating set to on, and nearly 90% in the evening peak. Interestingly, slightly less than 60% of boilers were on in the morning peak, and just under 50% in the evening peak, with around 25-30% on at any given time in-between (as the authors note, boilers do not necessarily run continuously when the heating is 'on'). Across the sample, a 2-period programme setting was most common for 'in' periods, with a 1-period programme occurring about 1/3 as often, 3-period programme about 1/5th as often, and other patterns (3+ periods, always on, or always off) being relatively rare. Differences in the relative rates of occurrence of these different patterns between weekdays and weekends and over the heating season were not investigated. Using 2013 data, do Carmo *et al* 2016 [12] also investigated diurnal heating patterns, applying k-means cluster analysis to the hourly maximum heat demand loads of 139 heat-pump heated homes in Denmark, to identify common patterns. They identified two patterns of heat demand, one with a fairly flat profile but a soft morning peak and some increase in the evening, the other with a more substantial trough between a morning peak and evening rise in demand. Both variants occurred over the weekday and weekend, and across homes with varying levels of overall demand.

Further work by Huebner *et al* 2015 [13] using the same dataset described above in [7] identified four clusters of diurnal heating pattern in a stratified random sample of 275 English homes. These clusters were identified based on room temperature data rather than active heating durations, but as the data were taken purely from winter months (over the 2007-2008 winter season), there is likely to be substantial correspondence between the two. The most commonly identified cluster was a two-peak temperature pattern (40.0% of homes) - this is the most similar to the weekday pattern assumed in the SAP model (although the variation between homes in length and timings of the morning and evening peaks was not described in detail). The next most common pattern (30.9% of homes) was a flat line, with largely steady day and night temperatures. The two remaining clusters both showed nighttime declines in temperature of differing degrees until early morning, followed by rises of differing degrees until around 21:00. No analysis of variation by weekday vs. weekend was presented.

Kane *et al*'s 2015 work [9] also identified variation in diurnal patterns (again based on room temperature data, over the 2009-2010 winter season), with a double heating pattern over the day again being most common (51% of homes analysed). Single peaks were also common (33%), whilst multiple peaks (5%) and others uncommon patterns were also identified. Also identified were 11% of homes with patterns 'too inconsistent to categorise'. On average, the single and double heating period times corresponded fairly strongly with the SAP two-period weekday and single-period weekend heating patterns, with 'the median heating times [being] 07:00–23:00 (15 h) for single heating periods and 06:00–09:00 and 15:00–22:00 (10 h in total) for double heating periods'. However, there were variations in start times of several hours between homes (correlating with occupancy numbers and employment status), e.g. afternoon start times in the double heating period homes varied between 13:00 and 16:00. The authors did not find significant differences in heating durations between weekday and weekend however, and the full-week average daily heating duration of 12.6 hours fell midway between the SAP's assumed weekday and weekend durations. There was a large variation between homes 'with daily heating durations in individual homes ranging from 4 h to 22 h' (standard deviation 3.5 hours). Finally, they investigated the start of the heating season, finding broad consistence with the SAP assumption for an October start, but with a large variation between homes, between 1 September and 22 October.

Watson *et al* 2019 [14] also found evidence that heat demand varies by external temperature, and by date. They estimated heat demand for a mean sample size of 6,400 dwellings from across Great Britain covering 1 May 2009 to 31 July 2010, using half-hourly smart meter data and splitting the energy use data into space heating, water heating and other uses based on averaged figures for their proportions. They found that heat demand varied greatly over the SAP heating season, but was markedly low outside it and higher within it. Demand was also lower during more mild conditions. Peak demand was at 18:00; and the highest 'ramp rate' (increase in demand between time points) was at 07:00. The variation identified in demand over the day was not inconsistent with the two-peak SAP times, although there were not sharp transitions in demand between the SAP heating and non-heating times.

Hughes et al 2016 [5] meanwhile report data on the duration and timing of the heating season, based on the EFUS survey data. They report the mean self-reported heating season as being 5.7 months (S.D. 2.07, median 5.0) compared to the 8 months assumed by the SAP. The proportion of the sample responding that they heated their home varied per month, with the large majority using heating in the months November to February (varying between 92% and 100% of respondents), and October and March being transition months in terms of the proportion of respondents using their heating (69% and 44%, respectively). 20% or fewer heated their homes outside those months.

Wei et al's 2014 literature review [11], finally, reports that three papers reviewed consistently reported correlation between type of room and patterns of heating, with living rooms heated the most often, while all of four studies found heating less likely to be on at any given time in warmer climate areas and/or on warmer days.

Overall, the existing literature finds evidence for considerable variation between homes in diurnal durations and timings of heating use. Whilst a two-peak pattern similar to the SAP-assumed weekday pattern is common, single peak and continuous heating patterns are also identified in different works, as well as other homes showing more diversity and inconsistent patterns. Unlike the SAP assumption, there appears not to be a strong weekday-weekend differentiation in heating patterns or durations, while there is consistent evidence of variation between rooms, across the heating season and by external temperature (as well as by other factors) that are not modelled by the SAP. There is also evidence that the heating season is for many households substantially shorter than modelled by the SAP, although the degree to which this is shaped by weather conditions rather than by time of year is unclear.

## 3. Data preparation

This paper presents a variety of descriptive analyses of ambient room temperatures and durations and patterns of radiator usage for rooms from a sample of homes from the region in and around Edinburgh, UK. The derived dataset used in this paper contains the following for each home in the sample: for each room, the ambient temperature and radiator status (on or off), at a 10 minute granularity; for each day for each room, a categorical classification, based on a cluster analysis, representing the pattern of heating in that room over that 24 hours.

This section presents information about the source dataset and the processing undertaken to it to prepare the derived dataset analysed in this paper. Meanwhile, the methods of analysis of the derived dataset to produce the results presented later in this paper are described inline throughout the Results section.

### *3.1. Dataset*

The source dataset drawn upon in this paper is the IDEAL Household Energy Dataset. The data has recently been published open access [15] along with a full data descriptor [16].

The IDEAL dataset includes sensor data collected from a sample of homes from the region in and around Edinburgh, Scotland, UK (specifically Edinburgh, Lothians and south Fife), between August 2016 and June 2018. Data was collected from participating homes for between 55 and 673 days, with a mean of 286 days, median 267 days, and with the total number of homes increasing over the course of the observation period due to ongoing recruitment of households, reaching a maximum of 255 homes.

The data was collected as part of two projects funded by the UK Engineering and Physical Sciences Research Council[1]. The projects had various aims, principal among them to develop a "long-life, battery-powered, wireless sensor system providing high frequency measurements" as part of the development and evaluation of a home energy

[1] Intelligent Domestic Energy Advice Loop (grant reference EP/K002732/1) and Data-Driven Methods for a New National Household Energy Survey (grant reference EP/M008223/1).

monitoring and digital feedback system, and to "investigate residential energy demand patterns, drivers and outcomes" [16], with this current study forming one of the outputs of that work.

All homes in the projects had a range of sensor and survey data collected from them as part of their participation. As well as a range of other sensor data that were collected and are published in the dataset (notably for electricity and gas usage), of relevance for this current article are the wall-mounted sensors fitted in each room to detect ambient temperature and humidity. These sensors reported wirelessly at 12 second intervals to a basestation in the home, which then sent the data (encrypted) via the home's internet router to a secure server for the project. A subset of 35 of the homes also had 'enhanced' sensor systems installed, which included, among others, additional sensors fitted to the inflow and outflow pipes of radiators in each room to monitor radiator usage, also reporting at 12 second intervals. Ambient room temperature and humidity data were collected using standard calibrated sensors (Sensirion SHT21) integrated into the PCB of the project-designed sensorboxes, whilst radiator pipe temperatures were collected using temperature probes (DS18B20, with TRS plug) connected to additional project sensorboxes [16]. Sensors were fitted in homes by trained project technicians following a set of criteria to maintain data quality. For ambient room sensors these included placing them at around shoulder height wherever possible, and locating them to "avoid factors that could reduce their accuracy", including "avoiding placement above a radiator, close to openable windows or on external walls, or in direct sunlight" [16].

A range of other data is also provided in the dataset that is drawn on in the research presented here, including building and occupant characteristics collected via the surveys and by the project technicians who installed the sensors systems in participants' homes, and secondary data on weather conditions including outdoor temperature from local weather stations.

### 3.2. Sample characteristics

All homes in the study had gas central heating as their primary heating source, with radiators in all or the majority of rooms in the home. Homes with supplementary heating sources, e.g. electric heaters or solid-fuel or gas fires, were accepted into the project if they confirmed these were not used as major heating sources. Participating households had a variety of dwelling and occupant characteristics, including a mix of flats and houses, construction eras, numbers of rooms, numbers of occupants and incomes and age bands. Figure 1 provides a summary of these characteristics of the homes and occupants. Edinburgh is a city with a large proportion of flats and historic buildings, particularly 19th century properties, so is atypical of the wider UK housing stock. The sample itself also has a larger ratio of flats to houses than is typical for the sample area, and also of buildings from the 1850-1899 period. Also notable is that there are relatively fewer homes in lower income bands, despite the efforts of the project team to recruit households from the full range of income bands.

| | Monthly mean temperatures, °C | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Summary period | Jul | Aug | Sep | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May | Jun |
| 2017-2018 | 13.0 | 12.9 | 11.0 | 9.6 | 3.5 | 2.3 | 1.7 | 1.1 | 1.9 | 6.4 | 10.7 | 13.0 |
| Mean of most recent 10 years | 13.7 | 13.2 | 11.3 | 8.2 | 4.9 | 3.4 | 2.3 | 2.8 | 4.3 | 6.0 | 9.1 | 12.0 |
| 2017-2018 minus 10-year mean | -0.7 | -0.3 | -0.3 | 1.4 | -1.4 | -1.1 | -0.6 | -1.7 | -2.4 | 0.4 | 1.6 | 1.0 |

**Table 1:** Monthly mean temperatures in East Scotland. Top row shows monthly mean temperatures for the 2017-2018 period covered in the analyses here. Middle row shows the means for the most recent 10 years of data available (for July-December, this is 2011-2020; for January-June, this is 2012-2021). Bottom row shows difference between the two. (Data from [17] and authors' own calculations)

This paper focuses on the data collected in the final 12 months of the study, July 2017 to June 2018, when participant recruitment was more progressed and more homes' data is as such available in the dataset. Table 1 compares the mean outdoor temperatures in the region for those months to the means of the monthly mean temperatures over the most recent 10 years of data available at the point of writing - for July to December, these are the monthly means for 2011 to 2020; for January to June, they are for 2012 to 2021. The data were derived from the mean temperature values available

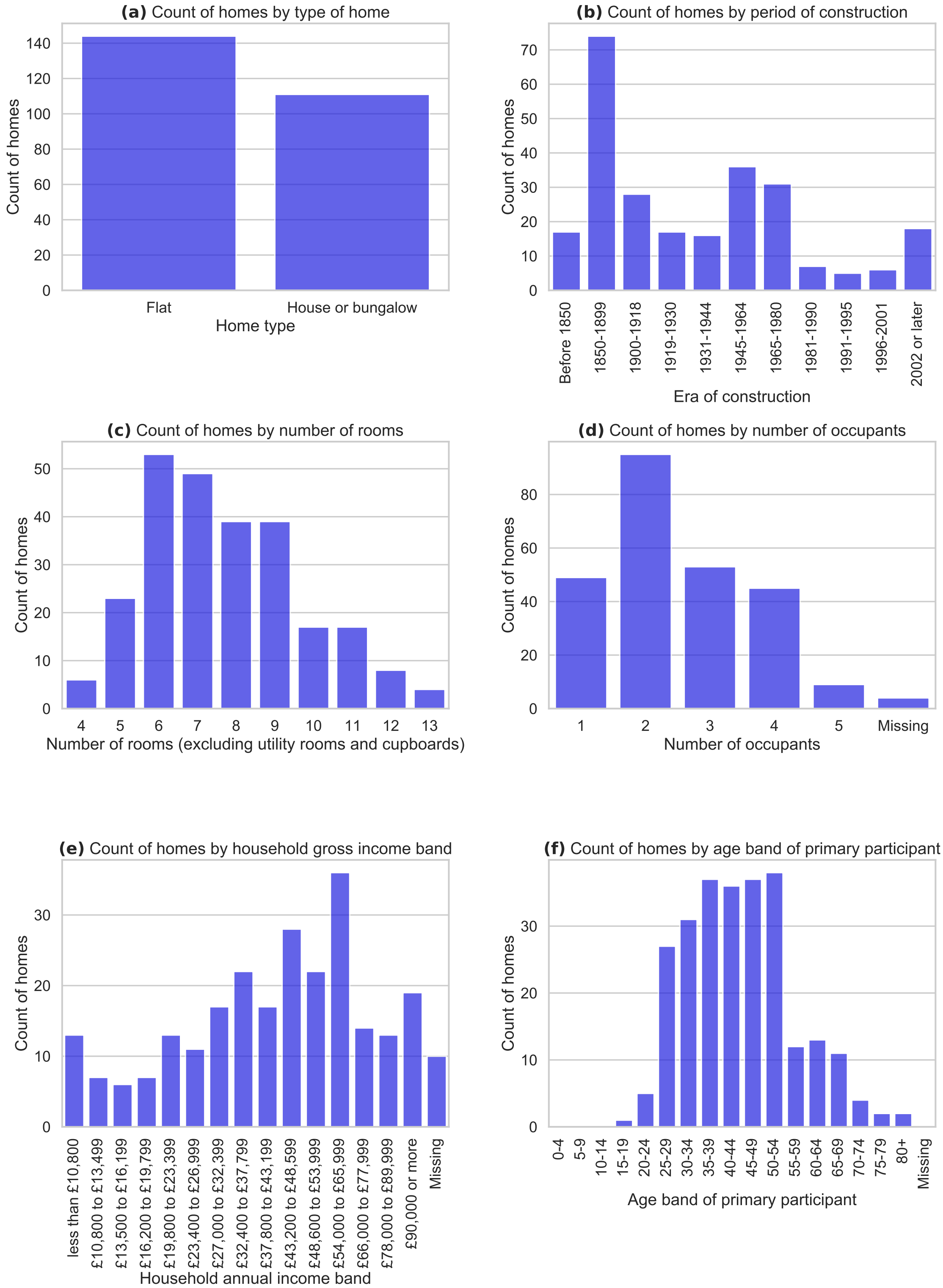



**Figure 1:** Selected building and occupant characteristics of the sample of homes included in this paper. Note varying y-axis scales. (Adapted from [16]).

from the Met Office, the UK's national meteorological service, for the 'Scotland East' region, the smallest geographic region available that encompasses our dataset's sampling area [17]. They show that the core of the heating season, November to March, was consistently colder than the 10-year average, October and April to June were somewhat warmer, while July to September were also slightly colder.

### 3.3. Preparation of room temperature and radiator usage data

We undertook various stages of post-processing of the IDEAL Household Energy Dataset to generate a dataset comprising the final set of features used in the analyses presented here. First, we downsampled the 12 second data to a 10-minute granularity, by taking the mean of the reported value. If no value was reported during a 10 minute period then we set the value to missing. Missing data points were then filled if they were within 3 data points (30 minutes) forwards or backwards of a non-missing data point. Imputed values were computed by linear interpolation between the readings immediately before and after a gap. Thus mid-sequence gaps of up to six time points (60 minutes) were completely filled, while larger gaps had three imputed values at each end (30 minutes at each end) while retaining missing data elsewhere.

For the 35 homes with enhanced sensor systems with sensors measuring radiator pipe temperatures, we used this data to label a radiator as either *on* or *off* at each 10-minute time point. Following [7], we defined a radiator to be on if its temperature was above room temperature, in this case by 5°C or more, using the mean value of the input and output pipe temperatures if both were available. To extend this data further, we inferred radiator on and off times in the rooms of the remaining homes in the dataset that lacked direct radiator pipe temperature measurements. To do this, we developed and applied a new Machine Learning methodology for inferring domestic radiator use - a deep, dilated convolutional neural network model. The model takes the available room temperature and humidity and external temperature and humidity data as inputs, and for each room and 10-minute time interval produces a label of whether its radiator was on or off, in the same format as the labels for radiators in homes that had radiator pipe temperature sensors. The Machine Learning model we used is based on approaches that have had success analysing time series data [18], including in the building energy domain [19], and is computationally efficient when there are likely to be variable time lags to be considered between the variable being predicted (in this case, the status of the radiator in a room as either on or off) and the variables used to make the predictions (in this case, room temperature and humidity and external temperature and humidity). The model was trained and validated on the homes with enhanced sensor systems, and the full methodology and its performance evaluation are described in our methods paper [20]. Briefly, the model was evaluated for its ability to predict if the heating was on or off for each 10 minute time period (*bins*) and to predict longer contiguous periods when the heating was on (*events*). Over the heating season, it achieved an overall precision and recall of 0.74 and 0.81 respectively per *bin*, and a higher precision and recall of 0.83 and 0.82 respectively per *event*. Overall, the model gave a good prediction of the average duration of heating events and the quartiles and overall distribution of heating durations, with "some underestimation of the proportion of days with short and long heating durations" - short heating durations were more likely to be missed; long were more likely to be slightly underestimated in duration. Performance was fairly consistent between rooms and, with the exception of slightly poorer performance for kitchens, between room types. The model was most likely to fail to predict short heating periods of less than an hour, presumably as they are too short to increase the ambient room temperature sufficiently. These short missed events reduce the model's precision and recall, however they represent heating events with comparatively little effect on room conditions and so are of less empirical interest. A further factor is that the model "detects heating of any kind, whereas the labels used with the demonstration dataset are exclusively for radiator use"; as such, true heating events detected by the model that arose from "additional heat sources, such as electric radiators, open fires, heat transfer from other rooms through open doors and heating from direct sunlight entering the room" would be counted, erroneously, as false positives, lowering the reported precision of the model below what it actually should be [20].

The resultant dataset used in this study therefore comprises a blend of homes with direct and inferred measures of room radiator on and off times. Adding inferred measurements increases the level of error in the dataset to a degree, but greatly increases the number of households for which data are available. The overall relatively strong performance of the inference model and its consistency across room types and at inferring all but the shortest and longest heating events, mean that on balance the inferred heating data for the 220 homes without direct radiator temperature measurements represent a valuable addition to the sensor-measured data for the other 35 homes in this study, with the nature of the errors introduced meaning they are likely to have minor substantive impact on the results and conclusions in this current study.

### 3.4. Clustering of room-day heating patterns

The final feature produced for the dataset used in this study is a label for each day for each room (each 'room-day') describing its diurnal pattern of radiator usage, i.e. the pattern of the radiator being on and off over the course of the day, starting from midnight. A cluster analysis was undertaken to identify common patterns of heating of rooms over the course of individual days, and to label each room-day of data with the cluster into which its heating pattern fell. As such, a room can potentially change clusters from one day to the next, and rooms within a home on any given day may potentially fall into the same or different clusters.

We undertook the clustering with HDBSCAN [21], which is a hierarchical density based algorithm. A wide range of clustering algorithms exist; we selected HDBSCAN because it has two characteristics that make it well-suited to the current study: firstly, it does not require the number of clusters to be specified *a priori*, and secondly, it incorporates a concept of noise - that is, some cases can be considered too different from any of the identified clusters to be allocated to any of them. These characteristics of HDBSCAN are valuable for the current study because, based on the literature reviewed earlier, we firstly do not have a strong theoretical or empirical basis for deciding the 'correct' number of clusters in advance, and secondly, alongside a limited set of commonly occurring heating patterns, we would also expect a wide range of heating behaviours that occur only occasionally, which will be represented as noise by this algorithm rather than being allocated to clusters that they only distantly resemble.

Three input features were created for clustering upon, derived from the 144 10-minute resolution time-steps (*bins*) that represent radiator on- and off-times across the course of each day for each room. The features were: (1) the total heating duration per day (the sum of all bins for the room-day when the heating was classed as 'on'), (2) the average duration of heating *events* (where an *event* is defined as a contiguous period of 10-minute bins during which a particular room is continuously labelled as having its radiator on ), and (3) the *centre of mass* of the 144 bins, defined as the median time across the day when the room's radiator was on. For example, a day heated for the full 24 hours, or with no heating at all, would have a centre of mass at 12:00 (midday). A home with heating from midday to midnight only would have a centre of mass at 18:00. These three features were standardised (subtracting the mean and scaling to unit variance) before applying the clustering algorithm. Producing these three features from the original 144 bins is an important step in enabling the clustering algorithm to identify underlying similarities and differences between room-days, i.e. to identify clusters. Using the 144 bins directly would mask the clusters, as including large numbers of variables in a cluster analysis prevents clusters being identified, as with more variables, each data point increasingly appears equally (dis)similar to each other data point, the so-called "curse of dimensionality" [22]. The choice of the above three input features is intended to retain information about important aspects of the heating patterns in each room-day.

We ran the algorithm with the minimum cluster size set to 1,000 (i.e. no clusters were permitted if they comprised fewer than 1,000 room-days) and the "minimum number of samples" to 45. The minimum number of samples is a parameter of HDBSCAN that effectively controls the level of noise by defining how many points have to be within a given distance to be counted as "core points", as defined by the DBSCAN terminology - the higher the value, the more cases will be classed as noise, and clusters will be progressively restricted to more densely populated areas of feature space (see [21] and [23] for detailed information). The input data was all the room-days falling into the 2017-2018 heating season[2]. The 2016-2017 heating season was omitted as the majority of homes do not have data for that period.

The clustering algorithm returned eight clusters plus one noise "cluster", which are shown in figure 2.

Room-days falling into cluster 0 used their heating throughout the day and night, while room-days in cluster 1 used heating from around 7am to 10pm. No heating is observed for cluster 2. Room-days in clusters 3 and 4 have their heating turned on either in the evening or in the morning, respectively. Clusters 5, 6, 7, and 8 are characterised by a two-peak pattern of heating in the morning as well as in the evening, with varying degrees of heating use during the day. While these eight clusters emerge from the clustering as different, we manually grouped them into four groups. This was based on our judgement of how similar these are with respect to our understanding of behavioural patterns and is further corroborated by inspecting the feature distribution of each cluster (c.f. figure 3). We describe the four

[2] In our analyses, we identified a core heating season from the beginning of November to the end of March. October and April were apparent as transition phases where heating was used to some degree, while the remaining months were periods with minimal levels of heating use. Throughout the paper, where we summarise for the heating season, we use an empirically-driven definition, taking it to span from October to April inclusive, rather than the SAP assumption of October to May, so that it includes the core and transition heating periods found in our dataset but excludes periods with very little observed heating use. Section 4.2 presents the relevant results on heating usage per month.

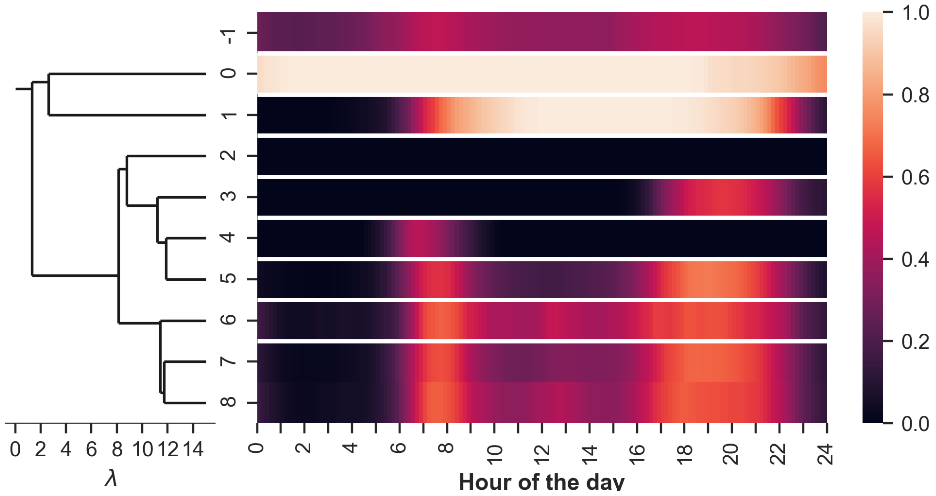



**Figure 2:** Heating pattern clusters. The cluster labels as returned by HDBSCAN are shown on the y-axis (-1 denoting the noise cluster). The horizontal bars for each cluster are heatmaps showing the proportion of rooms-days in the cluster that were being heated at each time point across the day, drawing on the 10-minute radiator data for all the room-days within each cluster. The dendogram shows the hierarchical splits undertaken by the algorithm and the corresponding $\lambda$ value when each cluster split off, which provides an indication of closeness or similarity between each cluster (splits at higher $\lambda$ values indicate more closely related clusters).

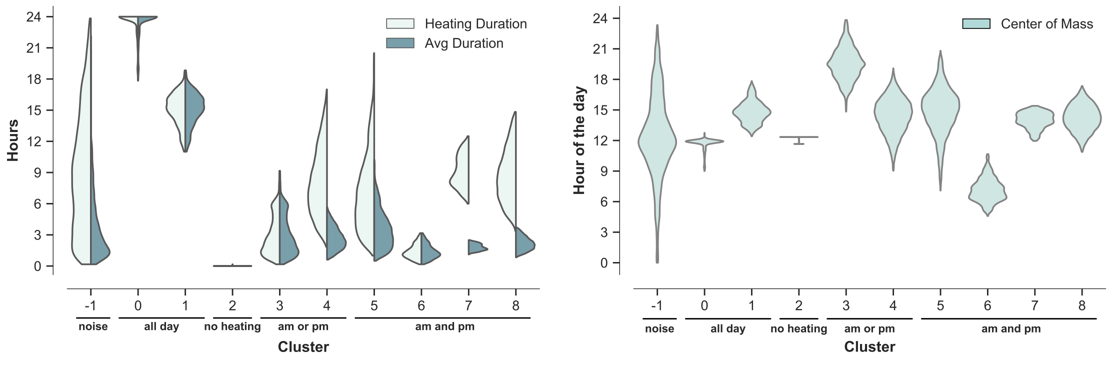
groups as: (i) *all day heating*, (ii) *no heating*, (iii) *am or pm heating*, and (iv) *am and pm heating*, plus the *noise* cluster. The group assignment is summarised in table 2.

| Cluster group | Cluster IDs |
|---|---|
| All day heating | 0, 1 |
| No heating | 2 |
| am or pm heating | 3, 4 |
| am and pm heating | 5, 6, 7, 8 |
| Noise | -1 |

**Table 2:** The nine clusters returned by HDBSCAN were each manually assigned to one of four groups (plus a *noise* group). The table indicates the descriptive name allocated to each group, along with their respective clusters.

### 3.5. Graphical presentation of results

A range of graphical approaches are used in the figures in this paper to present the results. Where the values of a single variable are being discussed, either for the sample as a whole or for subsamples, figure styles are tailored to the key characteristics of interest. To present totals and differences, *bar graphs* (Figures 1 and 14), or *stacked line graphs* (Figure 12) are used. *Boxplots* are used when means and spread (e.g. standard deviations) are also of interest (Figures 4 and 5). Where more detail of the distribution is required than can be revealed by a boxplot, such as when a variable's values deviate strongly from a normal distribution, then *line graphs* are presented that present similar information to a histogram but smoothed based on an estimate of the underlying distribution using a kernel density estimator (KDE) (Figure 8). These can be further enhanced into *violin plots*, which allow distributions calculated in the same manner for multiple variables to be plotted side by side or mirrored for a single variable (Figures 3 and 7) and can additionally present further information on the means and ranges of the values (Figures 11 and 14). Where correlations between two continuous variables are presented, we utilise *line graphs* (Figure 10). Finally, where the correlation between three




(a) (b)

**Figure 3:** Violin plots showing the distributions of values for each of the features used for clustering, across all room-days in each cluster. (a) shows the distributions of the total heating durations over full room-days (left hand side of each violin), and the distributions of the average durations of the individual heating events within each room-day (right hand side); (b) shows the distributions of the "centres of mass" - the median times of the 10-minute bins when the heating was on each room-day, counting from midnight. The group assignments as used in this study are indicated below HDBSCAN's cluster labels.

variables is being discussed, we utilise variants of *heat maps*, which use a colour scale to show the value of a variable across its range of values as it varies against two other variables, which are plotted on the x and y axes (Figures 2, 6, 9 and 13).

## 4. Ambient room temperatures and heating usage

### *4.1. Achieved ambient room temperatures*

Here we explore the temperatures achieved in living rooms[3] across our sample. The SAP assumes 21°C is achieved for nine hours per day on weekdays (and 16 for weekends), and so we focus on the achieved temperatures for the warmest nine hours of each day, irrespective of where in the day these data points occur (i.e. they may not be contiguous, or overlap with the precise periods of the day the SAP assumes to be actively heated). To achieve this we rank data points for each room-day by temperature. The minimum temperature reached during nine hours of the day then corresponds to the 62.5 centile ($1-9/24$), while the median temperature over the warmest nine hours corresponds to the 81.25 centile. We focus on the warmest periods of each room-day rather the specific heating times assumed in the SAP model, as this is a simplifying assumption in the model and real periods of heating use will vary between households and between days. Additionally, we analyse weekend and weekday data together here, as we find, consistent with other literature, little difference in heating durations between weekday and weekend (see 'Levels of active heating per day', below).

Figure 4 shows boxplots of the minimum temperatures for the warmest nine hours of each day for living rooms across all homes, split by month of the year. The figure also shows boxplots of the mean temperatures achieved during those same nine hours of each day. Across the heating season, it can be seen that the average minimum temperature is around 19°C rather than 21°C. The mean temperature is also below 21°C, at around 20°C, indicating 21°C is commonly reached for less than half of the time assumed by the SAP. The actual amount of time room-days are at a temperature of 21°C or above is shown in the boxplots in figure 5, which demonstrates that on very few room-days are living rooms heated to 21°C or above for the full nine hours assumed in the SAP. In fact, across the heating season, the figure indicates that the majority of room-days achieve 21°C for no more than an hour or even less, and that rooms heated to 21°C for nine hours or more are outliers.

[3]Note, in this paper we take the living area to be the living room/lounge for all homes in the study, consistent with the SAP definition.

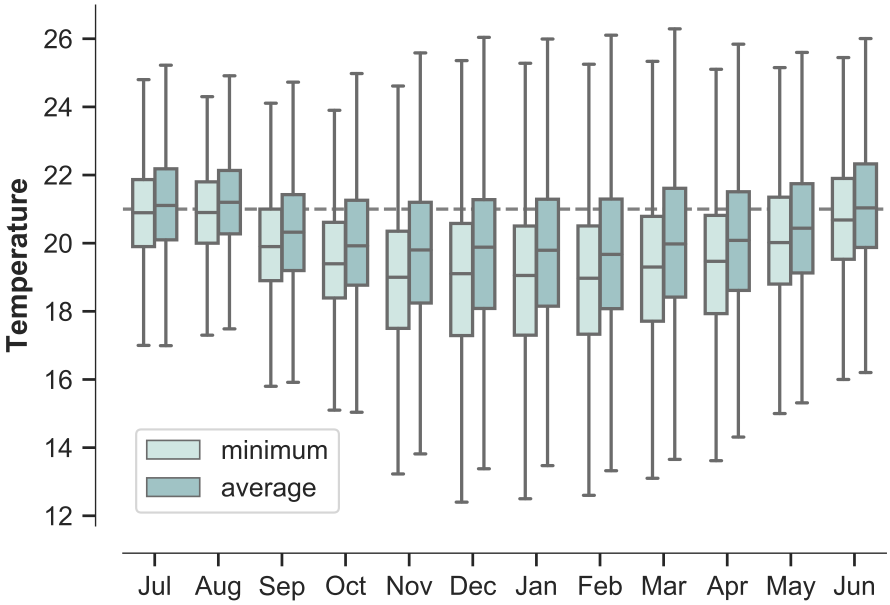



**Figure 4:** Boxplots showing the distribution of minimum temperatures reached for at least 9 hour per day (62.5 centile), and average (mean) temperatures over those same 9 hours, in living rooms in 2017/2018. The dashed line in the background indicates the 21°C assumed by SAP.

The two figures show that the actual achieved temperatures show a high level of consistency across the core heating season (November to March), i.e. little variation in the minimum and mean temperatures achieved for nine hours per day, or in the duration of time rooms are heated to 21°C or above. An increase in the average temperature reached, and the duration of time spent at 21°C or above, is only observed for the warmer months of the year (May to September) which are no longer considered to be part of the heating season.

Although the SAP model assumes the achieved temperature is unaffected by heating patterns or outside temperature, we find some relationships between these. This is highlighted in figure 6. Figure 6a shows a hexbin plot of the 62.5 centile room temperatures against the number of hours the radiator was used during the same day. It can be seen that for long heating periods, particularly above 15 hours per day, the achieved temperature starts to rise. There could be a range of explanations for this. Occupants could either desire, or be indifferent to these higher temperatures, or they may have difficulty controlling their heating system. For shorter heating durations of less than 15 hours per day, the minimum achieved room temperatures during the warmest nine hours of the day is usually below 21°C.

Figure 6b meanwhile shows a hexbin plot of the 62.5 centile room temperatures against mean outside temperature for the same day. This reveals greater spread in the achieved temperatures at lower outdoor temperatures, and some signs of relative overheating on warmer days.

### *4.2. Duration of active heating per day*

The SAP model assumes that rooms are heated for nine hours in total per day on weekdays, and 16 hours per day at weekends, with no difference in these figures between different room types or over the heating season.

Figure 7 shows violin plots of the distributions of heating durations between weekday and weekend and by room type. The left-hand two plots show heating durations over the heating season split by weekdays (left) and weekends (right). There is no substantive difference between the two distributions, and the mean value is 6.1 and 6.0 hours of heating per day for weekday and weekend respectively. The right-hand plots present heating periods over the heating season by rooms: firstly between the living area (mean 6.6 hours) and non-living area (mean 6.0 hours) and secondly between room types. Overall, the distributions indicate that non-living areas are more likely to be left unheated,

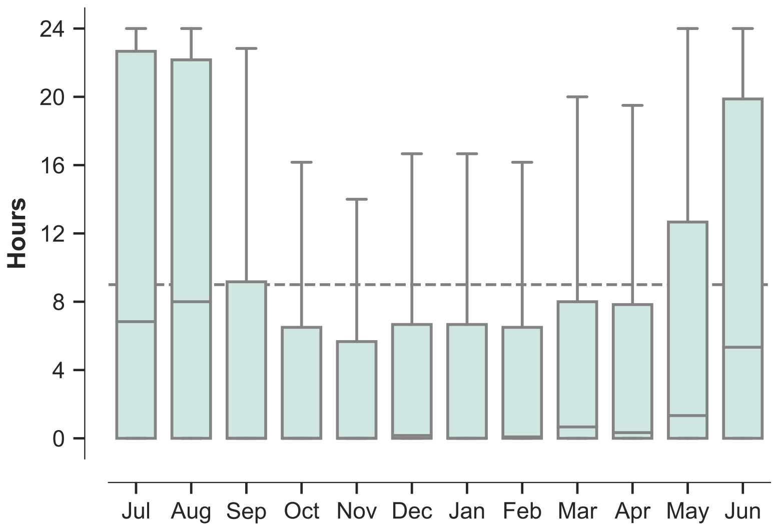



**Figure 5:** Boxplots showing the distribution of hours per day reaching 21°C or above, in living rooms in 2017/2018. The dashed line in the background indicates the 9 hours assumed by SAP.

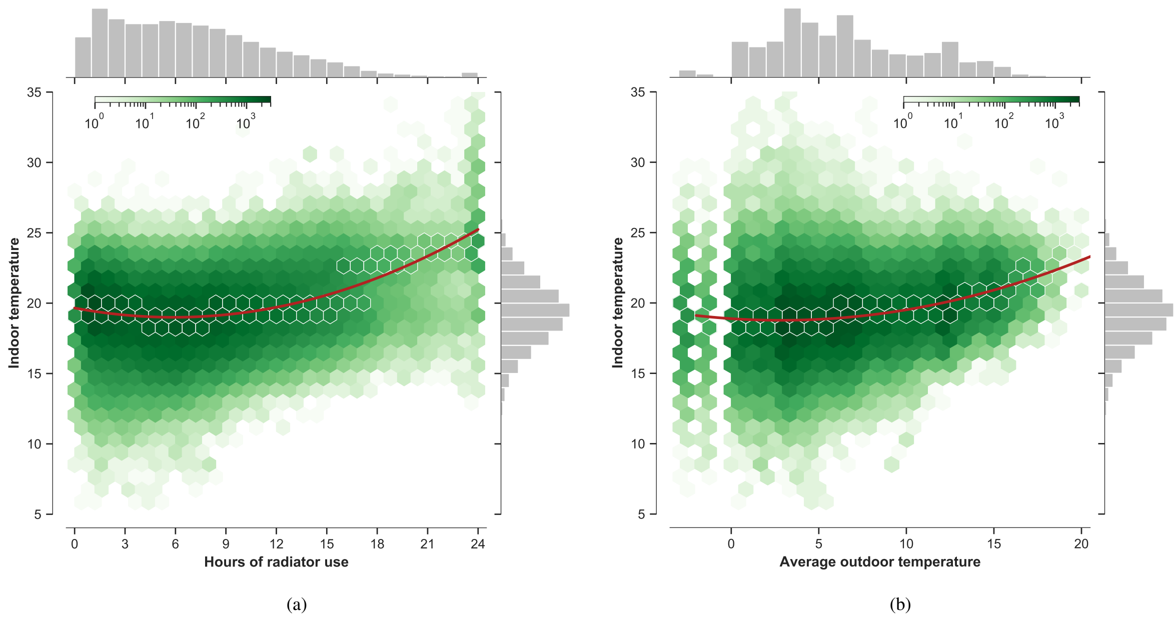



**Figure 6:** The relationships between (a) minimum indoor temperature reached during nine hours per day and hours of radiator use, (b) minimum indoor temperature reached during nine hours per day and average outdoor temperature. Only days during the heating season and days for which active heating was observed are included. The hexagonal bin colours indicate the number of room-days across the sample falling at that point. Bins with the maximum number of observations along the y-axis are indicated with a white border. The red lines show cubic interpolations for these bins.

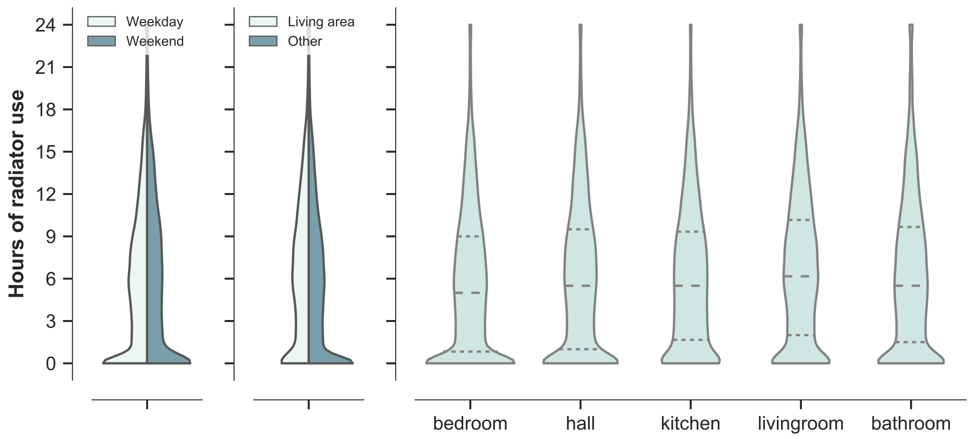



**Figure 7:** Distribution of heating durations per day for the heating season. The two-toned plots on the left show the distribution split by weekday and weekend, and by living area and all other room types, respectively. The plots on the right show a breakdown by room type. The dashed lines represent the quartiles of the respective distribution.

particularly bedrooms, but the differences in distributions between rooms are small, indicating little in the way of zoning of the duration of heating.

Breaking heating durations down into separate months of the year highlights substantial differences in the distribution of heating hours over the heating season, as can be seen in figure 8 (which pools data from weekdays and weekends and from all rooms). The figure also shows that heating is significantly used from around November, with October being a transition period where some heating is already observed. From around April, households transition to no longer requiring heating, leading to very little observed heating from May, about a month earlier than assumed in the SAP.

Figure 9 indicates that this seasonal trend is at least in part related to the corresponding changes in outdoor temperatures. The figure shows the correlation between mean outdoor temperature and hours of radiator use per day for all room-days across the heating season. As might be expected, generally lower levels of radiator use are found on days with higher outdoor temperatures. Meanwhile, when outside temperatures are lower, there is a large spread in hours of radiator usage. This increasing spread with decreasing external temperatures is unexplained. It may be explained by diversity in occupants' physiology through variation in the width of their thermal neutral zones, or in variations in their behaviour and thermal comfort practices. Occupants who wear more clothes in winter may be equally comfortable at lower internal temperatures. It might indicate the effects of occupants zoning - heating different rooms to different levels, such as for energy efficiency motivations. It may also be due to lower income or fuel poor households using less heating than higher income households because of cost factors. It also demonstrates that it is highly likely that heating periods will vary substantially from year to year based on annual variations in weather conditions.

### *4.3. Diurnal patterns of active heating*

The SAP model assumes heating to be on from 07:00-09:00 and 16:00-23:00 for weekdays, and 07:00-23:00 for weekends over the heating season, with no variation between rooms or across the heating season.

Figure 10 (top) plots the proportion of rooms in the study which were actively heated at different times of the day, across the whole heating season, showing weekday and weekend data separately. Whilst weekday peaks in heating coincide approximately with the SAP assumption, it can be seen that there remains substantial variation, with only around half of rooms across the homes in the sample heated at the peaks. Also, around a quarter of rooms remain heated during the middle of the day, outside of the two periods of heating assumed by the SAP. The weekend shows a similar pattern, but with lower peaks, more heating between the peaks, and a morning peak around half an hour later than the weekday one.

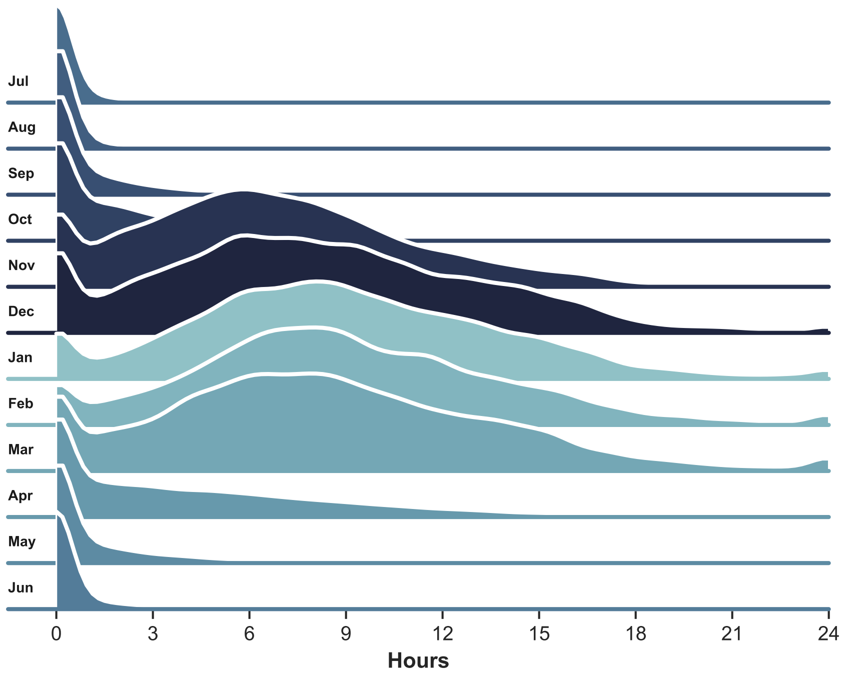



**Figure 8:** Distribution of the daily heating durations observed for room-days, split by month. The small "bumps" towards 24 hours of heating in colder months are the result of room-days with actual 24 hours radiator on-times.

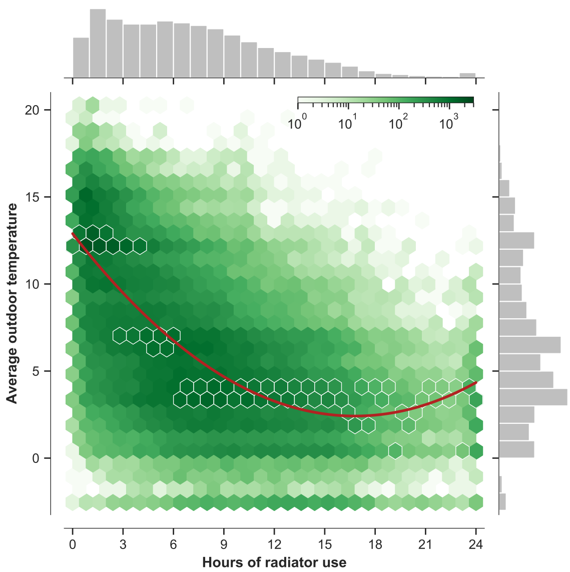



**Figure 9:** The relationship between average outdoor temperature and hours of radiator use. Only days during the heating season and days for which active heating was observed are included. The hexagonal bin colours indicate the number of room-days across the sample falling at that point. Bins with the maximum number of observations along the y-axis are indicated with a white border. The green line shows a cubic interpolation for these bins.

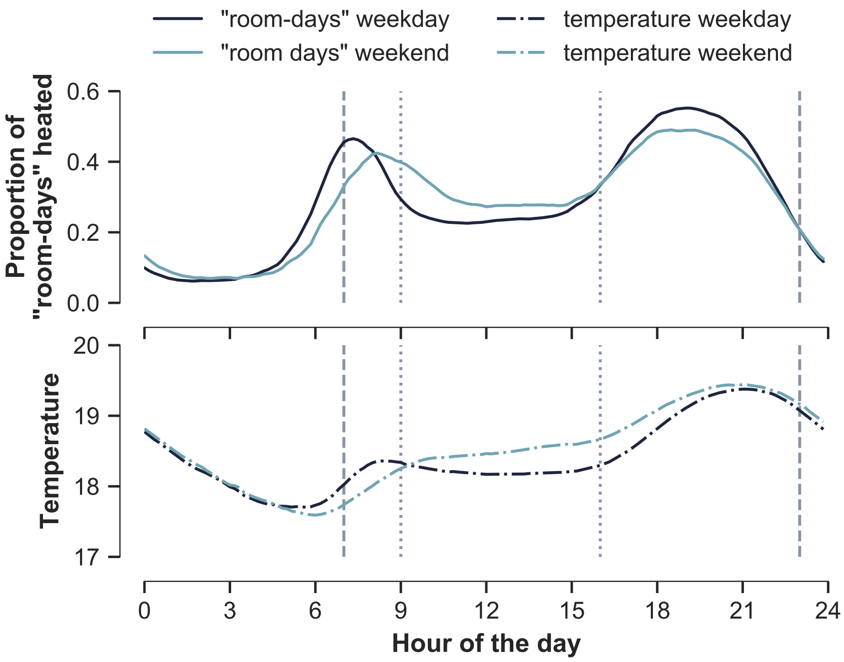



**Figure 10:** Proportion of rooms with heating on (top) and average indoor temperature (bottom) by hour of the day. The dashed and dotted lines indicate the heating periods per day as assumed by SAP.

The bottom of the figure shows the average indoor temperatures achieved across the day for the same set of room-days. These indicate that weekend achieved temperatures average a little higher than on weekdays during the heating season, except that there is a later start to the rise in temperature, which is likely explained by the observed differences in weekend heating patterns.

These aggregated figures reveal overall patterns but also obscure between-room-day variation. Our cluster analysis (described in section 3) identified four common patterns of daily heating. *All day heating* corresponds approximately to the SAP pattern of heating that it assumes is observed at weekends, although our cluster has a broader definition, encompassing days that are heated throughout the SAP heating period of 07:00-23:00 and which may or may not have further heating outside of those times; the *am and pm* cluster corresponds approximately to the SAP pattern assumed to occur on weekdays; while *No heating* is only assumed in the SAP model to occur outside the heating season; and the *am or pm* cluster has no direct equivalent in the SAP model. A further *noise* cluster captures a range of other patterns that each occur only infrequently and do not align sufficiently closely to any of the other clusters to be labelled as one of those.

| | *Assumed by SAP* | *Empirical results* | |
|---|---|---|---|
| Heating cluster | % of room-days | % of room-days | % of homes |
| No heating | 0% | 15 | 98 |
| am or pm | 0% | 11 | 100 |
| am and pm | 71% (weekdays) | 50 | 100 |
| All day heating | 29% (weekends) | 4 | 73 |
| Noise | 0% | 21 | 99 |

**Table 3:** The percentage of room-days falling into each pattern of heating, and the percentage of homes having at least one room fall within that cluster on at least one day.

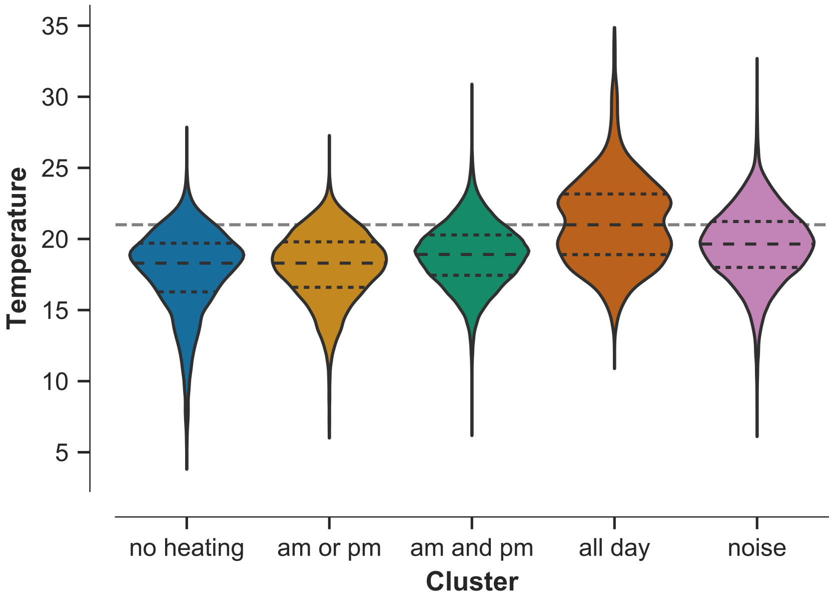



**Figure 11:** The distribution of the 62.5 centiles of the daily temperatures is shown. This corresponds to looking at the minimum temperature which is reached during nine hours per day. The dashed line in the background indicates 21°C. Mean and quartile ranges of the data are shown by the dashed lines in the violinplots.

Table 3 presents data on how commonly each of the heating pattern clusters occur in the homes, as a percentage of total room-days and as a percentage of homes in which that cluster is present at least once. The SAP model effectively assumes 5/7th of room-days (71%, all weekdays) fall into the *am and pm* pattern and 2/7th (29%, all weekends) fall into the longer *all day* pattern. We find that a substantially lower proportion of room-days, in this case 50%, falls in the two-peak, *am and pm*, pattern. Only 4% of room-days fall into the *all day* cluster. 11% of room-days have heating just in the *am or pm*, 15% have *no heating*, and a further 21% are in the *noise* cluster. Similarly to previously published empirical work, we did not find that the cluster into which a particular room-day fell correlated substantially with whether that room-day was on a weekday or a weekend. With the exception of the *all day* heating cluster, virtually every home had at least one room-day in each of the other clusters.

While the results of this study confirm that the most prevalent heating periods are observed in the morning after people tend to get up and in the late afternoon and evening, it further highlights that there remains a substantial degree of heating occurring between these periods even on weekdays, and that there is substantial variation between room-days, with almost half having heating patterns that are neither the *all day* nor the *am and pm* patterns assumed in the SAP.

#### *4.3.1. Heating patterns and room temperature*

We investigated whether there was a correlation between heating clusters and achieved room temperatures. Figure 11 shows the distribution of minimum room temperatures reached during the warmest nine hours of each day for room-days within each heating cluster.

The distributions in temperatures reached are similar for most clusters, including the *no heating* one. The exception is room-days in the *all day* cluster, which achieve a higher temperature on average. This indicates that, while there could be rooms which need continuous heating due to insufficient insulation, rooms with continuous heating are instead more likely to be heated to a higher temperature. This in turn implies that people whose homes are heated more are (on average) achieving higher indoor temperatures and not always simply compensating for higher heat loss due to lower outdoor temperatures or poor insulation. The reason for such *all day* heating patterns is unknown. It could arise from occupant choice, occupant indifference, or an inability to control heating times.

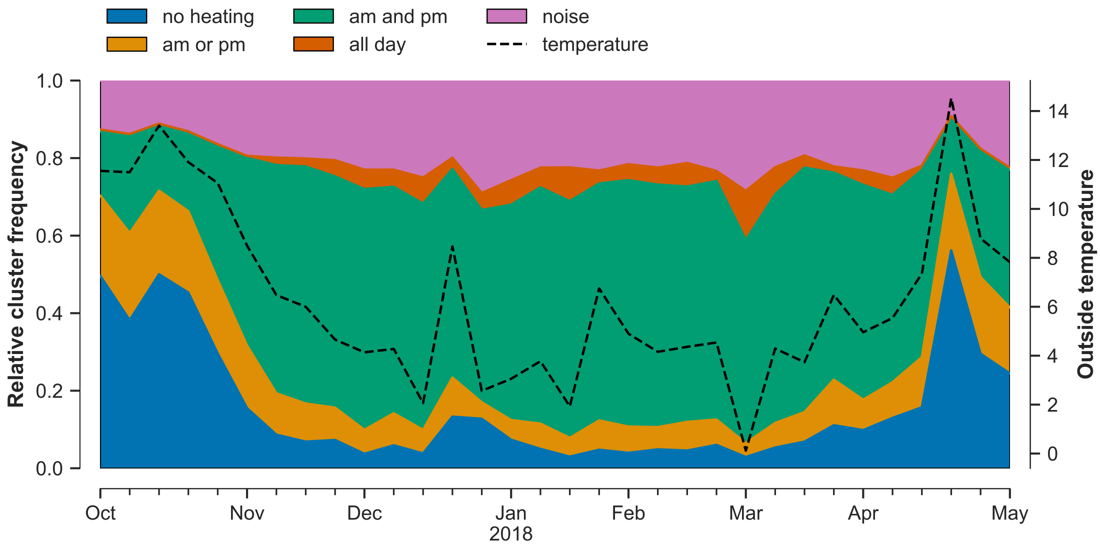



**Figure 12:** Relative frequency of room-days per week is shown for the heating season 2017/2018. The dashed line shows the average outside temperature as measured in the City of Edinburgh for that time.

### *4.3.2. Heating routines and change over time*

The data also demonstrates that heating patterns change over time. Figure 12 shows the changes in relative sizes of the heating clusters over the 2017/2018 heating season, as a proportion of room-days in each period. A clear adaptation of heating patterns to outside temperature is apparent. Room-days which are heated either in the morning or the evening are mainly found in the transition periods (October and April), characterised by higher average outdoor temperatures relative to the rest of the heating season, while room-days which are heated more (either continuously or in both the morning and afternoon) are predominantly found in the core heating period of November to March inclusive. It can furthermore be seen that during a particularly cold period in March, the number of room-days using heating throughout the day increased slightly. The *no heating* pattern is also strongly associated with temperatures, and as such is most common during the transition periods.

As well as these seasonal changes in heating patterns, we investigated patterns of change in heating patterns from one day to the next. Such changes could arise due to different householder schedules on different days, adaptions to rapidly changing weather conditions, and so on. We computed the transition probabilities of rooms switching between clusters from one day to the next. The full transition probability matrix is depicted in figure 13. Across nearly all the clusters, the most likely outcome for a room is for it to continue in the same cluster on one day as it was in the previous day. However, the likelihood varies by cluster.

Rooms falling into the *noise* cluster remain in the same cluster or switch to the *am and pm* cluster with roughly equal probability. This indicates that the *noise* cluster shares similarity with the *am and pm* cluster, leading to a fuzzy border separating these two clusters in the feature space. This is further corroborated by the relatively high switching probability from the *am and pm* cluster back to the *noise* cluster.

It can further be seen that *no heating* is the most stable pattern with a probability of 69% for a room to remain unheated on the next day.

Other transitions in the matrix correspond to an increase or decrease in the level of heating. Firstly, a switch away from *no heating* is observed with 17% probability to *am or pm*, with 9% probability to *noise*, and 7% probability to the *am and pm* cluster. Once in the *am or pm* cluster, there is equal probability of subsequently remaining in *am or pm* or switching to *am and pm* (30-31% each), and a lower probability of switching to *all day* or *noise* clusters (2% and 17% respectively). This indicates that *am or pm* is a relatively unstable heating pattern, which is consistent with the relatively low rate of occurrence of this cluster seen earlier.

A gradual decrease in heating use seems further to be reflected by the high transition probability of the *all day*

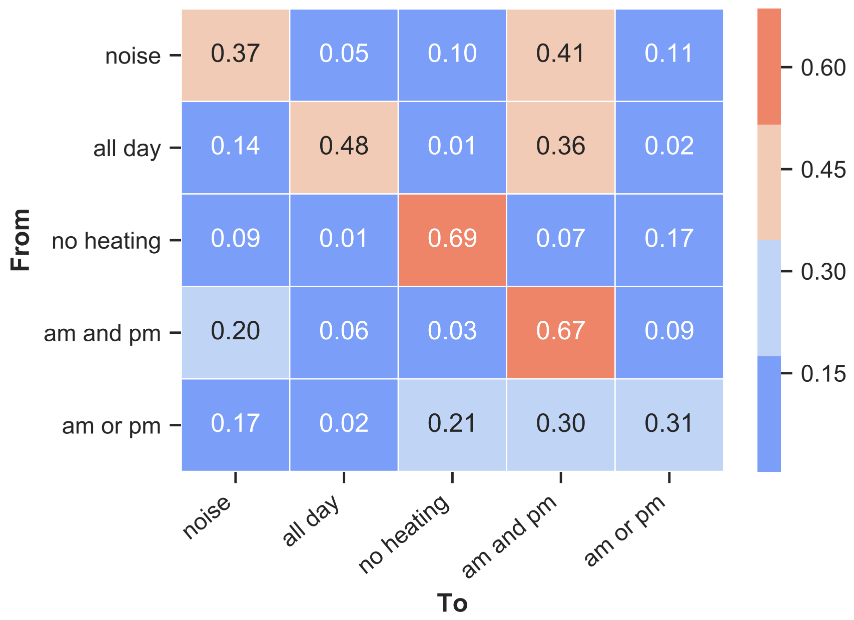



**Figure 13:** Transition matrix, indicating the probability of a room transitioning from a given heating cluster to any of the others between days (see main text for full details).

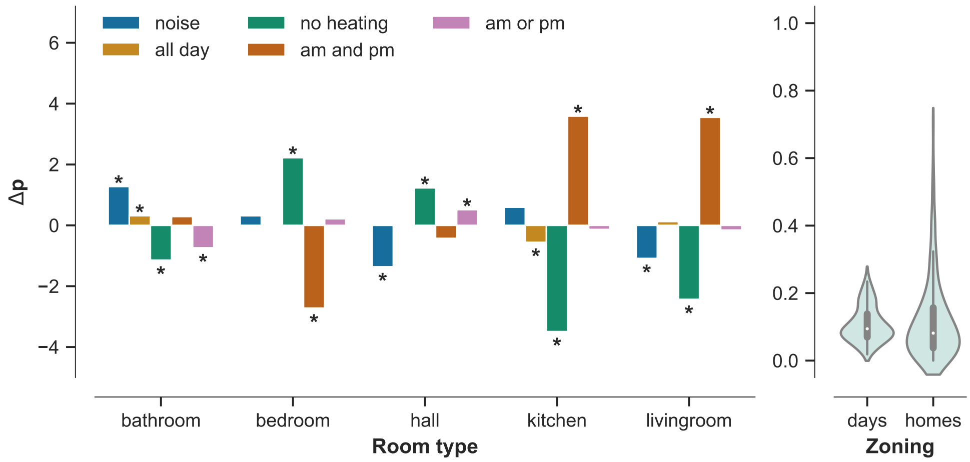
heating cluster to the *am and pm* heating pattern.

#### *4.3.3. Heating routines and "zoning"*

Where homes have different heating needs in different room types, householders might decide to "zone" their homes. Here, we investigate if room types show different prevalence to the various clusters as well as to what extent the heating patterns between bedrooms and living rooms differ.

If no interaction between cluster assignment and room type is present, we would expect the relative frequency of clusters (or probability that a random room-day falls into a specific cluster) to be equal between the complete dataset and each room type respectively. If, on the other hand, certain room types were more likely to be found in a specific cluster, this cluster would show a higher relative frequency for that room type compared to the complete dataset (and vice versa). Figure 14a shows the difference between the probability of a room-day falling into a cluster given the room type and the probability of a room-day falling into that cluster irrespective of the room type. A value larger than zero indicates a higher prevalence for the room type to be in the respective cluster compared to the complete dataset. The significance of this difference in probabilities is computed using a two-sided binomial test, assuming the true probability is as observed in the complete dataset and the outcome of the test is as observed for the respective room type. It can be seen in figure 14a that bedrooms have a higher probability of not being heated at all and a lower probability of being heated am and pm. This trend is reversed for kitchens and living rooms, which tend to be more likely to be heated am and pm and less likely to not be heated at all compared to other room types. While minor differences between room types are observed, there does not seem to be a striking difference in how rooms of different types are heated.

We further looked at differences in the heating patterns found in bedrooms and living rooms as an indicator of zoning. If the heating cluster which the living room was in for a particular home differed from the cluster at least one of its bedrooms was in on the same day, we assumed the householder to have performed some form of zoning. While this will not give a true indication of zoning as the temporal aspect of when heating is used is too coarse, it gives an estimate of differing heating patterns between these two room types that captures larger differences. We computed if a home was zoning as defined above, for each day of the heating and transition periods. Days for which the living room as well as all bedrooms were found in the *noise* cluster were excluded from the analysis. Figure 14b shows the distribution of the percentages of homes that performed zoning per day (left) as well as the distribution of the percentages of days a




(a) (b)

**Figure 14:** (a) Difference in observed probabilities of falling into a cluster given the room type. The difference in the probability of observing a cluster given the room type and the probability of observing that cluster given a random room type is shown ($p(Cluster|RoomType) - p(Cluster)$). (b) Violin plots showing the distribution of zoning for two cases: *days* shows the percentage of homes that do zoning on any given day; *homes* shows the percentage of days a single home does zoning over the full heating season. The first gives an understanding of variability in the proportion of homes zoning on a given day; the second highlights the variability between homes' propensity to zone over the heating season. As data points of the violinplot represent a ratio, points with less than 25 underlying events are excluded from the analysis.

single home performed zoning over the full heating period (right). We found that on any given day only around 10% of homes performed zoning, as defined here. Similarly, any given home was found on average to perform zoning for around 10% of days over the heating season, although the spread between homes is larger compared to the variability over time. We also observed a small variability in zoning probabilities with respect to the time of year, with the average level of zoning observed being slightly higher during the transition periods (data not shown).

## 5. Discussion

This paper has presented new empirical data and analyses of room-level heating patterns and achieved temperatures for a sample of homes in the Edinburgh region of Scotland, UK. The results highlight some areas of concurrence and others of substantial difference with the simplifying assumptions in the SAP model, which are discussed here in more detail.

In terms of achieved ambient temperatures, the SAP assumes a consistent 21°C is achieved in living spaces over the active heating periods. Even focusing on just the warmest nine hours of each day over the whole week, i.e. not necessarily those heating times assumed in the model, our results indicate that very few room-days in the sample of homes maintained this temperature over the full period. Instead, the minimum temperature achieved during the warmest nine hours of each day in each room averaged around 19°C, with the mean temperature over those times slightly higher, broadly consistent with the findings in the existing literature for English and UK homes. Indeed, the majority of room-days over the heating season achieved 21°C for no or nearly no time at all. Whilst this result may appear slightly lower than found in previous literature, this may be because the results are not directly comparable. Averaged across all rooms in the sample, temperatures do rise with the onset of the SAP-assumed morning heating period (from around 07:00) and rise again over the evening period (16:00-23:00), peaking then at a higher temperature. Weekend temperatures also remain higher during the middle of the day than on weekdays. However, there is variation of around 1°C over that time between the morning and afternoon heating periods, rather than a consistent temperature being maintained across the period. Despite this, the average minimum temperature achieved for 9 hours per day is relatively consistent across the SAP-assumed heating season (October-May), consistent with the model assumption. What varies month by month over the heating season is the level of variance in this value across the sample, with more variation during the empirically observed core heating season of November to March and less during the transition months of October and April. This may indicate some underheating of rooms on cold days and some overheating, particularly on warmer days, in a subset of rooms and homes.

Turning to heating durations per day, the SAP assumes a standard 9 hours of heating on weekdays and 16 hours on weekend days, with no difference between different rooms in the home. Consistent with reviewed empirical work for English homes, we find little indication of difference between weekdays and weekends, nor substantive difference between room types. The SAP also assumes durations to be standard across the heating season and invariant to external temperature. We however found large differences in heating durations across the heating season, with far more hours per day, and a wider spread of hours per day, across a core heating season, and some but much lower levels of heating in the transition periods. Correspondingly, the analyses demonstrate that as external temperature falls, the average duration of radiator use increases, and the variation in duration across the sample also grows.

In terms of diurnal patterns of active heating, i.e. periods when radiators are on or off over the day, the SAP effectively assumes that there are two distinct heating patterns during the heating season: 07:00-23:00, and 07:00-09:00 and 16:00-23:00. These closely corresponding to the *all day* and *am and pm* clusters that we found in our data. However, our analysis indicates two other common patterns: *am or pm* and *no heating* (plus a *noise* cluster).

The SAP assumes that these heating patterns vary only by day of the week across the whole heating season, so that *all day heating* in effect occurs on 2/7th of days (29%, weekends), and *am and pm* occurs on 5/7th of days (71%, weekdays). By contrast, we did not observe a substantial weekday-weekend variation like this; rather, cluster frequency varied greatly over the heating season, strongly correlating with external temperature. Furthermore, the *all day heating* cluster occurred on only 4% of room-days in our sample, and the *am and pm* cluster on only 50%. This suggests that where SAP occupancy schedules are applied to building stock models, the estimates of energy demand for heating might be substantially improved by including the additional heating clusters and modelling how these vary by month and external temperature rather than by weekday/weekend. Our results also indicated relatively high levels of transition out of certain clusters to others from one day to the next, notably out of the *all day* cluster into the *am and pm* cluster, and out of the *am or pm* cluster into the *am and pm* or *no heating* clusters. These may be explained by occupant behavioural responses to changing external temperatures, or to changes in occupancy between workdays and non-working days, with the home left unoccupied during the day. It would be of interest to explore these patterns in future work to identify explanatory factors, particularly ones which might be further introduced as refinements to the SAP model assumptions.

Finally, we found zoning of heating patterns between rooms to be relatively uncommon: homes on average zoned about 10% of days over the heating season, but only 2.5% of homes zoned their heating on a frequent basis, and on any given day only 10% of homes on average were zoning.

The current study has some limitations that could be addressed in future work. The source dataset has a similar number of homes to those used in many of the previous related works reviewed in the literature, and contains data on homes with a wide range of building properties and occupant characteristics. The room-level temperature data and the addition of inferred radiator usage measures to the subset where this was measured directly with sensors, although adding some noise to the data, makes the dataset unusually rich for exploring SAP model assumptions. It is, like the datasets used in most previous work in the literature, nevertheless not a representative sample of British households (nor of the region of Scotland from which it is sampled), so comparisons of frequencies and percentages between the data and SAP assumptions should be treated with some caution. The Edinburgh region has a lower average outdoor temperature than England, and Edinburgh in particular has a higher than UK-average proportion of 19th century tenement flats, factors which are further emphasised in the sample of homes in the dataset used in this study (older flats are over-represented, and much of the core heating season was slightly colder than the 10 year average for the study area). The high level of concurrence between our results and previous work nevertheless adds confidence regarding the generalisability of the key findings related to the diversity of heating patterns, generally lower temperature levels attained, levels of zoning and relationships to external temperature and time of year.

## 6. Conclusion

This paper has provided detailed descriptive statistics of room-level heating patterns and resultant indoor ambient temperatures, and described relationships between radiator usage, internal temperatures, room type, external temperatures and time of week and year, for a sample of homes from Edinburgh and surrounding regions of Scotland, UK, using a mix of sensor-based and inferred measures of radiator use, as well as sensor-based ambient room temperature data and weather data, for the period from August 2016 to June 2018.

The work highlights areas of concurrence and also substantial differences with assumed patterns and outcomes of heating in the UK's Standard Assessment Procedure model of building energy performance, and is broadly consistent with previous empirical findings discussed in the literature review.

We have demonstrated considerable differences in achieved temperatures between homes and rooms but, on average, temperatures during periods of active heating are lower than the SAP model assumes. Furthermore, the achieved temperature is influenced both by external temperature and by patterns of heating the home, while the SAP assumes no such differences. Those patterns of heating themselves have been demonstrated to fall into four common clusters of daily demand profile during the heating season (plus a *noise* cluster), rather than the two assumed by the SAP. Also contrary to the SAP model assumptions, the cluster that a particular room-day falls into is shaped by external temperature and, correspondingly, by day of year, but varies little by weekday vs. weekend. Zoning, defined here as the living room and at least one bedroom being in different clusters on a given day, is apparent in a minority of homes on any given day of the heating season, and few homes zone frequently over the heating season, with the average home zoning on just 10% of days.

The results are broadly consistent with other published research and suggest areas where specific changes to the use of SAP occupancy schedules and achieved room temperatures in housing stock models could be made to increase the models' concurrence with empirical findings. In particular, assumptions could be amended relating to average achieved temperatures, the range of diurnal heating patterns included, and the factors which predict them: rather than being weekday/weekend, heating patterns are more related to external temperature and/or day of the year. The consistency of the results in this study with previous work focusing on other regions of the UK provide evidence that SAP model assumptions do not need to be differentiated by geographic region.

The type of model and purpose to which it is put affect the likely value of making such changes in the underlying assumptions. BREDEM-based building stock models that utilise the standard SAP assumptions to make predictions of the heating energy use of occupied housing stock would likely produce more accurate estimates if the assumptions were updated to better match empirical observations. The value of outputs of the SAP itself in its primary use as an energy rating tool would, by contrast, be largely unaffected by such changes, as the focus is on the difference in energy rating between dwellings, or before and after interventions, independent of occupancy effects. Even for the case of building stock models, any proposed alterations to the standardised occupancy schedule that they utilise would need to investigate what the benefits and costs of various approaches to updating the models would be from technical, policy and other perspectives.

The research described here points to areas for further work. In particular, substantial variation between homes, rooms and room-days was identified in room temperatures and diurnal heating durations and patterns, beyond that explainable by external temperature and season alone. This included some substantial deviations above and below normative temperatures of 21°C in living areas, potentially indicating energy intensive heating behaviours, poor control over heating in some contexts, and risk of health impacts or fuel poverty, respectively. It also included high levels of transitions between certain heating patterns from one room-day to the next. Further statistical modelling could be undertaken to investigate predictors of the observed heating patterns and temperature outcomes, including the range of variables identified as predictors in previous literature. Finally, the impacts of the changes in occupancy patterns brought about by covid-19 and policy responses to it are likely to have substantially changed the relative frequency with which different diurnal heating patterns occur, as well as potentially leading to new heating patterns, and changes in temperature outcomes and levels of zoning. Despite the considerable expense, future work might therefore consider gathering new data of similar room-level detail, ideally from a UK-wide representative sample of homes to increase confidence in the generalisability of findings. This would enable an evaluation of the scale and nature of the changes in heating patterns and temperature outcomes, including their implications for the SAP model assumptions. Such findings could be of substantial importance for effective ongoing planning of the energy system transition.

## Acknowledgements

This paper is based on research funded by the UK Engineering and Physical Sciences Research Council and undertaken in the projects Intelligent Domestic Energy Advice Loop (grant reference EP/K002732/1) and Data-Driven

Methods for a New National Household Energy Survey (grant reference EP/M008223/1). The funding body had no involvement in the study design; in the collection, analysis and interpretation of data; in the writing of this article; nor in the decision to submit this article for publication. The authors are grateful to the research funding agency and to all those who participated directly in these projects in addition to the authors: Prof. D.K. Arvind, Cillian Brewitt, Dr. Myroslava Dzikovska, Edmund Farrow, Elaine Farrow, Jonathan Kilgour, Prof. Heather Lovell, Dr. Janek Mann, Prof. Johanna Moore, Dr. Evan Morgan, Dr. Charles Sutton, Prof. Jan Webb, Dr. Lynda Webb and Dr. Mingjun Zhong.

## References

[1] C. on Climate Change, Net Zero: The UK's contribution to stopping global warming, Tech. rep., Committee on Climate Change, London (2019).
URL https://www.theccc.org.uk/publication/net-zero-the-uks-contribution-to-stopping-global-warming/

[2] J. Henderson, J. Hart, BREDEM 2012: A technical description of the BRE Domestic Energy Model, version 1.1, Tech. rep., BRE (2015).
URL https://www.bre.co.uk/filelibrary/bredem/BREDEM-2012-specification.pdf

[3] G. Sousa, B. M. Jones, P. A. Mirzaei, D. Robinson, A review and critique of UK housing stock energy models, modelling approaches and data sources, Energy and Buildings 151 (2017) 66–80. doi:10.1016/j.enbuild.2017.06.043.

[4] D. of Energy, C. Change, The Government's Standard Assessment Procedure for Energy Rating of Dwellings - 2012 edition, sAP 2012 version 9.92 (Oct. 2013).
URL https://www.bre.co.uk/filelibrary/SAP/2012/SAP-2012_9-92.pdf

[5] M. Hughes, P. Pope, J. Palmer, P. Armitage, UK Housing Stock Models Using SAP: The Case for Heating Regime Change, Science Journal of Energy Engineering 4 (2) (2016) 12. doi:10.11648/j.sjee.20160402.11.

[6] M. Shipworth, S. K. Firth, M. I. Gentry, A. J. Wright, D. T. Shipworth, K. J. Lomas, Central heating thermostat settings and timing: Building demographics, Building Research & Information 38 (1) (2010) 50–69. doi:10.1080/09613210903263007.

[7] G. M. Huebner, M. McMichael, D. Shipworth, M. Shipworth, M. Durand-Daubin, A. Summerfield, The reality of English living rooms – A comparison of internal temperatures against common model assumptions, Energy and Buildings 66 (Nov. 2013). doi:10.1016/j.enbuild.2013.07.025.

[8] C. Hanmer, M. Shipworth, D. Shipworth, E. Carter, How household thermal routines shape UK home heating demand patterns, Energy Efficiency 12 (1) (2019) 5–17. doi:10.1007/s12053-018-9632-x.

[9] T. Kane, S. K. Firth, K. J. Lomas, How are UK homes heated? A city-wide, socio-technical survey and implications for energy modelling, Energy and Buildings 86 (2015) 817–832. doi:10.1016/j.enbuild.2014.10.011.

[10] J. Hulme, A. Beaumont, C. Summers, Energy Follow-up Survey 2011. Report 2: Mean household temperatures, Tech. Rep. BRE report number 283078, Department of Energy and Climate Change (2013).
URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/274770/2_Mean_Household_Temperatures.pdf

[11] S. Wei, R. Jones, P. de Wilde, Driving factors for occupant-controlled space heating in residential buildings, Energy and Buildings 70 (2014) 36–44. doi:10.1016/j.enbuild.2013.11.001.

[12] C. M. R. do Carmo, T. H. Christensen, Cluster analysis of residential heat load profiles and the role of technical and household characteristics, Energy and Buildings 125 (2016) 171–180. doi:10.1016/j.enbuild.2016.04.079.

[13] G. M. Huebner, M. McMichael, D. Shipworth, M. Shipworth, M. Durand-Daubin, A. J. Summerfield, The shape of warmth: Temperature profiles in living rooms, Building Research & Information 43 (2) (2015) 185–196. doi:10.1080/09613218.2014.922339.

[14] S. D. Watson, K. J. Lomas, R. A. Buswell, Decarbonising domestic heating: What is the peak GB demand?, Energy Policy 126 (2019) 533–544. doi:10.1016/j.enpol.2018.11.001.

[15] N. Goddard, J. Kilgour, M. Pullinger, N. Berliner, D. Arvind, H. Lovell, J. Moore, D. Shipworth, C. Sutton, J. Webb, N. Berliner, C. Brewitt, M. Dzikovska, E. Farrow, E. Farrow, J. Mann, E. Morgan, L. Webb, M. Zhong, [Dataset] IDEAL Household Energy Dataset, Edinburgh DataShare (2021). doi:10.7488/ds/2836.

[16] M. Pullinger, J. Kilgour, N. Goddard, N. Berliner, L. Webb, M. Dzikovska, H. Lovell, J. Mann, C. Sutton, J. Webb, M. Zhong, The IDEAL household energy dataset, electricity, gas, contextual sensor data and survey data for 255 UK homes, Scientific Data 8 (1) (2021) 146. doi:10.1038/s41597-021-00921-y.

[17] M. Office, [Dataset] UK and regional series. year-ordered statistics for region scotland e, parameter: Mean temp, Tech. rep., Met Office (2021).
URL https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-and-regional-series

[18] C. Zhang, M. Zhong, Z. Wang, N. Goddard, C. Sutton, Sequence-to-point learning with neural networks for nonintrusive load monitoring, arXiv:1612.09106 [cs, stat] (Dec. 2016). arXiv:1612.09106.

[19] G. S. Georgiou, P. Christodoulides, S. A. Kalogirou, Implementing artificial neural networks in energy building applications: A review, in: 2018 IEEE International Energy Conference (ENERGYCON), 2018, pp. 1–6. doi:10.1109/ENERGYCON.2018.8398847.

[20] N. Berliner, M. Pullinger, N. Goddard, Inferring room-level use of domestic space heating from room temperature and humidity measurements using a deep, dilated convolutional network, MethodsX 8 (2021) 101367. doi:10.1016/j.mex.2021.101367.

[21] L. McInnes, J. Healy, S. Astels, hdbscan: Hierarchical density based clustering, The Journal of Open Source Software 2 (11) (Mar. 2017). doi:10.21105/joss.00205.

[22] M. Verleysen, D. François, The curse of dimensionality in data mining and time series prediction, in: J. Cabestany, A. Prieto, F. Sandoval (Eds.), Computational Intelligence and Bioinspired Systems, Springer Berlin Heidelberg, 2005, pp. 758–770. doi:10.1007/11494669_93.

[23] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: KDD, Vol. 96, 1996, pp. 226–231.
URL https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf