# **Data cleaning, temperature and gas demand weighting, and determining the DHW demand**

# This document provides supplementary information for the paper ‘*Decarbonising domestic heating: what is the peak GB demand?****’*** Supplementary Information. Available at: https://doi.org/10.17028/rd.lboro.6795395’, by Watson SD, Lomas KJ and Buswell RA, Energy Policy, 2018.

# **S1: Data cleaning**

As noted in section 2.2 of the paper, it was necessary to have a consistently large sample of homes with no significant data errors or breaks in the data. It was also desirable that, on any given day, all the homes had a complete set of 48 half-hourly values, although the number of homes with complete data may differ day-to-day.

## **S1.1: Replacing missing data**

Between 1st August 2008 and 31st March 2009 there was little difference between the daily maximum and minimum sample size, but after 1st April 2009 a significant constant difference appeared (Figure 3). This was because approximately 1800 dwellings joined the sample at this time, and all of them had at least one missing half-hourly reading per day. For these dwellings, there were more than 20 times as many missing readings at 01:00, when the gas demand was very low[[1]](#footnote-1), than at any other time of day. As a correction, the missing value was replaced by the mean of the readings at 00:30 and 01:30. This correction was applied to 1995 dwellings in total[[2]](#footnote-2), 452 occurrences per dwelling on average. In 62% of cases the inserted value was zero.

Throughout the whole period there were numerous occasions where the sample size suddenly and briefly dropped (Figure 3). On many occasions, whole days of data were missing for some, but not all, dwellings. On other occasions, data was lost on one or more of the 48 half-hours in the day. On three occasions, affecting nine days, all the dwellings suffered simultaneous half-hourly data loss. On all these occasions, the data from the preceding or following day, whichever had the closest outdoor effective air temperature (see below), was used to replace all the faulty days’ data.

These measures produced a sample of between 5133 and 7664 dwellings all of which had complete days of data for the period from 1st August 2008 to 31st July 2010.

## **S1.2: Selecting the period of reliable data**

The reliability of the data in the period 1st August 2008 to 31st July 2010 was investigated using three metrics: the daily Mean Load Factor (MLF); the daily Peak Coincidence Factor (PCF); and the correlation (R2) between the daily gas consumption and the EDRP-weighted effective outdoor air temperature (see S2.1).

The load factor for a particular day and dwelling is the ratio between the mean half-hourly gas demand on that day and the peak half-hourly demand (Equation S1.1). The MLF for the particular day is the average of these values across all the dwellings in the sample.

Whereas the daily MLF indicates the ratio between the mean and the peak demand for each individual dwelling, the daily PCF examines the extent to which the peaks in the demand of different dwellings coincide. It is calculated for a given day by summing the half-hourly gas demands at each half hour for all the dwellings and finding the maximum of these 48 half-hourly values. This value is divided by the sum of the maximum recorded half-hourly demands of each dwelling on the same day (Equation S1.2).

Equation S1.1

Where:

*LFh,d* = Load Factor of dwelling h on day d

*EGh,t,d* = Gas demand of dwelling h, at time t, on day d

*t* = time of day in half hours, *d* = day and *h* = dwelling.

Equation S1.2

Where:

*PCFd* = Peak Coincidence Factor of sample on day d

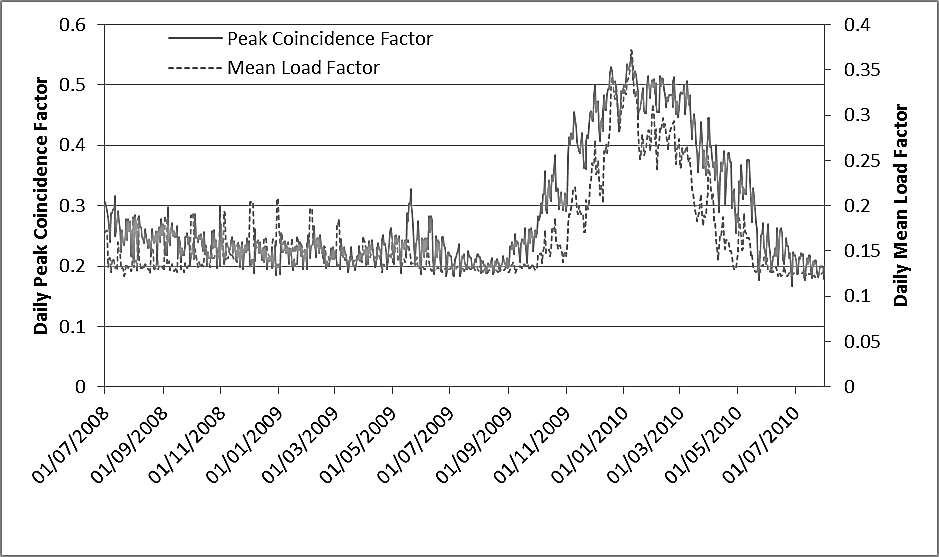
*EGh,t,d* = Gas demand of dwelling h, at time t, on day d

*t* = time of day in half hrs., *d* = day, *h* = dwelling, *n*= no. dwellings in sample.

In homes heated by gas, one would expect the both the MLF and the PCF on winter days to be much higher than in summer. In winter, gas is used for both space heating and DHW giving extended periods of gas use which are likely to occur at similar times in different dwellings, whereas in summer, gas use is mainly for water heating, which imposes short but energy intensive gas demands which may occur at different times of the day in different dwellings. Thus, the daily MLF will be higher in winter because the extended periods of gas use raise mean demand without necessarily increasing peak demand, and the daily PCF will be greater because the periods of gas use in different dwellings tending to overlap and coincide in the winter.

During the summers of 2009 and 2010, both the daily MLF and the daily PCF are low but both increase considerably through the winter of 2009-10 (Figure S1.1), which is the expected behaviour. During the winter of 2008-09 there is little change in either the daily MLF or the daily PCF, suggesting that the data is not reliable.

To investigate further, the period was split into two, before 1st May 2009 and from 1st May 2009 to 31st July 2010. For each period the correlation between the daily gas consumption of each dwelling and the outdoor effective air temperature was plotted. The gradient of the plot should, of course, be negative because the energy used for space heating will increase as the weather gets colder. The coefficient of determination (R2) of the relationship between daily gas demand and the outdoor effective air temperature was calculated for each dwelling for both periods.



**Figure S1.1: Mean Load Factor and Peak Coincidence Factor, 1st August 2008 to 31st July 2010**

The R2 values before 1st May 2009 were very variable with a mean of just 0.39 and only 20% of R2 values greater than 0.6. In contrast, after 1st May 2009, the mean R2 value for the whole sample was 0.74 with 86% of dwellings produced values greater than 0.6. Inspection of the raw data for the first period also revealed unexpected peaks in the total gas consumption at 23:30 for all the dwellings, which was clearly indicative of a problem with the data collection process.

Taken together, the MLF, PCF and R2 analyses suggest that data with the expected temporal behaviour was produced for the 15-month period from 1st May 2009 to 31st July 2010. Throughout this period, the sample size was between 5,707 and 7,664 dwellings, with a mean of 7,202.

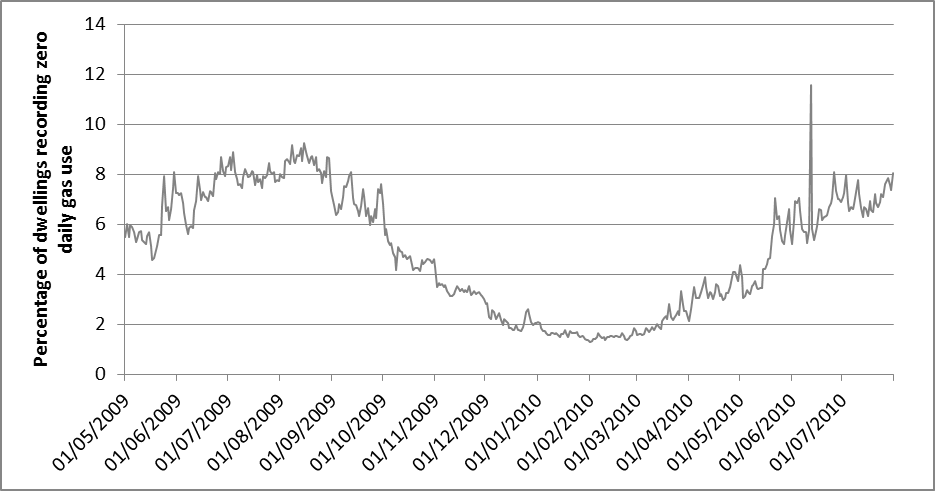
## **S1.3: Identifying and removing implausible data**

During the period from 1st May 2009 to 31st July 2010, some gas readings exceeded 70 kW, which is implausibly high for the vast majority of UK domestic boilers. When such readings occurred, the entire day of data was removed from the affected dwelling; 251 dwellings were affected and, on average, 2.6 days of data were removed.

There were 752 dwellings in the sample that yielded less than 20 days of valid data. Including such dwellings would have complicated the analysis unnecessarily so they were removed. (Although the number of dwellings removed is large the amount of data they represent is small).

On any given day between 1% and 9% of the remaining dwellings recorded zero gas use[[3]](#footnote-3). There are several possible reasons for this: the dwelling was unoccupied; occupants were genuinely using no heat[[4]](#footnote-4); other fuels were being used for space or DHW heating, most obviously electricity; or there were errors in the data. Since far fewer dwellings had zero daily gas use in winter (Figure S1.2), it is likely that many dwellings used gas for space heating but not DHW. As a pragmatic solution, dwellings with more than 15% of the days (i.e. 64 days in the 15-month period) with a daily gas consumption of zero were removed from the sample (883 dwellings). Dwellings with fewer than 64 days of zero gas use were retained in the sample and the zero readings ascribed to occupant absences.

Finally, data from around 388 dwellings produced a very poor correlation between the effective outdoor air temperature and daily gas use, R2<0.4 (Figure 5). Manual inspection of these dwellings’ data showed that they contained much spurious data, so these dwellings were removed from the sample.



**Figure S1.2: Percentage of dwellings recording zero gas use on each day, 1st May 2009 to 31st July 2010**

# **S2: Energy demand and effective temperature weighting**

## **S2.1: Weighting effective outdoor air temperature**

To model the variation of gas demand with weather, the National Grid (National Grid, 2012) use a Composite Weather Variable (CWV). The main component of this CWV is the mean daily effective outdoor air temperature (*ET*), the others being wind chill and various seasonal effects. *ET* is an exponentially weighted moving average of outdoor air temperature. The moving average takes account of the effect of the thermal mass of a building, which slows the change in heat demand in response to changes in outdoor temperature. GB is divided into 13 Local Distribution Zones (LDZs) by National Grid. The mean daily outdoor air temperature in each LDZ (*Tz,d*) was obtained from the weather stations used by National Grid (UK Meteorological Office, 2017) and the ET was calculated for each of LDZ (Equation S2.1).

|  |  |
| --- | --- |
|  | Equation S2.1 |

Where:

*ET* = Effective temperature

*T* = mean daily outdoor air temperature

*α* = coefficient (0.5 in National Grid’s method)

z = Local Distribution Zone number.

*d* = day

In order to produce a national CWV, National Grid multiply the CWV in each LDZ by a weighting which reflects the proportion of GB national (NDM) gas demand in that LDZ (Table S2.1, *PNz*). The ET from each LDZ was multiplied by these weightings to give a National-Grid-weighted ET for GB, *ETNG* (Equation S2.2). This temperature is representative of the temperatures experienced by the GB housing stock.

|  |  |
| --- | --- |
|  | Equation S2.2 |

Where:

*ETNG* = National Grid-weighted Effective Temperature

*PNz* = National Grid LDZ weighting (Table S2.1)

*z* = Local Distribution Zone number

*d* = day

However, the proportions of homes in each LDZ in the EDRP sample (Table S2.1, *PEz*) was very different to the National Grid weightings. Therefore, an EDRP-weighted ET (*ETEDRP*) was calculated by multiplying the ET from each LDZ by the proportion of the EDRP sample in that LDZ (Equation S2.3). This temperature is representative of the temperatures experienced by the EDRP sample.

|  |  |
| --- | --- |
|  | Equation S2.3 |

Where:

*ETEDRP* = EDRP-weighted mean daily effective outdoor air temperature

*PEz* = percentage of EDRP sample in each LDZ (Table S1.1)

*z* = Local Distribution Zone number

*d* = day

When developing the regression model of daily heat demand from EDRP data, the EDRP-weighted temperature (*ETEDRP*) was used, since this reflects the mean outdoor temperatures experienced by the EDRP dwellings. When the model was used to give predictions of GB daily heat demand, the National Grid-weighted temperature (*ETNG*) was used as an input, since this represents the temperatures experienced by the GB housing stock.

**Table S2.1: The percentage of EDRP dwellings in each LDZ and the National grid temperature weightings GB dwellings and the subsample**

|  |  |  |  |
| --- | --- | --- | --- |
| LDZ | | National grid  LDZ weightings\*  PNz  (%) | EDRP  sub-sample  PEz  (%) |
| EA | East Anglia | 8 | 0 |
| EM | East Midlands | 10 | 38 |
| NE | North East | 6 | 0 |
| NO | North | 6 | 0 |
| NT | North Thames | 11 | 0 |
| NW | North West | 12 | 0 |
| SC | Scotland | 9 | 18 |
| SE | South East | 10 | 8 |
| SO | South | 7 | 9 |
| SW | South West | 6 | 1 |
| WM | West Midlands | 9 | 20 |
| WN | Wales North | 1 | 0 |
| WS | Wales South | 4 | 6 |
| \*The National grid temperature weightings reflect the  relative proportions of NDM gas demand in each LDZ.  For LDZ locations see map (Figure 4) | | | |

## **S2.2: Weighting gas demand**

During the winter from 1st December 2009 to 28th February 2010, gas demand profiles for each Acorn category indicated that the gas demands decrease as the Acorn category number increases (Figure S2.1, top), with homes in Acorn Category 1 using 60% more gas on average than homes in Category 5 (Table S2.2, Column 8). This is not surprising because the Acorn category may reflect the size of dwelling and the income of the household, both of which are known to affect energy demand (Morris et al. 2015) and, perhaps, the pattern of heating.

The gas demand profiles for all Acorn categories were however similar, displaying twin peaks, one in the morning at around 08:00 and the other in the evening at around 17:30 (Figure S2.1, top). This is consistent with the findings of internal temperature monitoring studies, which have shown that two-period heating on both weekdays and at weekends is the most common pattern (Kane et al, 2015; Huebner et al, 2015).

The normalised profiles (Figure S2.1, bottom) indicate that homes in Acorn categories 4 and 5 have relatively lower morning peak demands than the homes in the other three categories.



**Figure S2.1: Half-hourly gas demand of the EDRP sub-sample averaged over the period from 1st December 2009 to 28th February 2010: Top, average half-hourly gas demand by Acorn Category; Bottom, average half-hourly demand normalised such that the daily total of half-hourly gas demand is one.**

Because homes in different Acorn categories use different amount of gas, any differences in the relative proportion of homes in each Category must be accounted for when scaling up from the EDRP subsample to the GB housing stock as a whole. In fact, homes in Acorn categories one (Wealthy Achievers) and three (Comfortably Off) are over-represented in the EDRP subsample compared to GB as a whole (Cumbria Intelligence Agency, no date), see Table S2.2 columns 4 and 5.

**Table S2.2: Number of dwellings in each Acorn category in the EDRP sub sample and in GB, and the mean and total gas demand from dwellings in each Acorn category**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| Acorn Category | Number of dwellings in cleaned subsample  (1st May 09 to 31st July 10). | | Percentage of homes in each Acorn category | | EDRP winter gas use characteristics  1/12/09-28/02/10 | | | | |
| EDRP  2009-10 | GB households  2010\* | Mean Peak gas  demand per dwelling | | Mean daily gas  demand per dwelling | Mean Load  Factor | |
| Max. No | Min No. | % | % | kW | Time | kWh | - | |
| 1 Wealthy Achievers | 2361 | 1918 | 36 | 25 | 8.76 | 08:00 | 121 | 0.58 | |
| 2 Urban Prosperity | 286 | 194 | 4 | 12 | 7.47 | 17:30 | 106 | 0.59 | |
| 3. Comfortably Off | 2197 | 1774 | 33 | 28 | 6.88 | 17:30 | 93 | 0.56 | |
| 4. Moderate Means | 820 | 584 | 12 | 14 | 6.17 | 17:30 | 86 | 0.58 | |
| 5 Hard-Pressed | 973 | 719 | 15 | 21 | 5.59 | 17:30 | 75 | 0.56 | |
| Total | 6637 | 5189 | 100 | 100 | 7.02 | 17:30 | 97 | 0.58 | |
| \* A total of 20.4million GB homes were heated by gas in 2010. | | | | | | | | |

To create the regression model, the GB Acorn-weighted half-hourly gas demand per dwelling had to be calculated. To do this, the mean half-hourly gas demand per dwelling was obtained for each Acorn category, and this was multiplied by the proportion of GB households in that Acorn category and summed. This was multiplied by mean boiler efficiency to give heat demand (see Equation S2.4).

|  |  |
| --- | --- |
|  | Equation S2.4 |

Where:

*EGnat, t* = National ACORN-weighted gas demand per dwelling (kWh) on day d at time t

*EGh,c,t* = Gas demand of dwelling h, belonging to Acorn category c, on day d at time t (kWh)

*Nc,t*= Number of dwellings in Acorn category c with valid readings on day d at time t

*PGBc* = Percentage of GB households in Acorn category c (Table S2.2, column 5)

*h* = Dwelling number

*c* = Acorn category number

*t* = time (half hours)

*d* = day

# **S3: Separation of DHW and space heating demand**

Gas is used in homes for space heating, DHW and cooking. Cooking is not of interest here, but the gas used for cooking had to be removed from the EDRP data in order to separate out the other two uses of gas. Around 2% of annual domestic gas demand is for cooking (BEIS, 2017), and this was assumed to be distributed evenly through the year. This was subtracted from the EDRP data, leaving only gas used for space heating and DHW.

The DHW gas demand at any particular half hour day was calculated in a two part process in a similar way to that for the total gas demand. Firstly, the mean daily demand was calculated and then this was distributed through the day using a half-hourly gas demand profile. The half-hourly DHW gas demands were then subtracted from the total half hourly gas demand to give the half-hourly space heating demand.

## **S3.1 Determining the mean daily DHW demand**

The mean daily gas demand, after subtracting the demand for cooking, was separated into space heating and DHW heat demand in a three-stage process. Firstly, the cut-off temperature was determined above which negligible amounts of gas are used for space heating (thus all gas use is for DHW), secondly, the mean daily gas demand and mean outdoor air temperature on days that were warmer than this cut-off temperature were obtained, then thirdly, the daily gas use was scaled according to outdoor temperature to give DHW gas demand at all outdoor temperatures.

It is commonly assumed that GB dwellings do not require space heating at outdoor temperatures above 15.5°C (Carbon Trust, 2017), meaning that all gas use above this temperature must be for DHW and cooking. However, the relationship between heat demand and outdoor temperature does not exhibit a sudden change of gradient at any given outdoor temperature (see Figure 5). Instead, there is a gradual change in gradient, as occupants choose to switch their space heating on at a wide range of outdoor temperatures (Kane et al, 2015). Therefore, it is not clear at what outdoor air temperature, called here the cut-off temperature, no space heating required and all gas use is for DHW.

To determine a suitable cut-off temperature, a straight line was fitted to the relationship between the total daily gas demand of all the EDRP homes and the effective outdoor air temperature, for all days above a chosen cut-off temperature. This cut-off temperature was varied from 14°C to 20°C in steps of 0.1°C. The correct cut-off temperature was determined based on two criteria: the outdoor temperature at which theoretically no DHW heat is required, and the change in gradient of the straight line when the cut-off temperature is varied. The second criterion stems from the observation that, if the cut-off temperature is correctly chosen, then no space heating is included in the gas demand, so increasing the cut-off temperature by a small amount should have no effect on gradient. If, however, significant space heating were still included, increasing the cut off temperature would result in a noticeable change in gradient.

Concerning the x-axis intercept, the expected value should be close to that obtained from the equations in SAP2012 (BRE,2013), which give the relationship between the mean monthly outdoor air temperature (UK average) and DHW heat demand. This relationship is based on the DHW heat demand and outdoor temperature monitored for a sample of homes by Shorrock (2008). Using values derived from SAP2012, the outdoor temperature at which theoretically no DHW heat is required was obtained from a straight-line regression (R2=0.96) as 40°C (the x-axis intercept).

This value was compared to the x-axis intercept obtained from the EDRP data using different cut-off temperatures. Having the same x-axis intercept indicates that two lines are multiples of each other. Thus, if the x-axis intercept for the SAP2012 and EDRP lines are the same, the variation in DHW demand with temperature is the same for the two lines, although the quantities measured may be different (i.e. the values from SAP2012 may not be representative of GB homes, but they should show the correct variation with temperature).

It was found that when a cut-off temperature of 18°C was used, the x-axis intercept for the EDRP homes, was 44°C, which is close to the value from SAP2012. The gradient of the line, showed little change when the cut off was increased by small amounts above this value. The R2 value of was 0.42 and the sample size was 63 days. Kane et al (2015) also concluded, using data form different homes, that no space heating was required at outdoor temperatures above 18°C. The mean daily effective outdoor air temperature on all the days with a temperature above 18°C was 19.1°C, and the mean daily gas demand 9.8 kWh.

To scale the mean daily gas demand at 18oC in order to obtain the DHW demands at higher temperatures, it was assumed that the DHW demand varied with temperature according the equation in SAP2012. A scaling factor was obtained using the SAP2012 DHW equation by dividing the DHW demand at various temperatures by the DHW demand at 19.1°C. This factor was then used to scale the gas demand of the EDRP homes at 18oC, to obtain the DHW gas demand at for all days with outdoor temperature above 18°C. The relationship between scaling factor and outdoor temperature is shown in Figure S3.1.

This calculated DHW demand was then subtracted from total EDRP gas demand to give the space heating gas demand any particular temperature above 18oC.

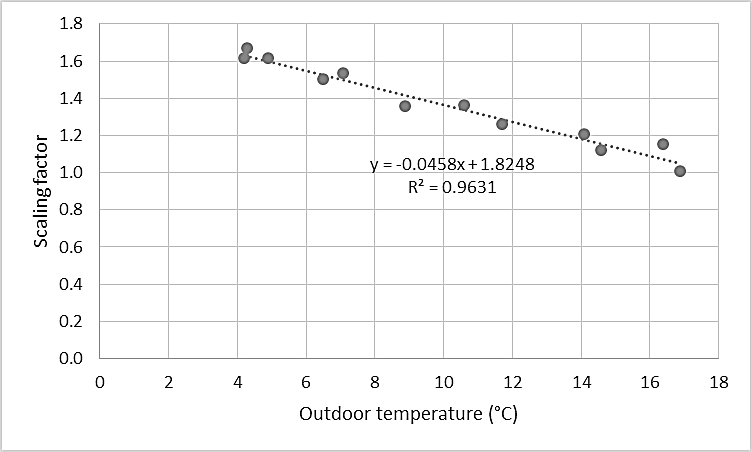


Figure S3.1: The scaling factor to calculate the DHW demand from the outdoor air temperature, derived from SAP2012 (BRE, 2013)

## **S3.2: Determining the half-hourly DHW gas demand profile**

To convert the mean daily DHW demand to half-hourly values a normalised half-hourly DHW demand profile was produced. This was simply taken to be the normalised profile of the mean half-hourly gas demands calculated using the data for all the days with an effective outdoor air temperature above 18°C. This profile is rather flat compared to the total gas demand profiles (e.g. Figure S3.2).

The absolute half -hourly gas demand for DHW at the prevailing outdoor temperature was determined by scaling (see above) the normalised half-hourly gas demand. For consistency with the approach for calculating the total half-hourly gas demands, the scaling was done to produce normalised DHW profiles for temperatures of 3°C, 0°C … 15°C.

The half-hourly gas demands for space heating were then obtained by simply subtracting the half-hourly DHW demand from the total gas demand at that temperature. As an example, the total gas demand profile for one house at 3°C is shown in Figure S3.2, top, with the DHW gas demand profile (middle) and the resulting net space heating gas demand (bottom).

In general, as shown here, calculated space heating gas demand profiles were similar in shape to the total heat profile, however, the load factor of the space heating profile tended to be slightly higher than for total heat at low temperatures, and slightly lower than for total heat at higher temperatures.

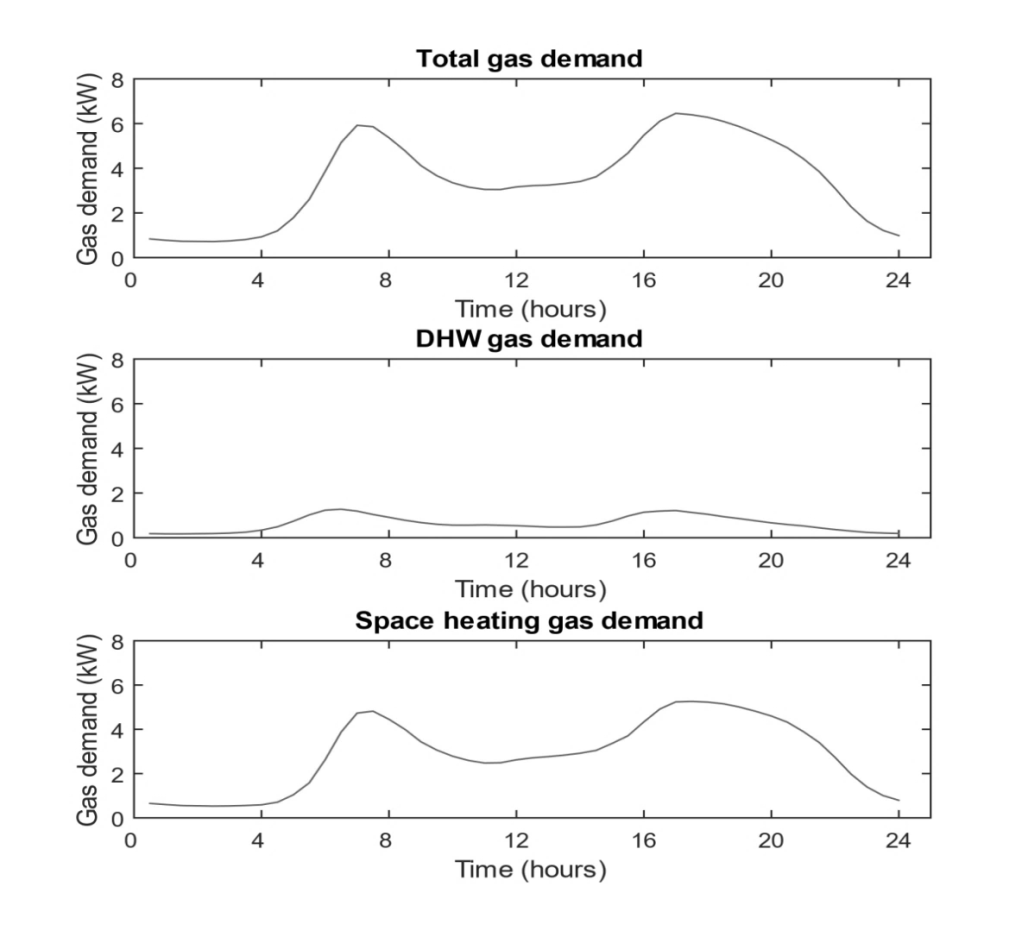


Figure S3.2: Profile shapes for total gas demand (top), DHW gas demand (middle) and space heating gas demand (bottom) at effective temperature 3°C

# **References**

BEIS, 2017. Energy Consumption in the UK. Department for Business, Energy and Industrial Strategy. London. Available at: https://www.gov.uk/government/statistics/energy-consumption-in-the-uk [Accessed 6th November 2017]

BRE, 2013. SAP 2012 – The Government’s Standard Assessment Procedure for Energy Rating of Dwellings. Building Research Establishment, Watford, on behalf of Department of Energy and Climate Change.

Carbon Trust, 2017. Degree Days. Available at: https://www.carbontrust.com/resources/guides/energy-efficiency/degree-days/ [Accessed 14th December 2017]

Cumbria Intelligence Agency, No Date. Profiling financially excluded groups in Cumbria. Available at: www.cumbria.gov.uk/elibrary/view.asp?ID=38608 [Accessed 8th February 2018]

Huebner, G.M. et al., 2015. The shape of warmth: temperature profiles in living rooms. *Building Research & Information*, 70, pp.298–305. Available at: http://dx.doi.org/10.1080/09613218.2014.922339.

Kane, T., Firth, S.K. & Lomas, K.J., 2015. How are UK homes heated? A city-wide, socio-technical survey and implications for energy modelling. *Energy and Buildings*, 86, pp.817–832. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0378778814008317.

Morris, J. et al., 2015. Benchmarking and tracking domestic gas and electricity consumption at the local authority level. *Energy Efficiency.* DOI 10.1007/s12053-015-9393-8

National Grid, 2012. Gas Demand Forecasting Methodology. Available at: https://www.nationalgrid.com/sites/default/files/documents/8589937808-Gas%20Demand%20Forecasting%20Methodology.pdf

Shorrock, L., 2008. Analysis of the EST's domestic hot water trials and their implications for amendments to BREDEM and SAP. Building Research Establishment, Watford.

UK Meteorological Office, 2017. Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853 -current). British Atmospheric Data Centre. Available at: http://artefacts.ceda.ac.uk/badc\_datadocs/ukmo-midas/.

1. On average 0.5% of daily consumption occurred during the half hour ending at 01:00 [↑](#footnote-ref-1)
2. 1995 rather than 1800 because some of these dwellings dropped out of the sample and others joined, the cohort thus changed on a day-to-day basis. [↑](#footnote-ref-2)
3. There is a spike of 11.5% on 12th June 2016. The reason for this is not known. [↑](#footnote-ref-3)
4. There were no significant correlations between Acorn category and the number of zero readings suggesting that readings of zero daily gas use are unlikely to be caused by occupants choosing not to use heat because of financial constraints. [↑](#footnote-ref-4)