

Smart meter consumption data: Data quality report

Creation date	2021-08-26
Edition	03
Author	Ellen Webborn, Jessica Few
Project	Smart Energy Research Lab (SERL)
Organisation	University College London (UCL)

Table of Contents

Data availability	2
Number of participants with data over time.....	4
Errors	8
Half-hourly data	8
Daily Data	10
Distributions of valid reads	10
Half-hourly reads.....	10
Daily reads	13

Data availability

The following tables summarise the amount of valid data among participants recruited in waves 1, 2 and 3. For more information about error definitions, see the Errors section below or the *serl_smart_meter_documentation_edition03.pdf* accompanying document. The data are presented as percentages on the next page. The abbreviation 'HH' stands for 'half-hourly'.

The first row of Table 1 shows the number of participants for which the read type above was possible. Around 6500 participants have at least 90% of all import readings valid and over 280 have at least 90% of all export readings valid. Some participants have "read type possible" for half-hourly data but not for daily. This is either because these participants withdrew consent after we started collecting half-hourly data but before daily reading collection began, or because they have a SMETS1 meter that doesn't collect daily reads.

Note that where daily readings are invalid or unavailable it is often possible to impute the reading with the sum of the half-hourly readings for that day. Where possible a valid half-hourly sum is provided for comparison/imputation with the daily reads.

Table 1: Number of participants with different minimum levels of valid import ('im') and export ('ex') readings.

	<i>Daily gas (im)</i>	<i>HH gas (im)</i>	<i>Daily elec (active im)</i>	<i>HH elec (active im)</i>	<i>HH elec (reactive im)</i>	<i>HH elec (active ex)</i>	<i>HH elec (reactive ex)</i>
Read type possible	9,404	10,202	12,328	13,320	13,320	806	806
100% valid	3,662	1,249	5,704	5,074	236	10	10
>= 99% valid	5,031	4,646	6,155	8,090	549	16	16
>= 98% valid	5,496	5,432	6,264	8,676	1,432	22	22
>= 95% valid	6,193	6,427	6,588	9,484	2,722	143	143
>= 90% valid	6,647	7,651	6,895	10,921	7,712	287	292

	<i>Daily gas (im)</i>	<i>HH gas (im)</i>	<i>Daily elec (active im)</i>	<i>HH elec (active im)</i>	<i>HH elec (reactive im)</i>	<i>HH elec (active ex)</i>	<i>HH elec (reactive ex)</i>
>= 75% valid	7,130	8,788	7,342	11,911	9,894	678	678
>= 50% valid	7,540	9,142	8,037	12,239	12,328	791	791
>= 25% valid	8,409	9,345	9,232	12,336	12,677	804	804
>= 10% valid	8,504	9,436	9,392	12,408	12,771	806	806
>= 5% valid	8,536	9,472	9,438	12,432	12,788	806	806
>= 0% valid	9,404	10,202	12,328	13,320	13,320	806	806

Table 2 shows the data from Table 1 presented as percentages. The first row of data ('Read type possible') shows the percent of all participants for which we expect to have each type of read. Approximately 77% of households have a DCC-enrolled gas meter (i.e. a SMETS2 gas meter or an upgraded SMETS1 meter). The percentages in the rows that follow are percentage of valid readings from those meters for which we expect the read type, e.g. 82.0% of households with a smart meter that we could collect data from have at least 90% of their half-hourly electricity active import readings available and valid. 6.1% of participants have at least 1 export read.

Table 2: Percentage of participants with different minimum levels of valid import ('im') and export ('ex') readings. 'Read type possible' indicates the percentage of participants with that type of read available.

	<i>Daily gas (im) (%)</i>	<i>HH gas (im) (%)</i>	<i>Daily elec (active im) (%)</i>	<i>HH elec (active im) (%)</i>	<i>HH elec (reactive im) (%)</i>	<i>HH elec (active ex) (%)</i>	<i>HH elec (reactive ex) (%)</i>
Read type possible	70.6	76.6	92.5	100.0	100.0	6.1	6.1
100% valid	38.9	12.2	46.3	38.1	1.8	1.2	1.2
>= 99% valid	53.5	45.5	49.9	60.7	4.1	2.0	2.0
>= 98% valid	58.4	53.2	50.8	65.1	10.8	2.7	2.7
>= 95% valid	65.9	63.0	53.4	71.2	20.4	17.7	17.7

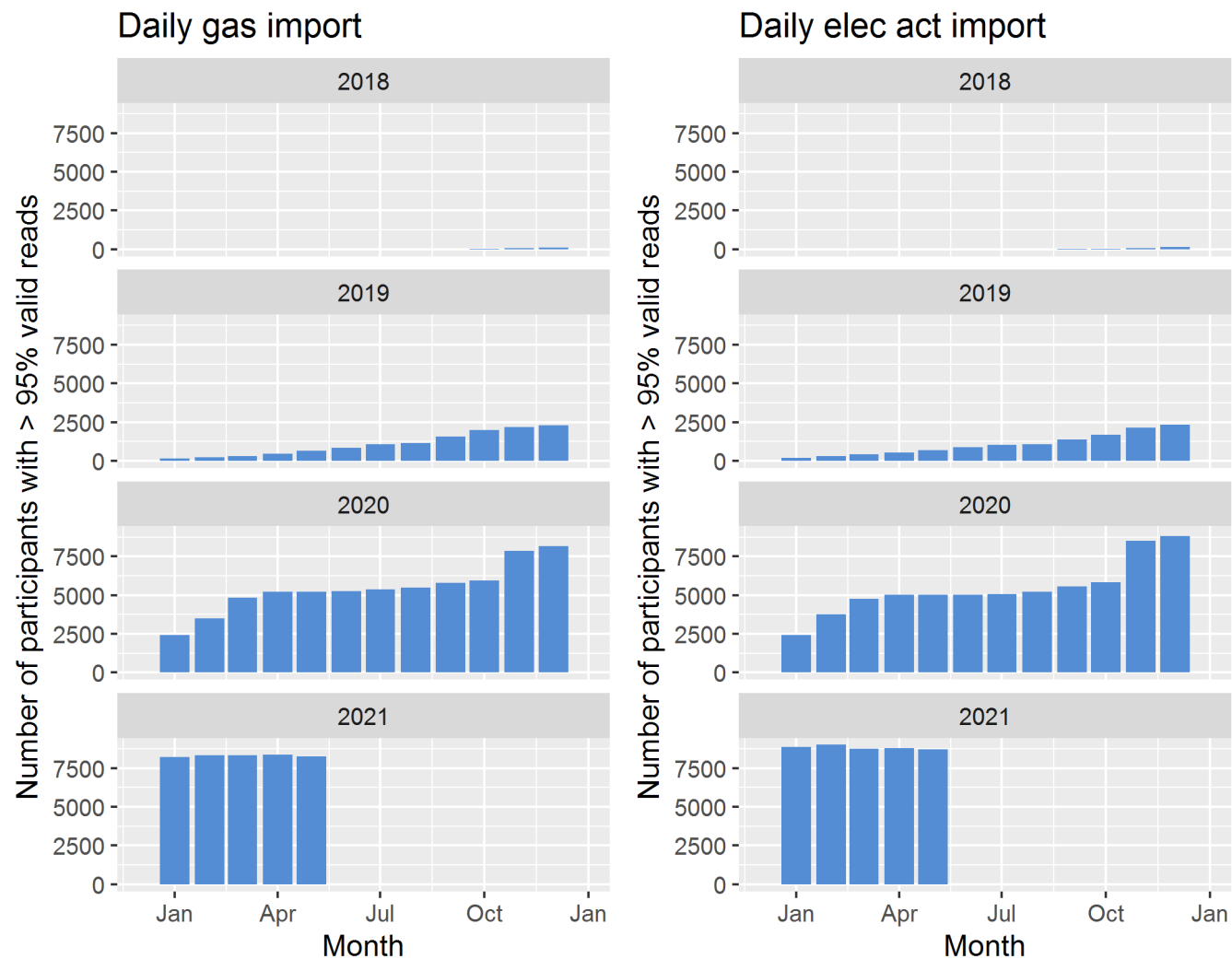
	<i>Daily gas (im) (%)</i>	<i>HH gas (im) (%)</i>	<i>Daily elec (active im) (%)</i>	<i>HH elec (active im) (%)</i>	<i>HH elec (reactive im) (%)</i>	<i>HH elec (active ex) (%)</i>	<i>HH elec (reactive ex) (%)</i>
>= 90% valid	70.7	75.0	55.9	82.0	57.9	35.6	36.2
>= 75% valid	75.8	86.1	59.6	89.4	74.3	84.1	84.1
>= 50% valid	80.2	89.6	65.2	91.9	92.6	98.1	98.1
>= 25% valid	89.4	91.6	74.9	92.6	95.2	99.8	99.8
>= 10% valid	90.4	92.5	76.2	93.2	95.9	100.0	100.0
>= 5% valid	90.8	92.8	76.6	93.3	96.0	100.0	100.0
>= 0% valid	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Number of participants with data over time

The following figures show how many participants have at least 95% of each month's data available and flagged as valid. This means no more than 1 daily read missing/invalid and no fewer than 75 half-hourly reads in a month. A participant's earliest read date depends on the date they signed up (we can collect up to 1 year of historic data from consent date (3 months for export data)), when their SMETS2 meter was installed (or upgraded if SMETS1), and when they moved into the house (if they moved in within the year preceding sign up).

We see a steady increase in the numbers from August 2018 (1 year before the first sign up) until January 2021, reflecting the availability of historic data for participants recruited in different waves. The flattening in 2021 reflects the fact that all historic data for participants has been collected and collection is now in ongoing/longitudinal mode.

Figure 1: Number of participants in the SERL Observatory with at least 95% of daily data available and valid, by month.



Half-hourly data collection dates back to the start of 2019 and increases in a similar way to the daily data in the previous figure. The half-hourly reactive data is only stored on the meter for 3 months which is why the trends are different, and increase sharply around 3 months before each recruitment wave.

Figure 2: Number of participants in the SERL Observatory with at least 95% of half-hourly import data available and valid, by month.



Export data is only stored for 3 months on the meter, and collection for the first wave did not begin as early as for the other datasets, hence the delay in the increases in numbers of participants in these figures.

Figure 3: Number of participants in the SERL Observatory with at least 95% of half-hourly export data available and valid, by month.



Errors

More details about the error flags/codes can be found in the *serl_smart_meter_documentation_edition03.pdf* accompanying document.

Half-hourly data

The half-hourly data runs from 2019-01-13 to 2021-06-01 and 13,070 participants have half-hourly reads. For each half-hourly read, for each participant, the following tables show the number of errors and valid reads in the half-hourly electricity and gas datasets. Abbreviations: 'act' (active), 'react' (reactive) and 'elec' (electricity).

Note that 13,320 participants have an electricity meter and 10,202 participants have a gas meter (76.6%). The main issue for half-hourly electricity active import and export reads is missing data, which affects import reads much more than export reads (as a percentage of available reads for each).

Table 3: Half-hourly electricity active import and export readings: number and percent of readings with each error flag.

<i>Flag</i>	<i>Meaning</i>	<i>Act elec import (N)</i>		<i>Act elec import (%)</i>	<i>Act elec export (N)</i>		<i>Act elec export (%)</i>
1	Valid	280,433,417	88.5		10,236,704	87.0	
0	Missing	36,290,271	11.5		1,526,268	13.0	
-1	Max read	542	0.0		0	0.0	
-2	Very high but not max	182	0.0		9,778	0.1	
-3	Negative	0	0.0		0	0.0	
-4	Elec in kWh	0	0.0		0	0.0	
-5	Valid read but invalid read time	228	0.0		18	0.0	

The electricity reactive import and export reads only have issues with missing data and invalid read times (but a maximum valid read has not yet been established for the reactive read data (this may change in future)). The percentage of active half-hourly reads missing is similar for reactive for both import and export reads.

Table 4: Half-hourly electricity reactive import and export readings: number and percent of readings with each error flag.

<i>Flag</i>	<i>Meaning</i>	<i>React elec import (N)</i>		<i>React elec import (%)</i>	<i>React elec export (N)</i>		<i>React elec export (%)</i>
1	Valid	158,566,764	84.1		10,246,482	87.0	
0	Missing	29,901,735	15.9		1,526,268	13.0	
-1	Max read	2	0.0		0	0.0	
-2	Very high but not max	1	0.0		0	0.0	
-3	Negative	0	0.0		0	0.0	
-5	Valid read but invalid read time	186	0.0		18	0.0	

The half-hourly gas reads have greater availability than the half-hourly electricity active import reads, but suffer from a few incredibly high readings. The 'max read' error flag (-1) is when the read is the largest number storable/transmittable (all 1s in binary). For more details see the aforementioned smart meter data documentation.

Table 5: Half-hourly gas import readings: number and percent of readings with each error flag.

<i>Flag</i>	<i>Meaning</i>	<i>Gas import (N)</i>		<i>Gas import (%)</i>
1	Valid	208,866,909	86.6	
0	Missing	30,480,701	12.6	
-1	Max read	1,736,169	0.7	
-2	Very high but not max	7,622	0.0	
-3	Negative	0	0.0	
-5	Valid read but invalid read time	69,431	0.0	

Daily Data

The daily data runs from 2018-08-18 to 2021-05-31. A total of 13,321 participants are included in these datasets. Of these, 13,012 have daily reads. For each daily read, for each participant, the following tables show the number of errors and valid reads in the daily electricity and gas datasets. Note that in this table the gas and electricity reads are included together, as the only daily read types are gas import and electricity active import. There is greater availability of the gas data than the electricity data. Around 9% of the electricity data was suspected to be recorded in kWh not in Wh as prescribed by the Smart Energy Code.

Table 6: Daily electricity active import readings and daily gas import readings: number and percent of readings with each error flag.

<i>Flag</i>	<i>Meaning</i>	<i>Elec act imp (N)</i>		<i>Elec act imp (%)</i>	<i>Gas import (N)</i>		<i>Gas import (%)</i>
1	Valid	4,009,715	66.0		3,803,714	82.7	
0	Missing	1,467,930	24.2		749,785	16.3	
-1	Max read	0	0.0		37,880	0.8	
-2	Very high but not max	1,196	0.0		4,394	0.1	
-3	Negative	0	0.0		0	0.0	
-4	Elec in kWh	594,690	9.8		0	0.0	
-5	Valid read but invalid read time	6	0.0		1,260	0.0	

Distributions of valid reads

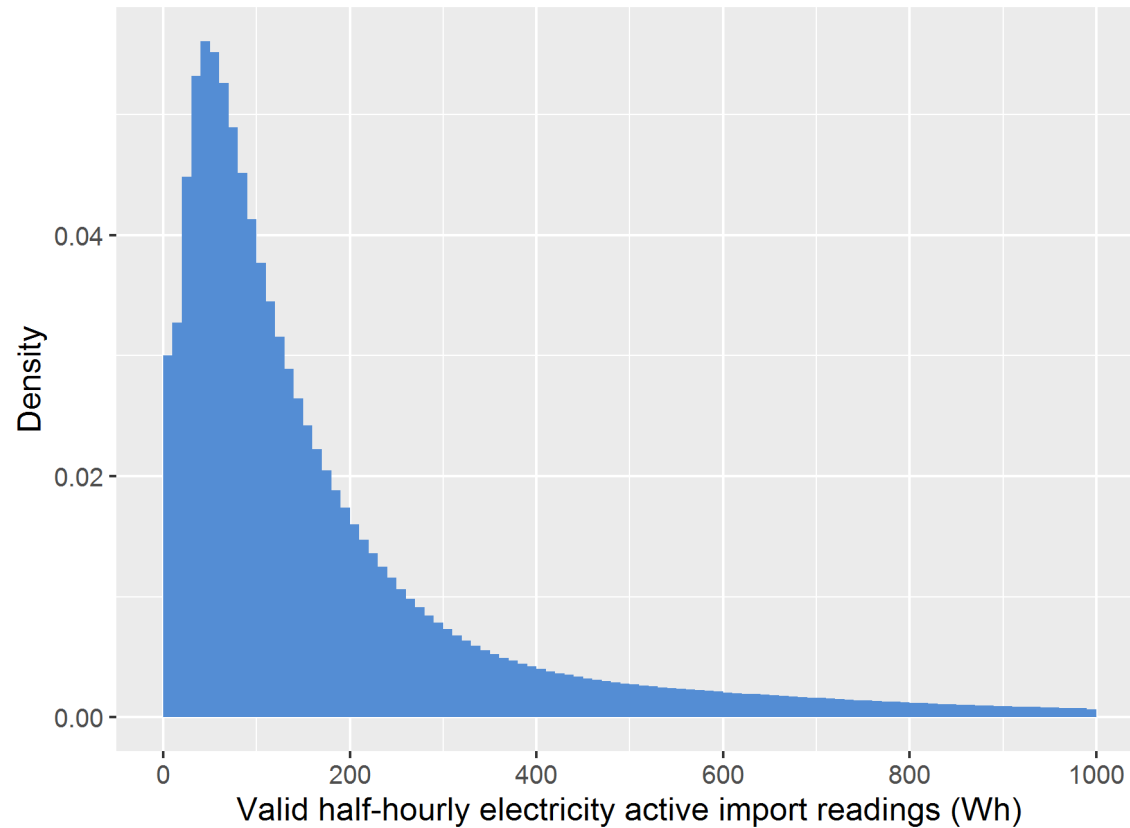
Half-hourly reads

The following figures show the half-hourly electricity active and gas import readings deemed to be valid. Reactive reads will be summarised in future documents.

Electricity active import

The electricity active import reads form a distribution with a shape that is typical in the literature for this type of data.

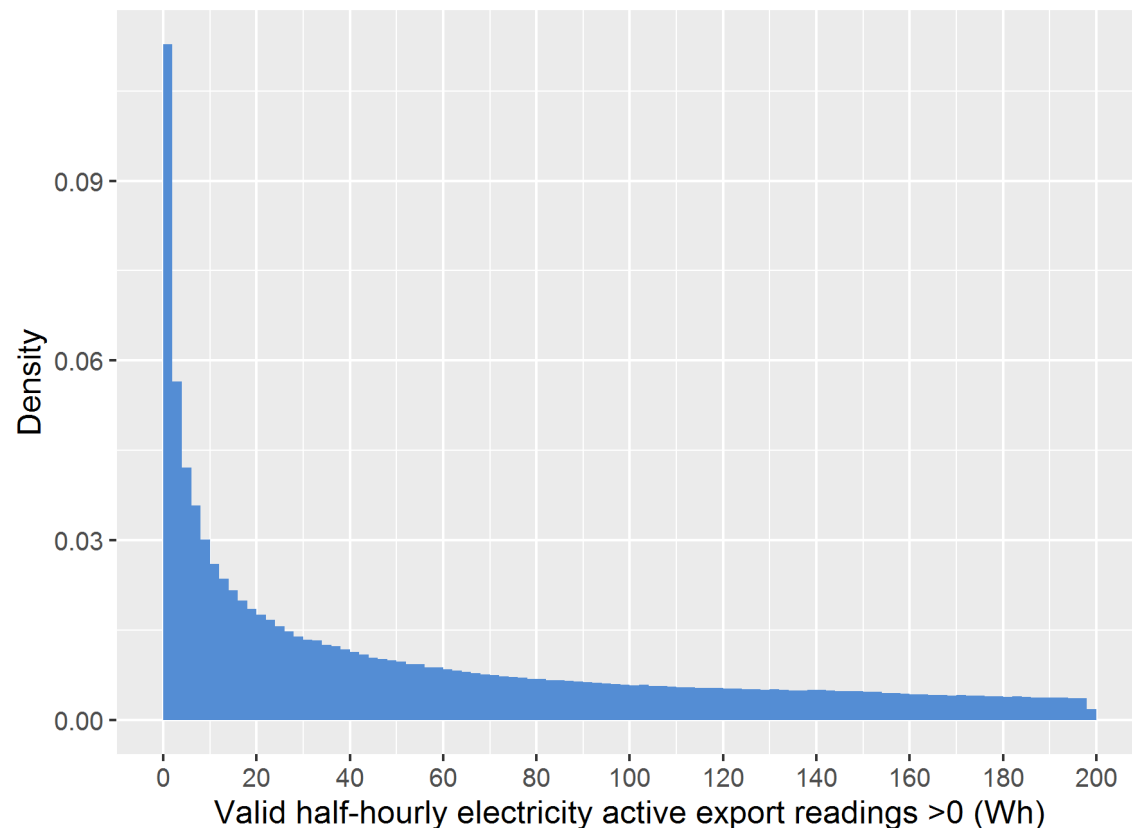
Figure 4: Histogram of half-hourly electricity active import readings, excluding those flagged as errors. Restricted to readings under 1000 Wh (1 kWh), bin width = 10 Wh.



Electricity active export

Figure 5 excludes readings equal to 0 because there are so many that the rest of the plot is difficult to see. There is still a spike of very low reads, as we would expect from times with low light such as during winter.

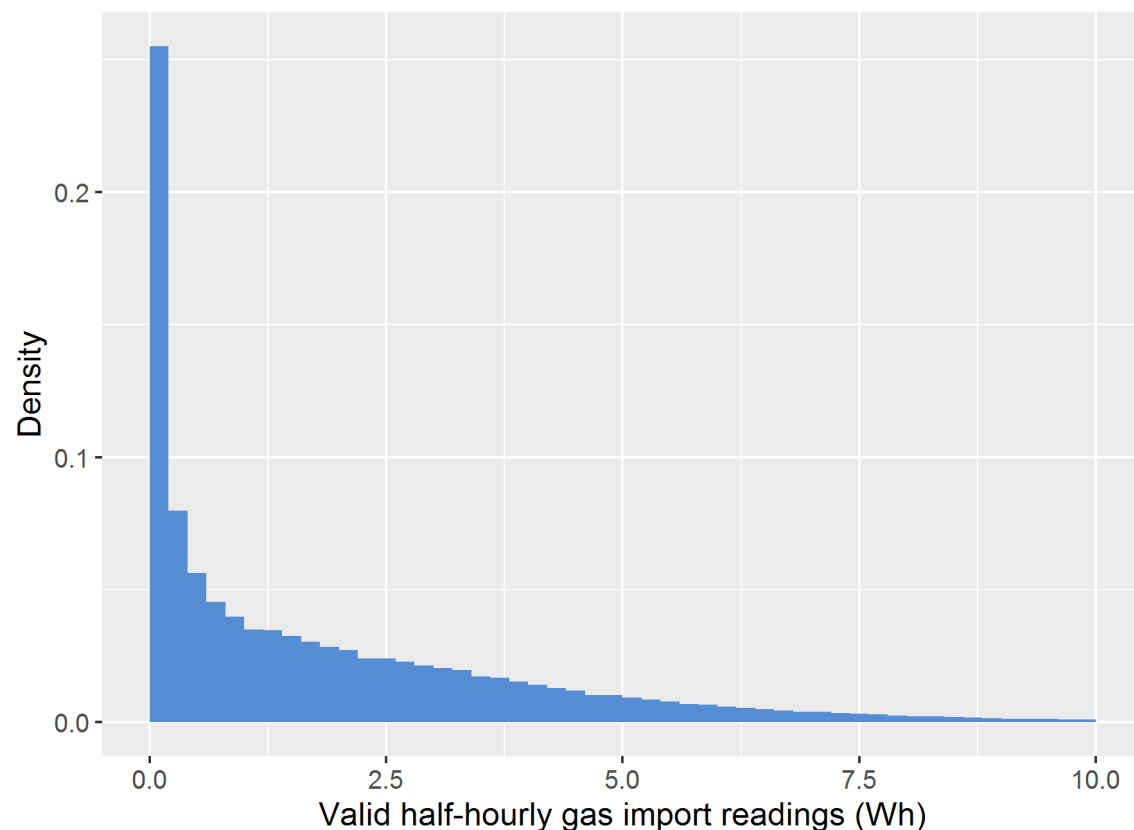
Figure 5: Histogram of half-hourly electricity active export readings, excluding those flagged as errors and excluding 0 reads as these dwarfed all others. Also restricted to readings under 200 Wh, bin width = 2 Wh.



Gas import

Gas usage is much higher than electricity usage as it is the most common source of heating in the UK. Therefore Figure 6 reports values in kWh rather than Wh. Note that gas data are recorded in m³ and converted into an estimate for kWh. Since the median half-hourly gas read is 0, 0 reads have been excluded from Figure 6 in order for the rest of the figure to be clearly visible.

Figure 6: Histogram of half-hourly gas import readings, excluding those flagged as errors. Restricted to readings under 10 kWh but strictly greater than 0, bin width = 0.2 kWh.



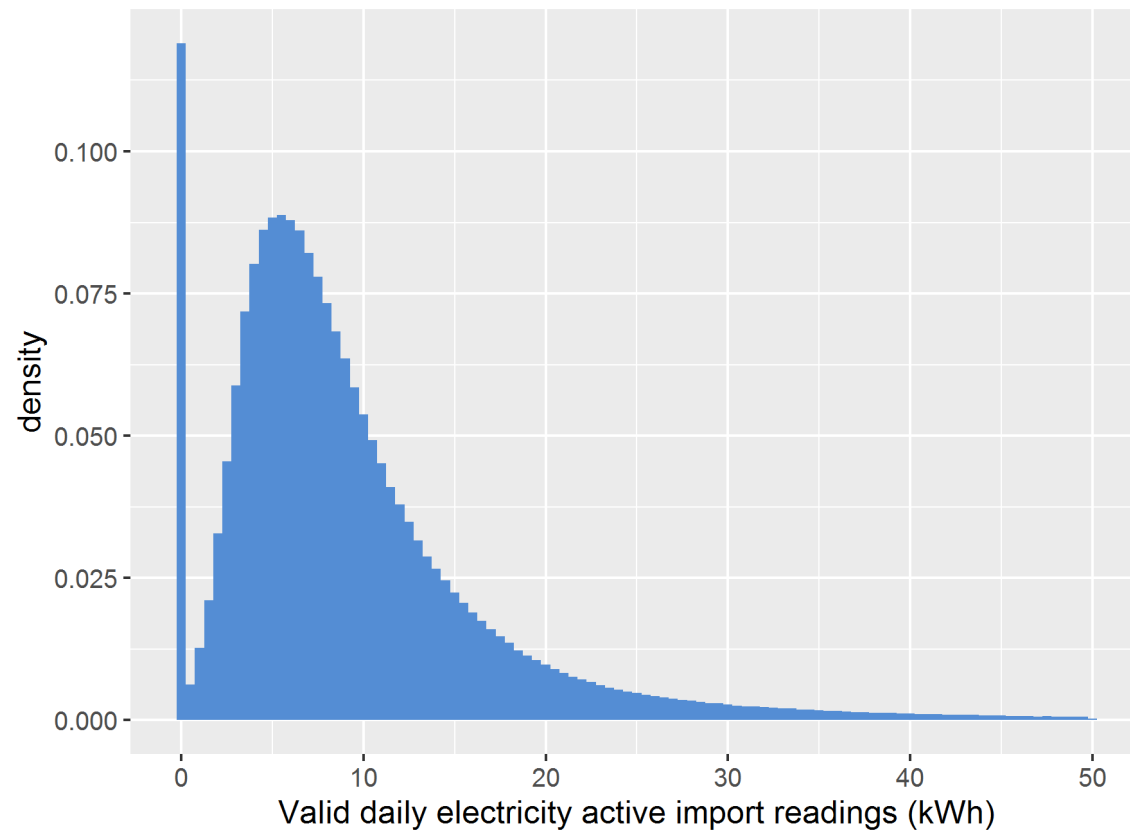
Daily reads

Electricity active import

While most of the distribution takes the expected shape for this type of data, there is a significant spike at 0. This could be a read error that has not yet been investigated, or some houses may have been left unoccupied for full days (however all electronic-consuming devices including fridges and freezers would need to have been switched off). It has not yet been

determined whether this is the result of a few houses with zero read data for many days or many houses with a few days with zero reads.

Figure 7: Histogram of daily electricity active import readings, excluding those flagged as errors. Restricted to readings under 50 kWh, bin width = 0.5 kWh.



Gas import

The daily gas import read distribution looks approximately as we would expect.

Figure 8: Histogram of daily gas import readings, excluding those flagged as errors. Restricted to readings under 200 kWh, bin width = 2 kWh.

