Smart meter consumption data:

Data quality report

Creation date	2020-08-21
Version	01
Author	Ellen Webborn
Project	Smart Energy Research Lab (SERL)
Organisation	University College London (UCL)

Table of Contents

Data availability

The following tables give some key statistics for the distribution of data quality across participants, and some basic statistics about the valid readings. For more information about error definitions, see the Errors section below or the *SERL_smart_meter_data_documentation_v2020_08.pdf* accompanying document. The data are presented as percentages on the next page. The abbreviation 'HH' stands for 'half-hourly'.

All of the import read types have some participants for which there were no valid readings available, however over 900 participants have at least 90% of gas and electricity active import reads available.

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

	Daily gas (im)	*HH gas (im)*	*Daily elec (active im)*	*HH elec (active im)*	*HH elec (reactive im)*	*HH elec (active ex)*	*HH elec (reactive ex)*
With a device	1277	1281	1703	1708	1708	110	110
100% valid	705	20	879	609	0	63	63
>= 99% valid	920	623	940	1055	65	86	86
>= 98% valid	1043	838	947	1258	72	96	96
>= 95% valid	1092	983	957	1319	860	99	99
>= 90% valid	1118	1067	992	1360	896	101	101
>= 75% valid	1159	1142	1009	1420	1538	103	103
>= 50% valid	1183	1185	1023	1449	1621	105	105
>= 25% valid	1203	1222	1285	1504	1655	109	109
>= 10% valid	1224	1235	1321	1528	1669	110	110
>= 5% valid	1228	1240	1328	1534	1672	110	110
>= 0% valid	1277	1281	1703	1708	1708	110	110

Table 2 shows the same data presented as a percentage. The first row of data ('With a device') shows the percent of participants we expect to have each type of read. Approximately 74% of households have a gas meter registered with the DCC (i.e. a SMETS2 gas meter or an upgraded SMETS1 meter). The percentages in the rows that follow are percentage of those With a device, e.g. 82.4% of households with a smart meter that we could collect data from have at least 90% of their readings available and valid.

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

	Daily gas (im) (%)	*HH gas (im) (%)*	*Daily elec (active im) (%)*	*HH elec (active im) (%)*	*HH elec (reactive im) (%)*	*HH elec (active ex) (%)*	*HH elec (reactive ex) (%)*
With a device	74.8	75.0	99.7	100.0	100.0	6.4	6.4
100% valid	55.2	1.6	51.6	35.7	0.0	57.3	57.3
>= 99% valid	72.0	48.6	55.2	61.8	3.8	78.2	78.2
>= 98% valid	81.7	65.4	55.6	73.7	4.2	87.3	87.3
>= 95% valid	85.5	76.7	56.2	77.2	50.4	90.0	90.0
>= 90% valid	87.5	83.3	58.3	79.6	52.5	91.8	91.8
>= 75% valid	90.8	89.1	59.2	83.1	90.0	93.6	93.6
>= 50% valid	92.6	92.5	60.1	84.8	94.9	95.5	95.5
>= 25% valid	94.2	95.4	75.5	88.1	96.9	99.1	99.1
>= 10% valid	95.8	96.4	77.6	89.5	97.7	100.0	100.0
>= 5% valid	96.2	96.8	78.0	89.8	97.9	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 a each month's data available and flagged as valid. 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).

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

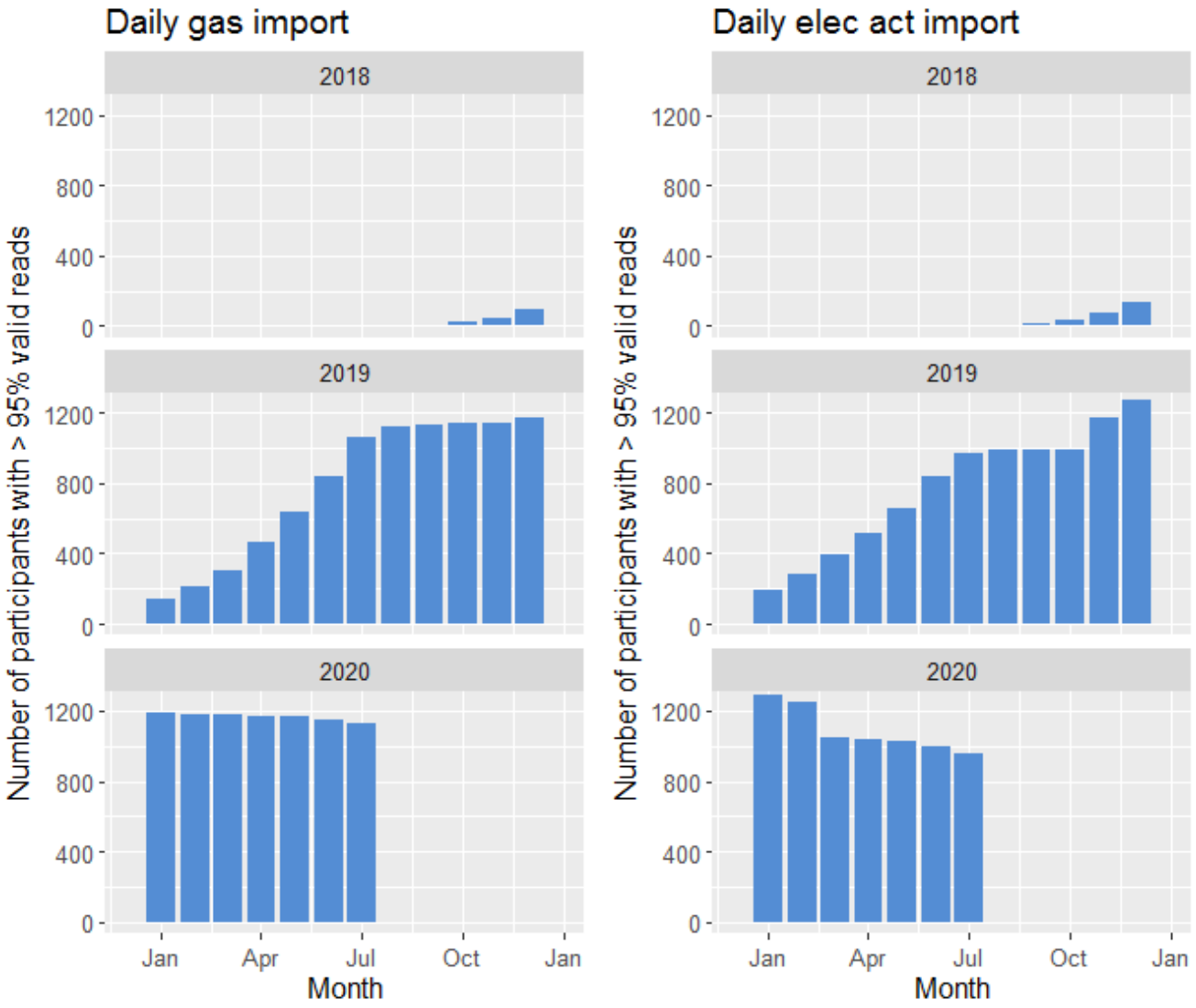
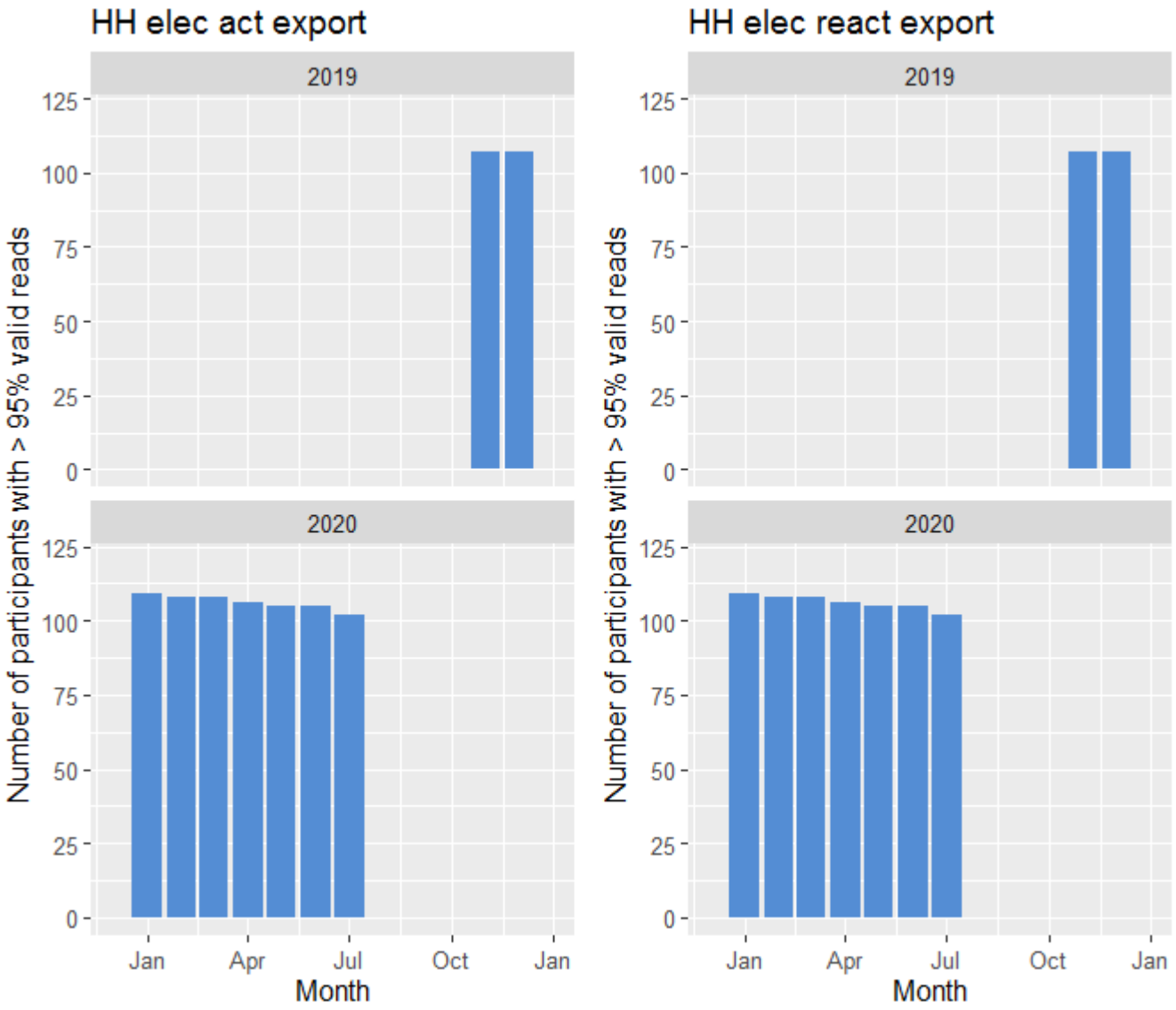


Figure 2: Number of participants in the SERL Observatory with at least 95% of half-hourly import data available and valid, by month.

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_data_documentation_v2020_08.pdf* accompanying document.

Half-hourly data

The half-hourly data runs from 2019-01-13 to 2020-07-31 and 1688 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 all 1708 participants have an electricity meter and 427 participants don't have a gas meter (25%). 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).

All participants have an electricity meter and 427 participants don't have a gas meter (25.00%).

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

Flag	*Meaning*	*Act elec import (N)*	*Act elec import (%)*	*Act elec export (N)*	*Act elec export (%)*
1	Valid	33190942	85.39	1347203	96.36
0	Missing	5614384	14.44	48975	3.50
-1	Max read	190	0.00	0	0.00
-2	Very high but not max	0	0.00	0	0.00
-3	Negative	0	0.00	0	0.00
-4	Elec in kWh	0	0.00	0	0.00
-5	Invalid read time	62196	0.16	1870	0.13

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)). Less of the reactive import data are missing than the active import data, but the same amounts of reactive and active export reads are missing.

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

Flag	Meaning	React elec import (N)	React elec import (%)	React elec export (N)	React elec export (%)
1	Valid	21977846	87.31	1347203	96.36
0	Missing	3132070	12.44	48975	3.50
-1	Max read	0	0.00	0	0.00
-2	Very high but not max	0	0.00	0	0.00
-3	Negative	0	0.00	0	0.00
-5	Invalid read time	62196	0.25	1870	0.13

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.

Flag	Meaning	Gas import (N)	Gas import (%)
1	Valid	26131765	90.98
0	Missing	2280097	7.94
-1	Max read	249363	0.87
-2	Very high but not max	149	0.00
-3	Negative	0	0.00
-5	Invalid read time	62162	0.22

Daily Data

The daily data runs from 2018-08-18 to 2020-07-31. A total of 1708 participants are included in these datasets. Of these, 1612 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 6% 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.

Flag	Meaning	Elec act imp (N)	Elec act imp (%)	Gas import (N)	Gas import (%)
1	Valid	553708	67.13	561375	92.28
0	Missing	215233	26.09	43703	7.18
-1	Max read	0	0.00	1611	0.26
-2	Very high but not max	5	0.00	93	0.02
-3	Negative	0	0.00	0	0.00
-4	Elec in kWh	54381	6.59	0	0.00
-5	Invalid read time	1540	0.19	1540	0.25

Read statistics

The following table shows some basic statistics for each read type. Only 'valid' reads were included for these statistics (error code 1), excluding those recorded in the wrong units, at the wrong time, or outside of the defined valid data range. These ranges were deliberately defined to be very conservative so as not to rule out plausible outliers at this first stage of data quality analysis. The distribution of reads for some of these read types are shown in the next section.

Table 7: Read statistics for valid readings of each type.

	Daily gas im (kWh)	HH gas im (kWh)	Daily act elec im (kWh)	HH act elec im (kWh)	HH react elec im (kvarh)	HH act elec ex (kWh)	HH react elec ex (kvarh)
Max	10314.00	10255.87	170.36	65.53	2.34	2.34	0.57
Mean	36.60	0.77	8.69	0.19	0.02	0.09	0.07
Median	20.49	0.00	6.93	0.10	0.00	0.00	0.06
Median exc. 0	23.33	0.90	7.25	0.10	0.02	0.16	0.06
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Standard Deviation	139.01	21.07	7.69	0.29	0.04	0.23	0.06

Distributions of valid reads

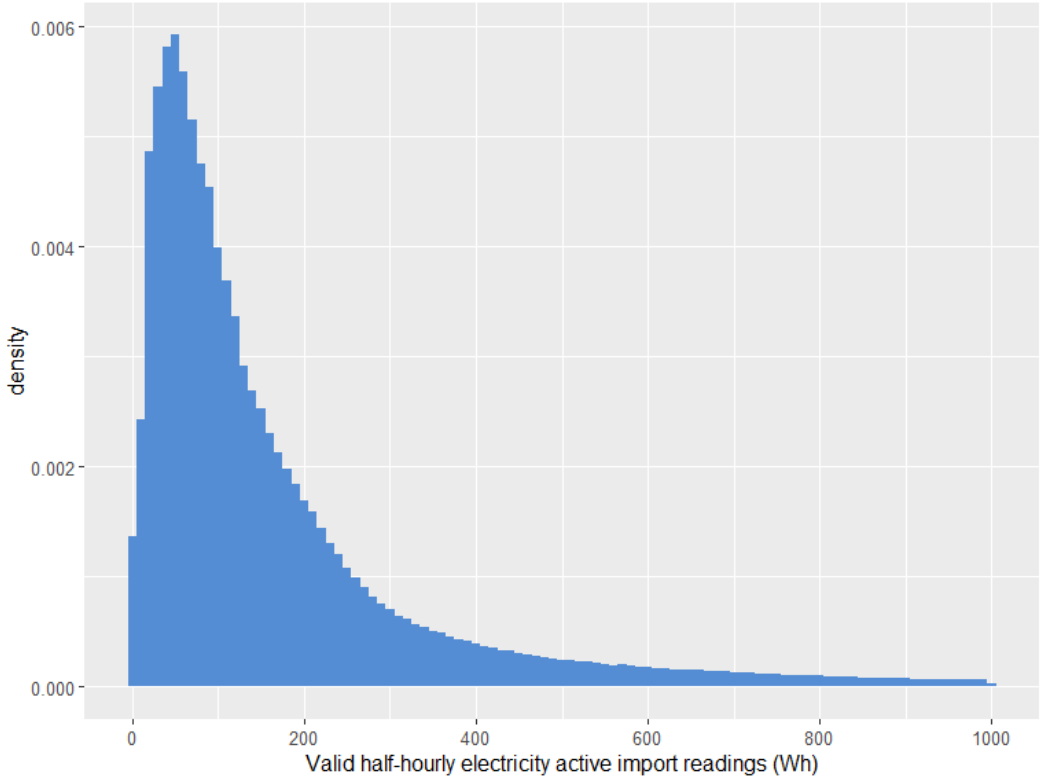
Half-hourly reads

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.

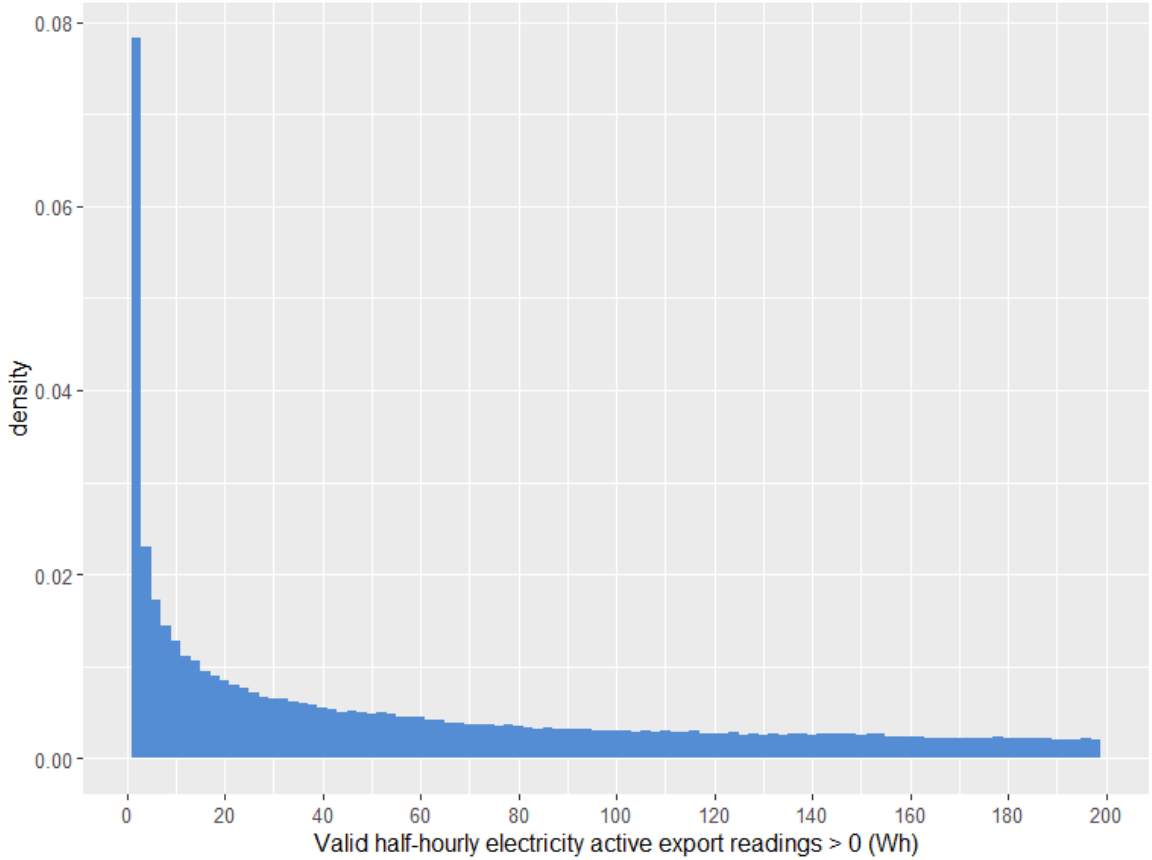
Figure 4: Histogram of half-hourly electricity active import readings, excluding those flagged as errors. Restricted to readings under 1000 Wh (1 kWh), bin width = 10 Wh.



Electricity active export

Figure 2 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.

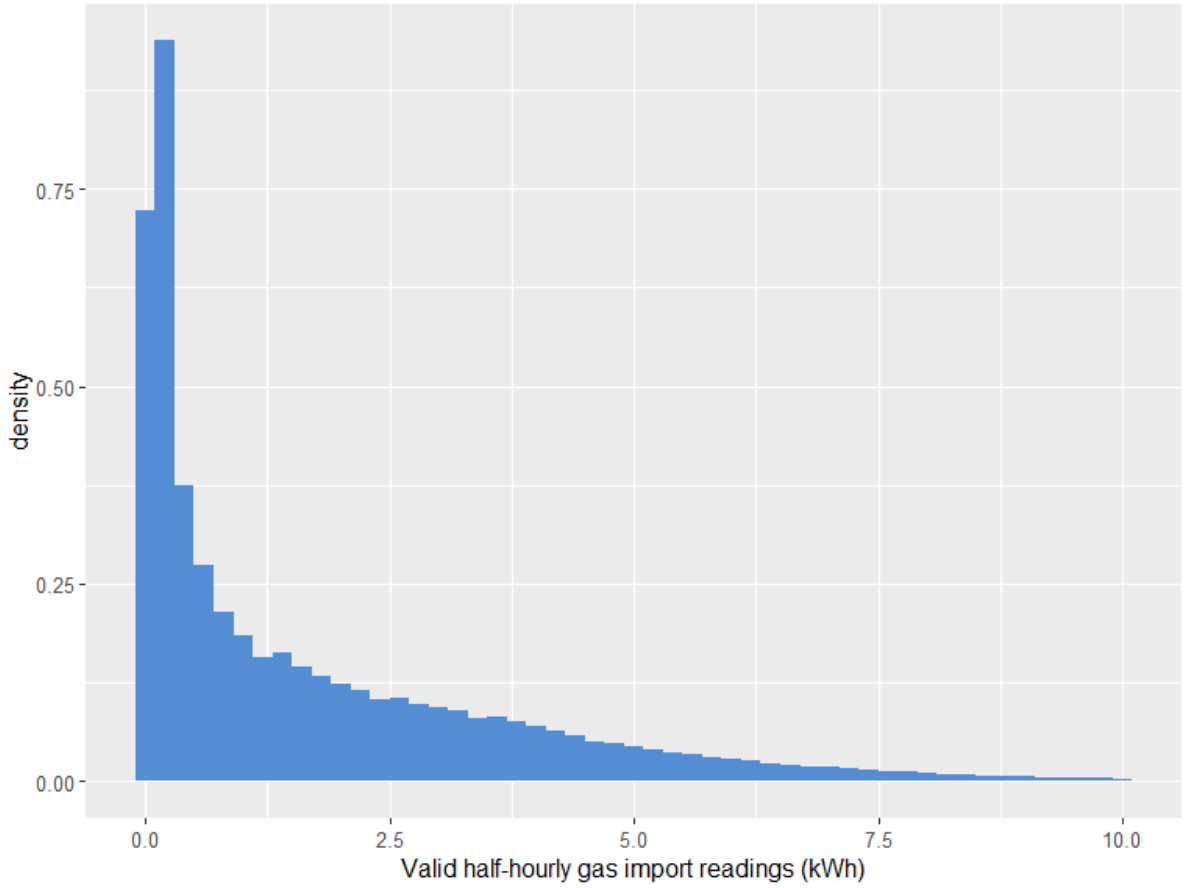
Figure 5: Histogram of half-hourly electricity active export readings, excluding those flagged as errors and excluding 0 reads as these dwarfed all others. Also restricted to readings under 200 Wh, bin width = 2 Wh.



Gas import

Gas usage is much higher than electricity usage as it is the most common source of heating in the UK. Therefore Figure 3 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 3 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.

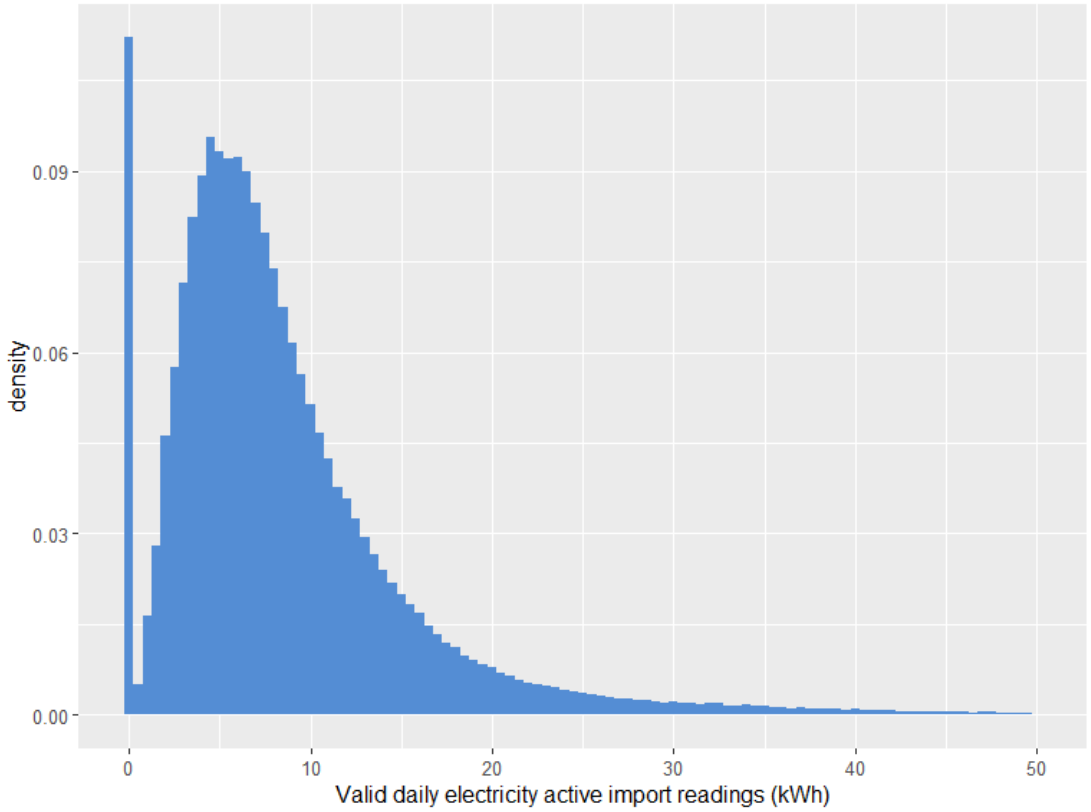


Daily reads

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. There is a large spike at 0 where heating is not used for a day.

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