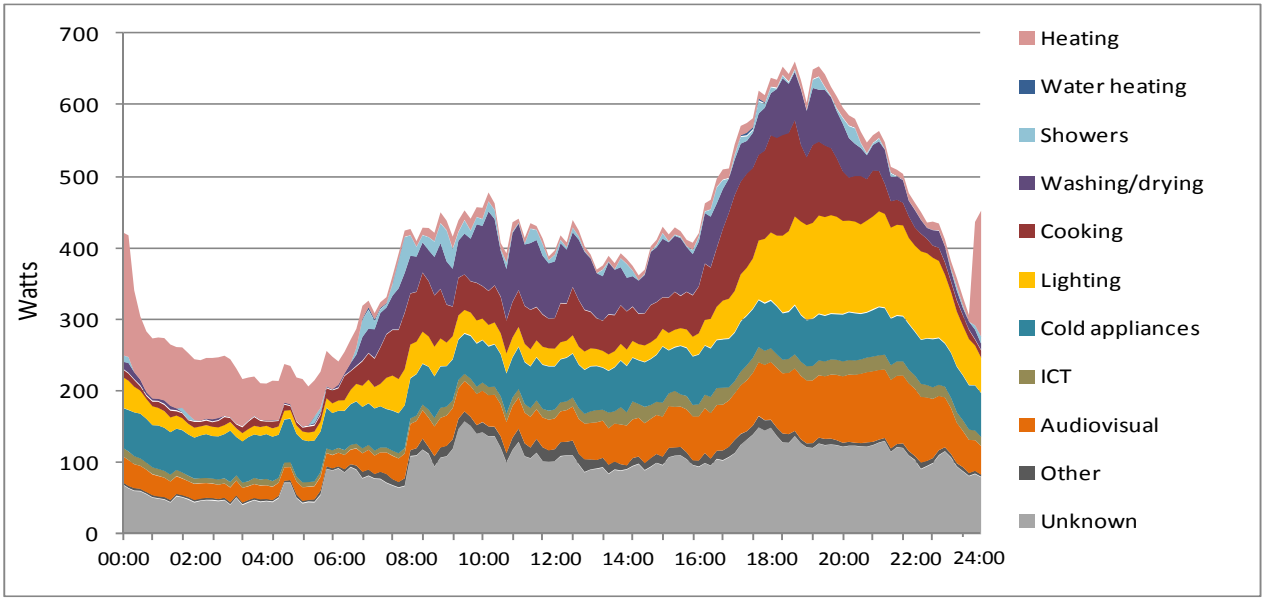
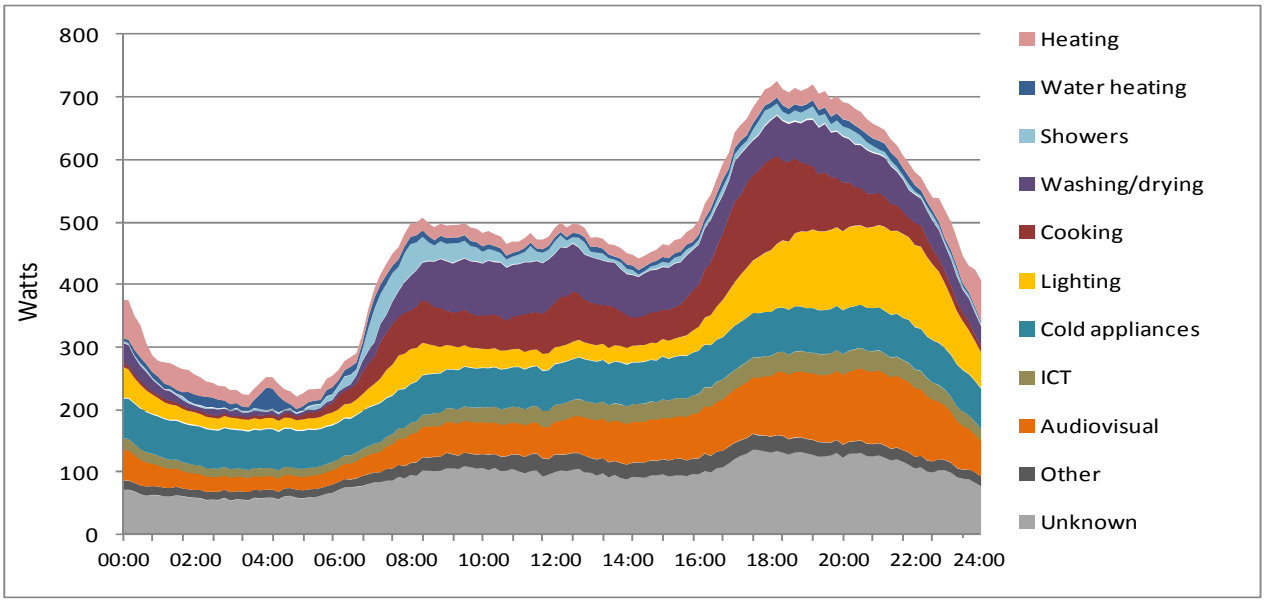
Household Electricity Survey:

Cleaning the Data

21 March 2013





Reference 475/09/2012

CAMBRIDGE
ARCHITECTURAL
RESEARCH
LIMITED

Contents

Cleaning the Household Electricity Survey Data

What sort of errors were there in the data?

The data from the HES survey had a variety of inconsistencies and errors in it that we have taken measures to correct. The kinds of errors we found in the original profiles were:

- Incorrect readings – where the measuring devices gave nonsense readings
- Misclassified appliances such as
    - A washing machine coded as a kettle
    - Appliances grouped on one profile, presumably when several devices were plugged into the same sockets, such as a profile coded as a desktop computer which also included a printer and monitor
- Overlapping profiles – for example duplicate profiles or where one profile included another
- Missing readings – usually for just a part of the day at the start and end of the monitoring period or sometimes a few readings dropped during the day
- Off the scale readings – one or two very high readings, sometimes as high as 10kW from an appliance which cannot possibly have used that power.

Also there were inconsistencies between appliance data and profiles – where there was appliance data for profiles that did not exist or vice versa.

Finally, data missing from the system meant it was impossible to calculate the total energy use for each household. Only 14 households had measurements taken coded as ‘mains’ and on some of these the device did not seem to be working most of the time. In most cases there were periods when the ‘mains’ recorded less than the sum of the measured appliances, which is not plausible.

For most households there were profiles described as ‘sockets’, which appeared to mean electrical circuits for wall sockets through the dwelling. However, it was clear from the data that not all circuits were monitored because the sum of appliance electricity use was often more than the readings from the ‘sockets’. We have done our best to determine which appliances were on unmonitored circuits by comparing the profiles of sockets and appliances. The appliances which were not monitored at the circuit level had to be added to the total of sockets and light distribution circuits to obtain an overall total consumption.

Identifying errors and fixing them

We identified missing readings by counting the number of readings in each profile per day. Where there was a significant part of the day missing we deleted those profile days so as not to distort statistics based on daily usage.

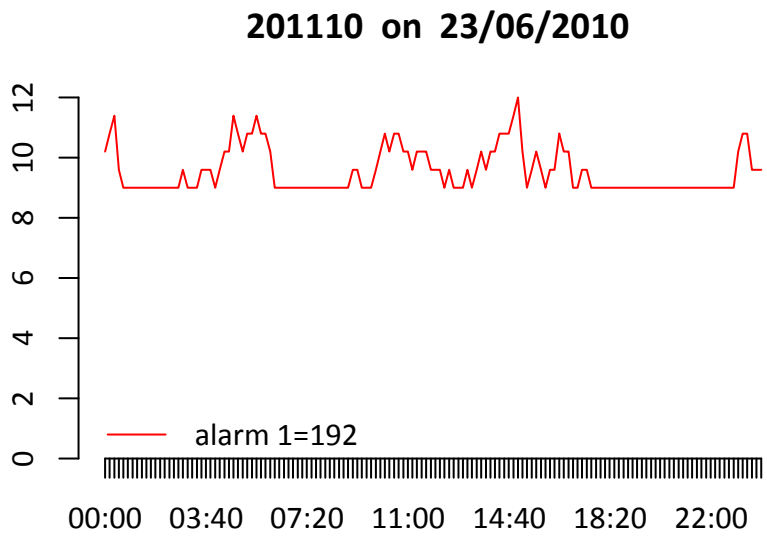
We identified some of the misclassified appliances by analysing the appliance data file. This file

records information about appliances such as the type, model and energy performance rating. The type of appliance can be compared with the appliance code used - if a kettle code is used for a washing machine, for example, there is a clear inconsistency. A program with simple heuristics was able to cull the number of possible inconsistencies down to 330, which were individually reviewed manually, revealing 45 errors plus additional inconsistencies between the appliance data and the profile data. However, the appliance data file only had information for 4115 appliances out of 4980.

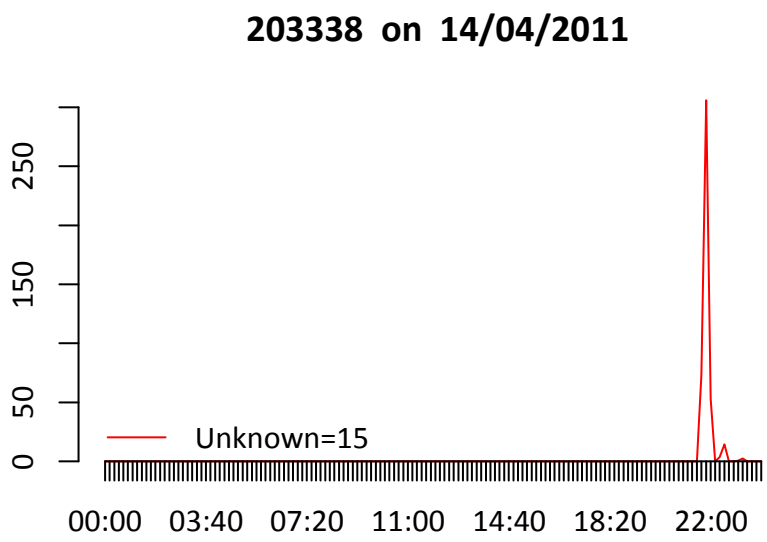
We took this opportunity to add some new codes – distinguishing a spin dryer from a tumble dryer and a slow cooker from a cooker.

We have identified other cases by inspecting daily usage profiles by eye. We have developed tools (see below) to facilitate comparing profiles from different houses for similar appliances and from the same appliance on different days through the year (when available). We have added appliance codes 'Unknown' and 'Unknown kitchen' for cases where we are sure the profile does not match the appliance it is supposed to be, but we can't be sure what it actually is.

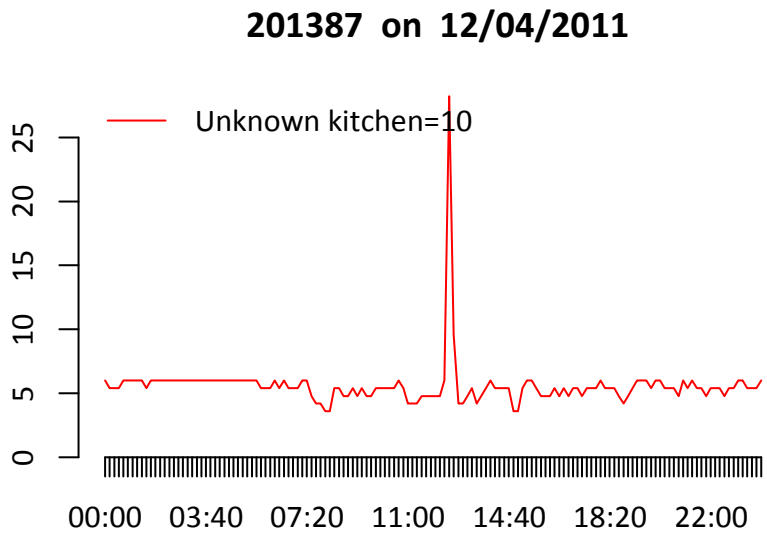
For example, below is a typical profile for an alarm system. Energy use is fairly steady, low power all the time.



Below is a profile also labelled ‘alarm’, which we concluded could not possibly be an alarm system, so we reclassified it as ‘unknown’.



The profile below shows an appliance that was supposed to be an electric cooker. The power use is never more than 30W, so we reclassified this as ‘unknown kitchen’.



Some profiles had single-readings of very high values (such as 6 kW on a TV) that we concluded must be noise in the meter readings. We edited them to replace those readings with the value from the previous correct reading.

Looking at daily appliance use over a period, high or low values can indicate problems with those readings – for example ‘mains 1’ for household 101010 recorded nothing at all for days at a time, and high values at other times led to identifying erratic readings during other periods.

We have prepared a data file which records all the reclassifications and other adjustments made to

the profiles. This file is included as the Appendix to this report.

Impact of the cleaning operations

Although the total number of appliances affected by the cleaning was small, the impact was significant in some cases. For example, twenty cookers (out of 144) were recoded as other appliances or deleted.

- One was a gas cooker, which used electricity only for its fan.
- Eight were recoded as extractor fans, slow cookers, or other appliances as they were described in the appliance data
- Six were deleted as they were not used at all, which is very unlikely for an electric cooker. They were more likely to be slow cookers, extractors or some other appliance.
- Five were recoded either as 'Unknown' or 'Unknown kitchen' because they had implausible profiles for a cooker, such as very low power use, up to only 30W or less.

Since all these profiles had much lower electricity use than a genuine cooker, removing the profiles increased the *average* energy use calculated for the genuine cookers.

Another high impact profile change was reclassifying the main lighting circuit for one household, which was recording very high readings throughout the year. There were only 26 households recorded for a whole year; we used them to estimate seasonality factors for each appliance type, which we needed to scale up data from the other households from the month of monitoring to a year. Removing one consistently high-energy lighting appliance from the set of annual readings affected the seasonality factors for lighting as a whole.

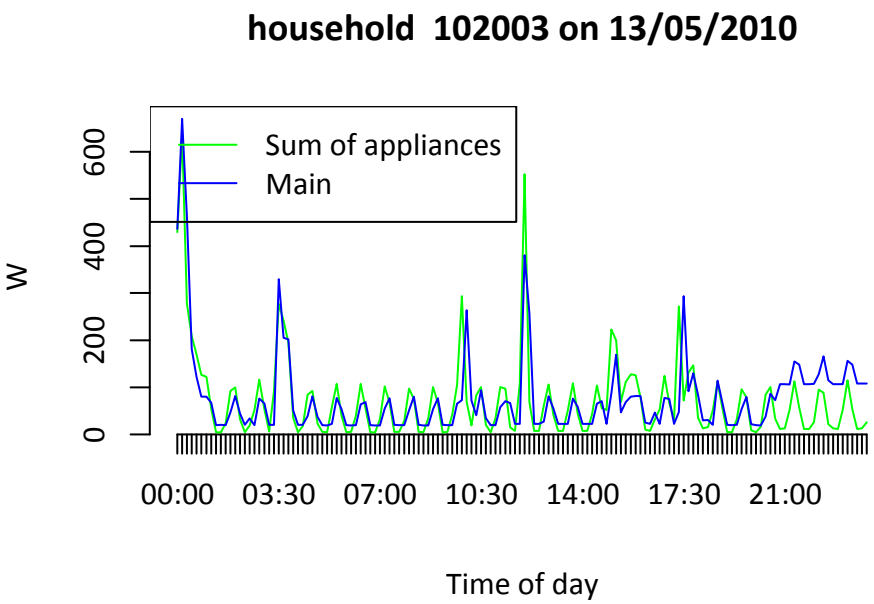
Finally, there were a few cases where there were very few examples of an appliance monitored, where adding or removing even one or two examples made a difference to the average. For example, recoding removed one out of only fifteen alarms and one out of six radios. Also recoding added seven kitchen extractor hoods for a total of forty eight, and added a cooking steamer to make a total of three steamers.

Identifying total household consumption

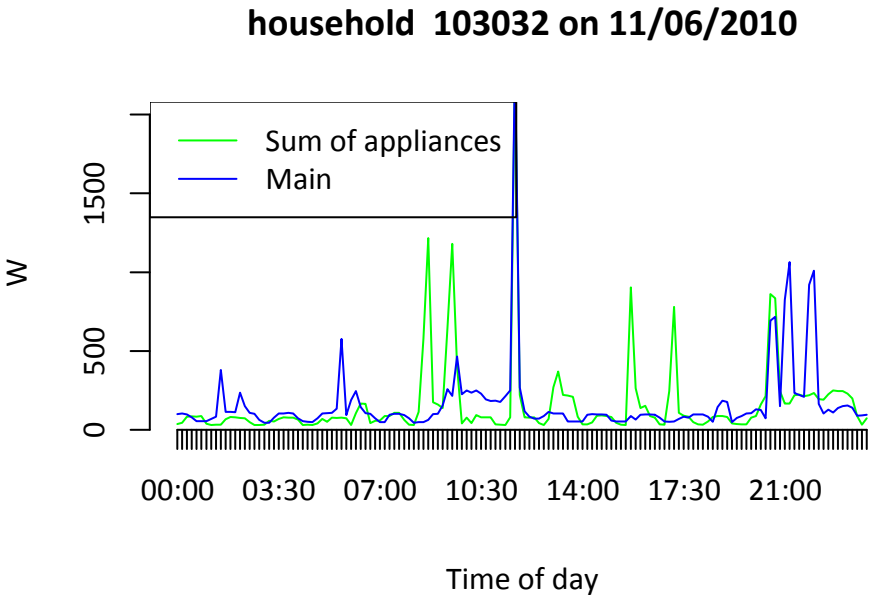
For households with profiles for 'mains' we compared visually the mains profile with the sum of identified appliances (not including 'sockets'), both over a range of days and for sample daily profiles. We were looking for:

- Correlation between peaks and troughs in 'mains' and appliance readings (though sometimes with a time delay)
- Periods when appliances use was higher than 'mains', which should never occur if the 'mains' is recording all appliance use.

Below is an example of good correlation:



Below is an example of bad correlation:

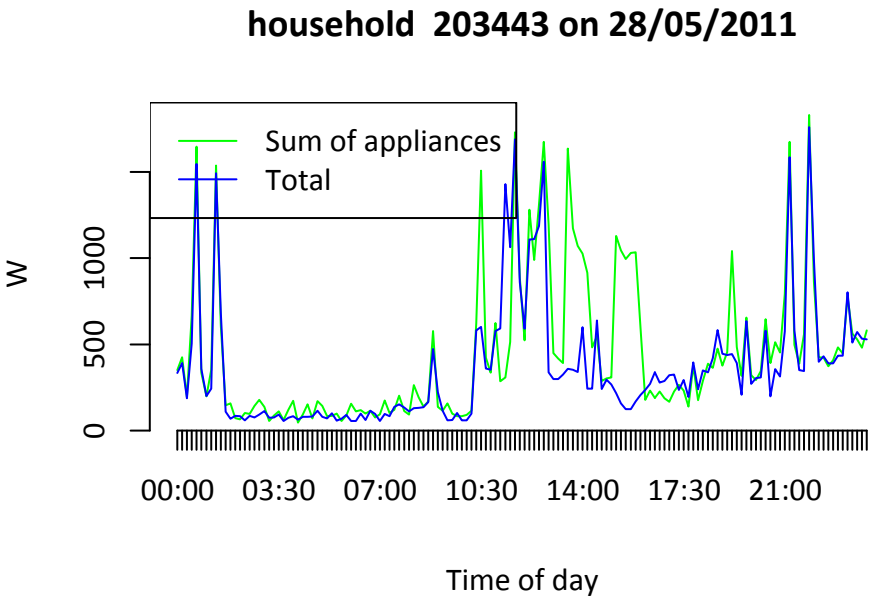


For five of the 14 cases where 'mains' was measured, we have sufficient confidence in the 'mains' profile to use this as the total consumption.

For cases where sockets have been measured (216 households) we compared the sum of sockets with the sum of specific appliances, and identified appliances that were not recorded by the sockets by looking for peaks in their use that did not show up in the socket profiles. For each household, we have a list of 'total_profiles', which had to be added up to derive the total household consumption – where the mains data was not available or not useful we started with the lighting distribution and socket profiles and added in the other appliances, which we conclude were not included in the sockets. For example, in most houses this included the cooker and some

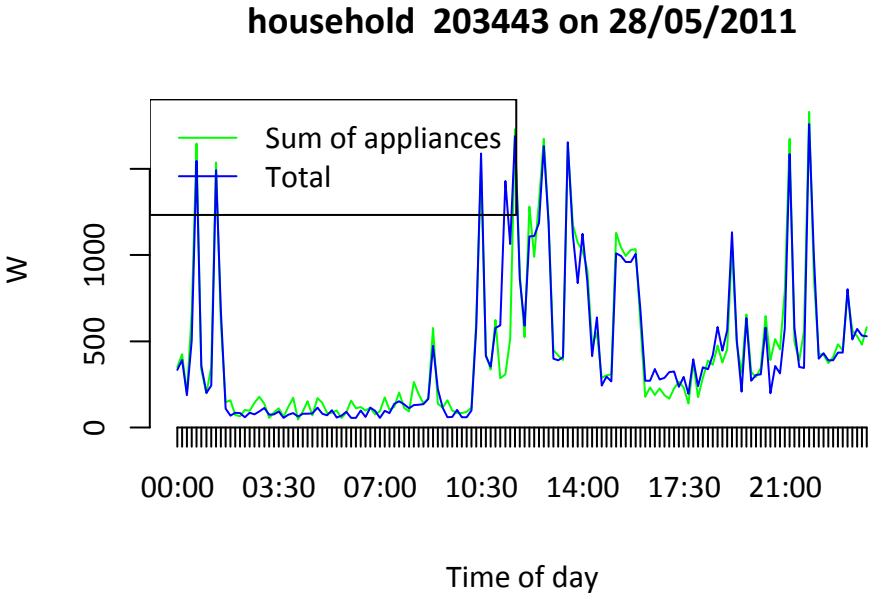
kitchen appliances, and often an immersion heater if there was one, because these are often on separate circuits in the consumer unit.

Here is one house starting with just sockets and lighting circuits in total_profiles:

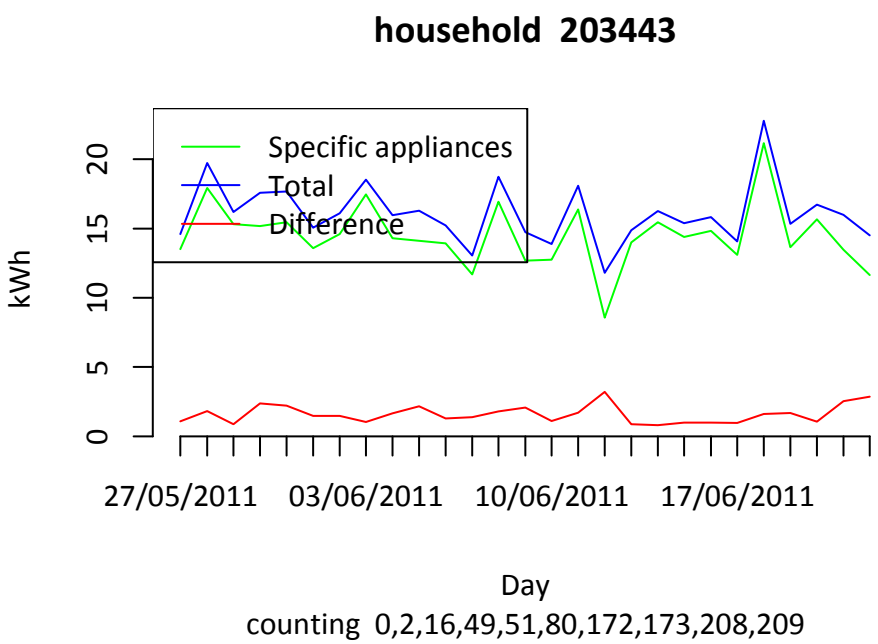


There was a peak around 11am, and more peaks in the early afternoon, which were not recorded by the sockets. By comparing these with individual appliance profiles we identified them as the cooker, tumble dryer and washing machine.

Adding these gives a fair correlation for the day:



We looked at several days for each household because not all appliances are used every day. We also compared the total usage on each day of the measuring period:



There are some occasions when the sockets still read less than the appliances, and it is not possible to be completely certain because not all appliances have a distinctive profile.

We have not tried to identify which appliances are on each socket (circuit) as this is not necessary to determine the overall consumption.

Having identified the set of profiles which make up the total household usage we populated daily and annual total tables with appliance 2000 representing the total and appliance 159 comprising the difference between this and the sum of the known appliances. In a few cases, because of inaccuracy in identifying total_profiles, the sum of the energy use by total_profiles was less than that of the appliances – we then recorded the unknown use as 0 and increased the total to be consistent with the sum of the appliances.

Seasonal Adjustments

Because most of the households were only monitored for a month, we needed to apply a scaling factor to estimate annual energy use and for this seasonal differences had to be taken into account. For example fridges and freezers use more energy in the summer but lighting is used more in the winter. We used the data from the 26 households monitored over a whole year to generate seasonality factors for each appliance type – for cold appliances, cooking, lighting, washing, AV, ICT, water heating and space heating. For water heating there was no significant difference between the seasons.

To calculate the factors we:

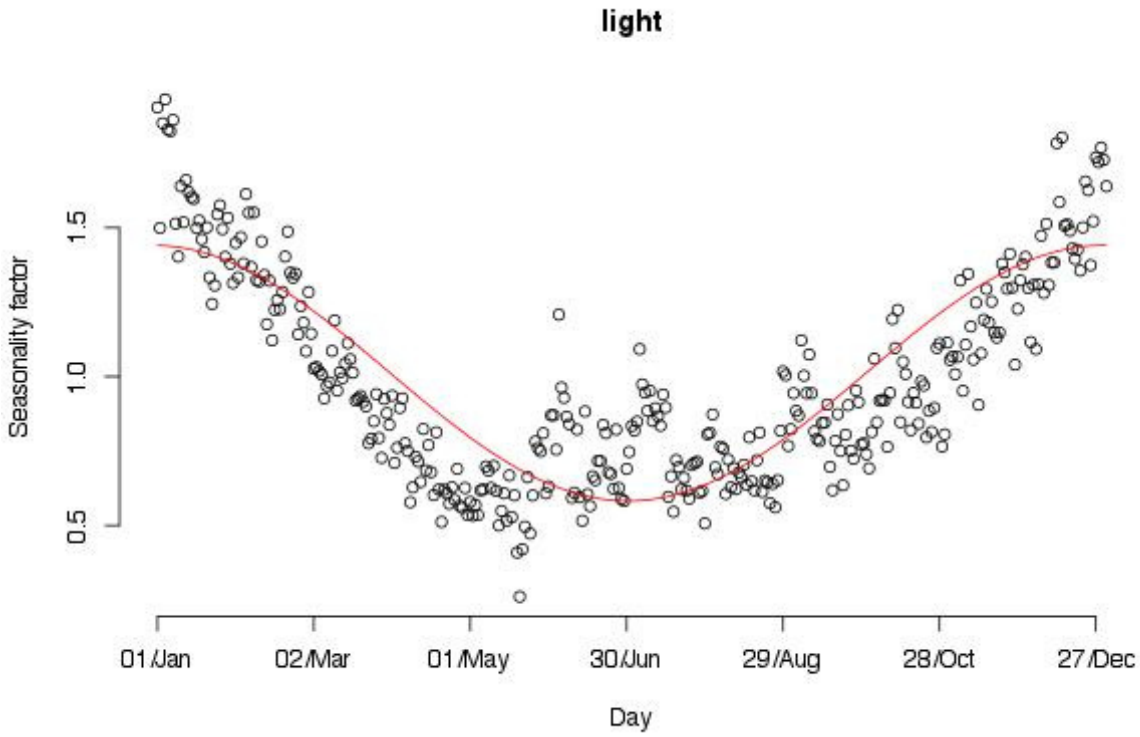
- calculated the electricity use on each day for each appliance type, averaging over the total usage by the 26 households
- normalised this by dividing by the total use over the year, times 365 to get a factor for each day.

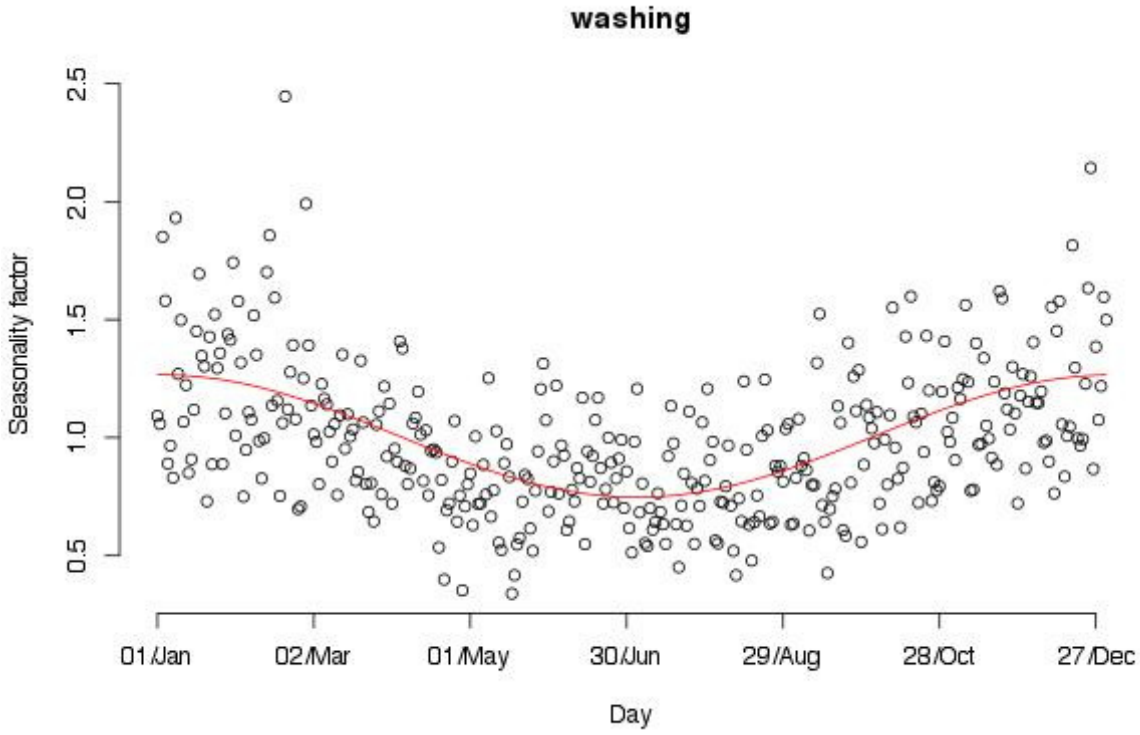
The results were very noisy, the sample size being rather small, so we used regression analysis to

find a best fit curve. Since we only had one year of data, but we expected the curves to be continuous from one year to the end, we doubled up the data to show two years i.e. two complete cycles, and fitted the curve over the whole period.

We tried using a polynomial formula based on day number (to order 5) but the result was poor. We did better with a curve based on sine and cosine functions, with 365 days scaled to 2xPi for the annual cycle. Since a good deal of energy use is based on day length and temperature, both of which cycle through the year, this seemed reasonable. We also tried using $\sin^2$ and $\cos^2$, but this produced bumps in the curve which were hard to explain; we decided they were more likely due to noise in the data than to real effects.

Ultimately we used a curve based on sine and cosine functions, and then computed least squares to fit it to the data , and the results are plotted below.





From the fitted curves we stored a seasonality factor for each day and appliance type. Then to scale up appliance use for each household from the period of monitoring up to a day, we:

- divided the electricity use on each day by its factor to get the normalised energy use
- found the average for all the monitored days, and
- multiplied by 365 to get annual use.

MPAN Data

We were given a dataset of annual meter readings for most of the households that were monitored for just one month (MPAN dataset). The data included total annual electricity consumption for the year 2010, which overlaps with the monitoring period of the survey. We used this to check the accuracy of our calculations for annual electricity use, which depend both on seasonality adjustments and on identifying appliances to include in total household usage ('total_profiles').

We matched the MPAN dataset to the households in the survey. There were 201 matching households. We compared the annual electricity use estimated from the survey with that recorded at the meters, and investigated cases with a serious mismatch. The following charts show the results, with points for outlying households labelled, each of which we is described in the table at the end of this section. Based on this we made some adjustments to our analysis, as indicated in the table on p14 below:

- In two cases we discarded socket readings, which were unreasonably high
- In one case we ignored seasonality factors for a lighting circuit
- From another case we decided that the use of a central heating (pump/controller) was not seasonal (which also affects some other households slightly).

Agreement for the households without electric heating was fair, for households with additional heating it was less good and for households with primary electric heating it was very poor indeed. Correlation was better when the households were monitored in the winter period (when a smaller seasonal adjustment is needed) but there was no obvious seasonal correction in the others. There were only seven matched households with primary heating for comparison, only one of which was monitored in the winter.

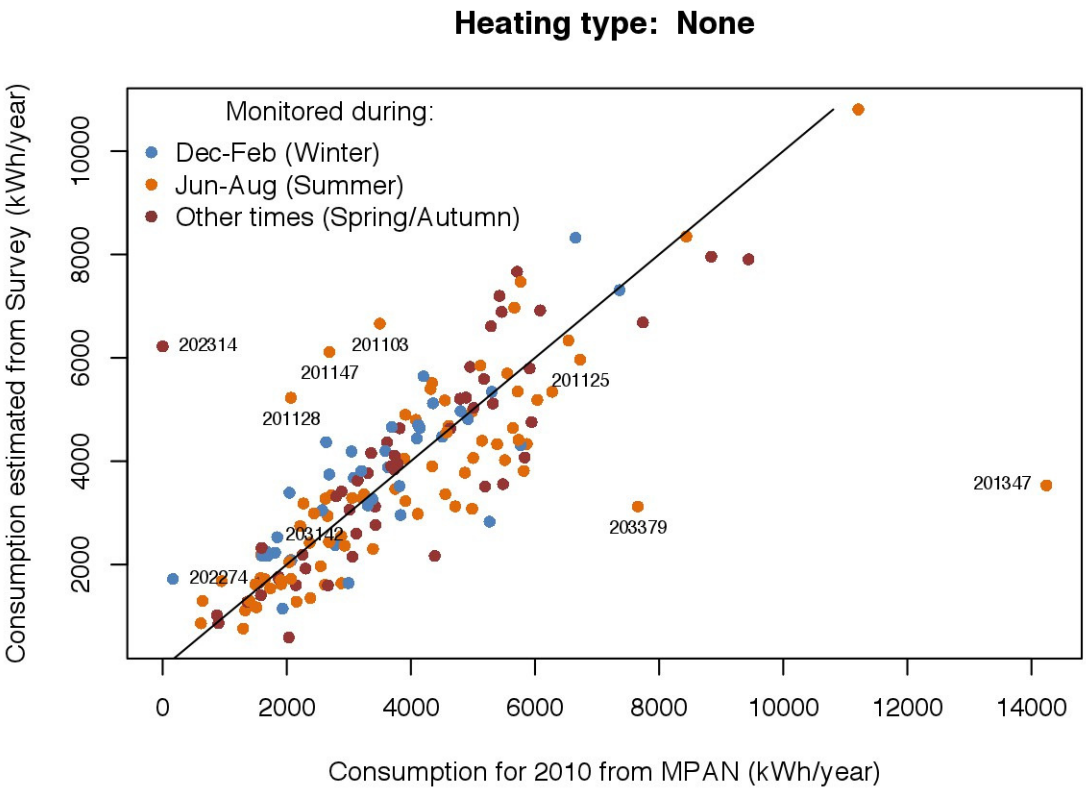
The table below compares mean electricity use for households by heating type.

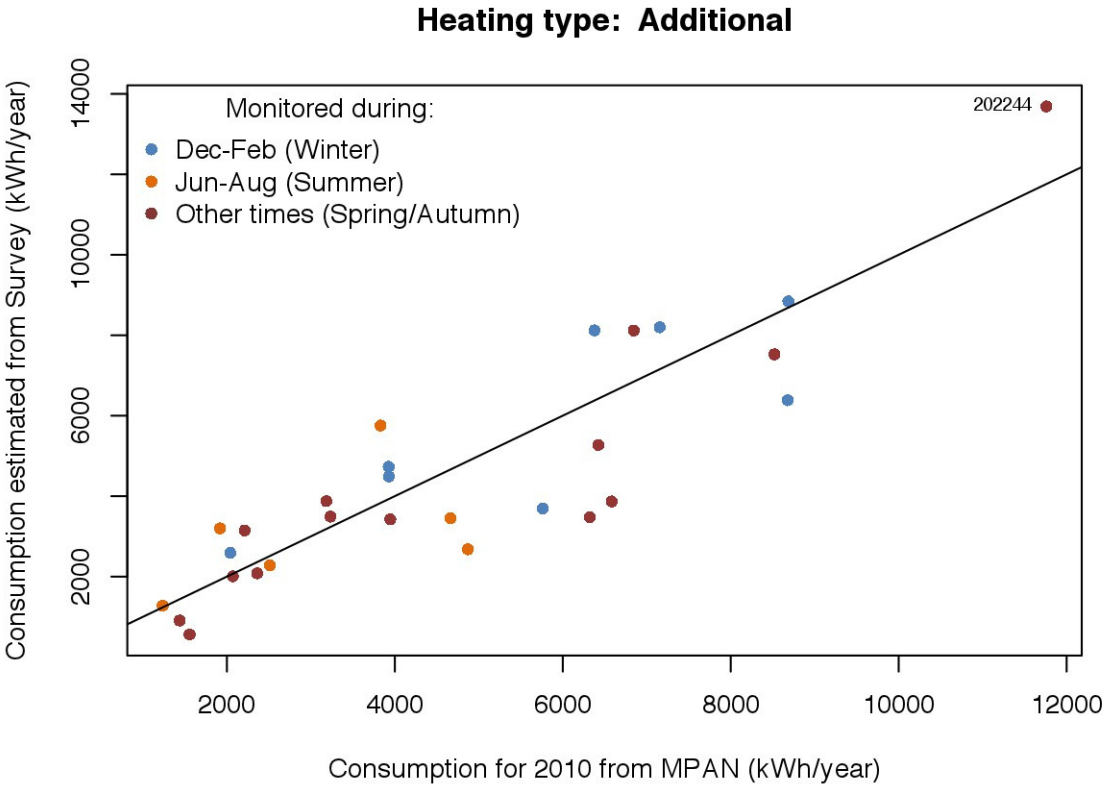
Heating Type	**MPAN dataset: mean kWh/year**	**Survey estimate: mean kWh/year**
None	3783	3734
Additional	4716	4541
Primary	7817	6984

However, we know that MPAN data itself is subject to errors and uncertainty. When DECC uses it for domestic analysis they exclude anything with a consumption of less than 100 or more than 25,000 kWh. They also exclude anything that looks like an estimate (primarily readings where a property is given the same consumption as the previous year, but also some values which seem to appear too often to be actual readings). There are also some cases where a household has been mis-matched i.e. consumption matched to the wrong property. The Department believes these are minimal, but they are currently doing internal work to look at this in more detail[1]. This means, overall, that you would expect a discrepancy between MPAN and our monitored data, even with a full 12-months of monitoring.

[1] Private communication with Mary Gregory, DECC Statistics, March 2013.

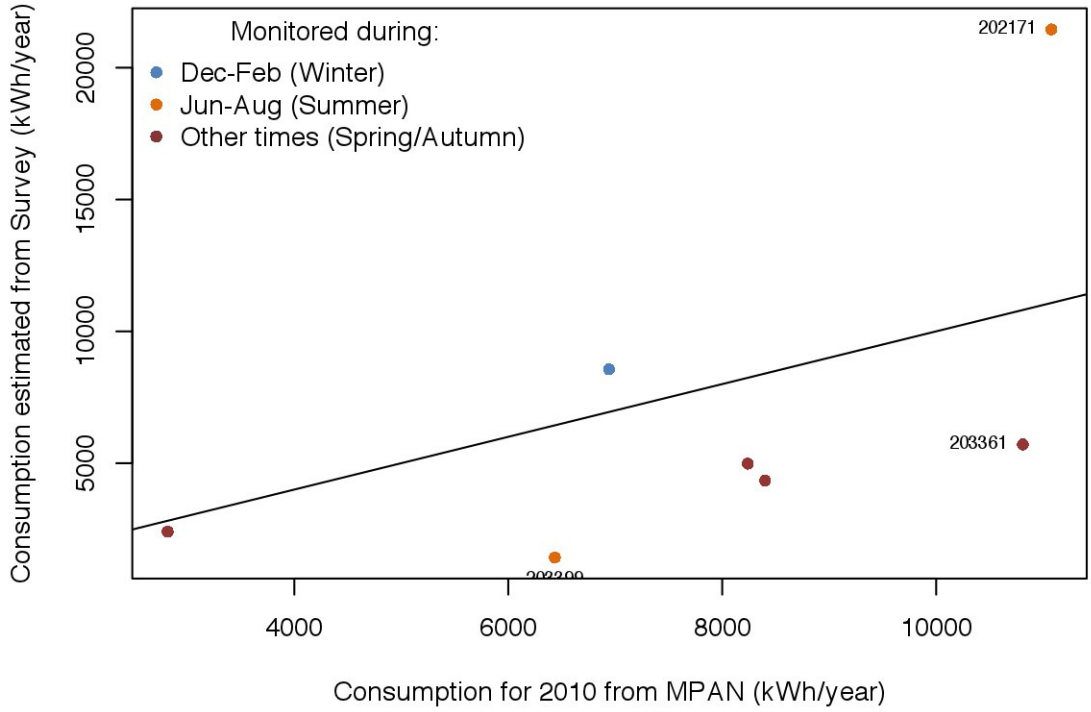
The charts below compare our estimate with MPAN data after adjustments have been applied.





There was good agreement between MPAN and our estimate for the single household with primary heating monitored over winter. In other cases it was poor: sometimes too low and sometimes too high.

Heating type: Primary



Household (Heating type N-none, A=additional, P=primary)	MPAN Consumption 2009	MPAN Consumption 2010	Our estimate kWh/year (before/after adjustments)	Comments/ Action
201103(N) Monitored Jun-Jul	6768	3501	6663	This house had a monitored mains recording 4709 kWh/year: already larger than the MPAN reading. Also some appliances had to be added to ensure that the total usage was more than monitored appliances. The total we found is not much different from 2009 so the MPAN dataset for 2010 could be wrong. No change.
201125 (N) Monitored Jul-Aug	5324	6726	9870/ 5967	Our total_profiles analysis found little correlation between sockets and appliances

				and had assumed many appliances were on unmonitored circuits, with a large proportion of use unknown. We changed this to ignore the sockets altogether.
201128 (N) Monitored Jul-Aug	6780	2067	5228	The MPAN reading for 2010 is very low: much lower than for 2009 and our estimate is in between so MPAN could be wrong. On the other hand, this house uses a lot of lights: 3 kWh/day on the light distribution circuit even in July. It is possible that for this household light use is less seasonal than usual, or there are loads on this circuit which are not in fact light. However, changing the seasonality factor for this appliance would not help enough to align these data. No change.
201147 (N) Monitored Jul/Aug	3937	2687	7930/6111	No sockets were monitored on this house but one light distribution circuit accounts for ¾ of use (recording 6-18 kWh/day even in August, mainly but not exclusively overnight). It may not be lighting after all, or it may be lighting which is on a timer and not seasonal. We adjusted the analysis to exempt this profile from seasonal adjustments but the overall use is still very high.
201347 (N)	11912	14249	3529	We are unable to

Monitored May/Jun				explain our low estimate. Only one socket circuit was measured and there were some unmonitored appliances including one TV , three set top boxes, a desktop computer and monitor. One of the appliances we did monitor is an aquarium, which could be seasonal, but this is unlikely to use more than about 1000 kWh/year. No change.
202155 (N) Monitored Aug/Sep	4006	4317	21,947/ 5398	The socket readings for this household were very high and the MPAN data suggests these are erroneous. We reset the analysis to ignore sockets. However, our estimate is still rather high.
202171 (P) Monitored Aug-Sep	11,309	11,077	21,450	One heater uses 4-6 kWh/day – in August/September. Seasonal adjustments lead to very high consumption on this one appliance. Possibly seasonality on this appliance is anomalous. However, if we assumed use was the same all year our reading would be too low. No change.
202244 (A) Monitored Nov-Dec	15,199	11,759	13,688	The MPAN value for use in 2010 is much lower than that for 2009 and survey data is not much higher than 2009. A large part of the demand comes from an appliance labelled 'garage' with no details – if this is seasonal and

				higher in winter this could explain why our estimate is high. No change.
202274 (N)	1845	166	1721	MPAN data for 2010 is wrong.
202314 (N)	7563	0	6216	MPAN data for 2010 is wrong. No change.
203142 (N) Monitored Aug-Sep	2828	2439	6240/ 2991	Most of the annual heating was 'central heating' which we concluded must be controller and pump. This house had the highest daily use for the appliance at any time of year (1.2 kWh/day) but there were only 6 other examples, 4 using any power at all. We adjusted the seasonality analysis so that this appliance was not regarded as seasonal, since pumps and controllers are needed all year.
203361 (P) (Single pensioner, in a flat, no heating controls) Monitored April-May	6558	10,810	5596	Our estimate includes one heater but less than a quarter of consumption is for the heater – it is hard to extrapolate from households monitored outside the heating season. No change.
203379 (N) monitored April-May	8295	7656	3123	This house has two storage heaters and a portable heater which were not monitored at all. It is possible these are used for supplementary heating in winter, though maybe not used during the monitoring period. No change.
203399 (P)	7916	6435	1425	No heating at all was

Monitored **May-June**	used during the monitoring period so we are unable to estimate heating energy, and this estimate excludes heating, hence the low reading. No change.

Appendix 1 – Reclassifications Table

This table shows the appliance classification changes we made during data cleaning. The diagnosis can be:

- Appdata: inspection of appliance data
- Profile: inspection of the energy use profiles

Household	FromApp	ToApp	Diagnose	Notes
101010	central heating 1	Unknown	Profile	Not central heating
101018	tv 2	fryer 1	Appdata	TV is fryer
101020	light distribution 3	Unknown	Profile	This light distribution profile changes during the year. From 2011-01-01 it has a high standby power and big peaks for about a month, then in February the standby disappears but the peaks remain
101038	microwave 1	Other cooking	Profile	From 23/11/2010 this microwave seems to be joined by something else on the same profile. It runs for an hour or more at 100W
101038	tumble dryer 1	Unknown	Profile	Too spiky to be a tumble dryer
102004	tv plasma 1	audiovisual site 3	Appdata	TV plasma to av site
102024	light distribution 1	Unknown	Profile	light distrib at up to 6kW and not seasonal.
103014	tv plasma 1	audiovisual site 1	Appdata	TV plasma to av site - lots of appliances
103029	shower 1	Unknown	Profile	Showers don't last 2 hours. Also very rarely used.
201103	cooker 1	Unknown kitchen	Appdata	This is a gas cooker - the only electricity is for the fan
201110	sockets 1	light 1	Appdata	socket to light - used only rarely and 10 W
201125	dvd 1	audiovisual site 1	Appdata	DVD to AV 1 - has set top box too
201125	set top box 1	delete	Profile	This is the dvd - which is included in the set top box
201125	cooker 1	Unknown kitchen	Profile	Only uses 3-15W, probably a gas oven
201128	cooker 2	extractor hood 1	Appdata	Cooker is an extractor hood
201129	dvd 1	Remove spikes	Profile	DVD 2 readings above 2kW on day 306, apart from that used very occasionally up to 6W
201132	cooker 1	delete	Profile	Not used at all - probably an extractor
201146	cooker 1	Unknown	Profile	Max 6W
201148	tv 3	dvd+vcr 1	Appdata	TV 3 to DVD + VCR - this is a VCR but the house has 2

				already
201189	tv 2	Unknown	Profile	TV with just one spike to 8W
201193	vaccum cleaner 1	Unknown	Profile	Vacuum cleaner running 20W continuous all day
201194	computer site 1	light 3	Appdata	Computer site to lamp
201194	computer site 2	light 1	Appdata	Computer site to lamp
201194	computer site 3	light 2	Appdata	Computer site to lamp
201194	computer site 4	light 4	Appdata	This computer site is a light
201206	light 2	other 1	Appdata	light to other - something to do with babies bottles
201206	tv 2	tv+dvd+set top box 1	Appdata	TV to TV + DVD + set top box
201206	tv 1	tv+settopbox 1	Appdata	TV to TV + set top box
201270	vaccum cleaner 1	Unknown	Profile	Vacuum cleaner cycles around 0.6 W max 6W
201276	light 1	Unknown	Profile	A light at 1.5 kW!
201276	tv 1	Remove spikes	Profile	Mostly about 60W in use but spike to 6kW
201282	tumble dryer 1	Spin Dryer	Appdata	New code for spin dryer
201290	light 3	other 1	Appdata	Light shares with cordless phone and radio
201297	cooker 2	Slow cooker	Appdata	New code for a slow cooker
201297	tumble dryer 1	Unknown kitchen	Profile	tumble dryer cycles all day.Shares with fridge?
201311	other 1	hair straightener 1	Appdata	Is a hair straightener
201325	cooker 2	extractor hood 1	Appdata	Cooker is extractor
201327	electric blanket 1	other 1	Appdata	2 electric blankets and also a lamp
201328	tv 3	microwave 2	Appdata	TV is a microwave
201387	cooker 1	Unknown kitchen	Profile	Cooker can't be max 30W - probably an extractor
201389	printer 1	washing machine 1	Appdata	Printer is a washing machine
201488	set top box 3	Unknown	Profile	There is more than just a set top box on here
202113	cooker 1	extractor hood 1	Appdata	Cooker is an extractor hood
202115	desktop 2	computer site 1	Appdata	printer and monitor as well as desktop
202127	cooker 1	iron 1	Appdata	Cooker is an iron
202136	cooker 1	food steamer 1	Appdata	Cooker is a steamer
202155	modem 1	Unknown	Profile	there is more than just a modem on here - up to 250W
202156	light 3	delete	Appdata	This is a hairdryer but there is an identical hairdryer profile already
202170	kettle 1	Unknown	Profile	kettle on for around 1 hour at 60W at midnight sometimes
202176	set top box 1	tv+settopbox 1	Appdata	TV/DVD shares with set top box
202213	cooker 1	delete	Profile	Not used at all - probably an extractor
202217	washing machine 1	washing/drying machine 2	Appdata	This is a second washer dryer - allocated a new code
202246	cooker 1	Unknown	Profile	Cooker used almost

				continuously, mostly around 150W
202314	ps2 1	delete	Profile	ps2 with nothing but one spike of 6 kW
202315	chest freezer 1	Unknown	Profile	off for hours each day and doesn't cycle like a freezer
202351	extractor hood 1	modem 1	Appdata	Extractor is a modem
202351	kettle 1	extractor hood 1	Appdata	Kettle is an extractor
202374	light 3	other 1	Appdata	There are different appliances on this profile but all hardly used and never more than 3 watts
202375	cooker 1	extractor hood 1	Appdata	Cooker is an extractor hood
202375	upright freezer 1	Unknown	Profile	off 5-10am and doesn't cycle like a freezer
202403	light distribution 1	sockets 2	Profile	Light distribution circuits don't even take 10 kW
202403	shower 1	Unknown	Profile	Shower has frequent extended use at just 80W
202434	radio 1	audiovisual site 2	Appdata	Radio as audio visual 2 - other appliances share
202448	audiovisual site 2	Unknown	Profile	av site shows nothing but frequent short spikes to 800W
202448	hair straightener 1	other 1	Appdata	Hair dryer and straightener share with ipod
202448	kettle 1	washing machine 1	Appdata	Kettle is a washing machine
203106	sockets 1	computer site 1	Appdata	Sockets is a computer site
203106	tumble dryer 1	Unknown kitchen	Profile	tumble dryer looks like a fridge
203107	tv 1	Unknown	Profile	TV but looks like a fridge
203108	cooker 1	extractor hood 1	Appdata	Cooker is an extractor hood
203109	cooker 1	extractor hood 1	Appdata	Cooker is an extractor hood
203164	cooker 1	Slow cooker	Appdata	New code for slow cooker
203164	tumble dryer 1	Spin Dryer	Appdata	New code for spin dryer
203204	tv lcd 2	dvd 1	Appdata	tv is actually a dvd - this and the next effectively swap
203211	desktop 2	computer site 1	Appdata	dvd is a TV
203229	cooker 1	delete	Profile	Not used at all - probably an extractor
203250	audiovisual site 6	other 3	Appdata	computer and set top box shares with light and paper shredder
203250	audiovisual site 3	other 1	Appdata	fans share with TV/DVD/hifi
203250	av receiver 1	other 2	Appdata	wii shares with AV receiver. Both used a lot.
203252	circulation pump 1	massage bed 1	Appdata	This is a circulation booster pad not a heating system pump
203258	light distribution 1	Unknown	Profile	Light distribution circuit showing just short blips 2-10kW
203258	oven 1	Unknown kitchen	Profile	oven on all day, complex

				pattern
203260	oven 1	Unknown kitchen	Profile	oven looks like a fridge to me
203320	dvd 1	Unknown	Profile	DVD cycles constantly 5-15W
203338	alarm 1	Unknown	Profile	An alarm showing variable power and up to 400W
203361	vaccum cleaner 1	Unknown	Profile	Vacuum cleaner running long hours up to 11 hours continuous
203385	cooker 1	delete	Profile	Not used at all - probably an extractor
203402	cooker 1	delete	Profile	Not used at all - probably an extractor
203406	shower 1	Unknown	Profile	Shower at only 25W
203407	tv+dvd 1	hair dryer 1	Appdata	TV is actually a hair straightener and hair dryer
203408	speakers 1	laptop 2	Appdata	Speakers is a laptop
203408	cooker 1	delete	Profile	Not used at all - probably an extractor
203422	cooker 1	delete	Profile	Not used at all - probably an extractor
203423	tv lcd 1	Unknown	Appdata	Appdata calls it a vacuum cleaner but it doesn't look like that either
203423	kettle 1	tv lcd 2	Appdata	Kettle is actually a TV
203426	shower 1	Unknown	Profile	Shower at max 40W
203427	microwave 1	light 4	Appdata	Microwave is a light
203442	washing machine 1	Unknown	Profile	washing machine used all day, peaks overnight.
203444	cooker 1	Slow cooker	Appdata	New code for slow cooker
203445	laptop 1	Unknown	Profile	laptop using up to 2kW
203446	kettle 1	Unknown	Profile	kettle max 0.6W. used as water filter?
203493	water heater 1	circulation pump 1	Appdata	Water heater is a gas boiler with circulation pump