Providing Electricity for the future

Participants:
Ángela Díaz López
Alba García González
Izaskun Oregui Bravo
Joseph C. Foster
Maksim Zhuravlev

Instructor:
Tamsin E. Lee

Index

1. Introduction

Electricity demand is rapidly evolving, but the infrastructure delivering our electricity is not. Distribution Network Operators have traditionally been focused on large consumers. Moreover, electricity use on the low voltage network (households and small-to-medium enterprises, SMEs) is changing, as more and more devices require electricity in some way. These all contribute to stresses upon an infrastructure, which is decades behind the technology it supplies.

Although each SME consumes considerably less electricity than large customers, in total their consumption must be taken into account for future predictions. Furthermore, little research has been done into their electricity use, despite the fact that they consume much more than domestic households.

Smart meters measure the electricity consumption for an enterprise, but their data is frequently unavailable or/and expensive for Distribution Network Operators. Because of this reason, the aim of this project is to predict the electricity profile for SMEs without a smart meter.

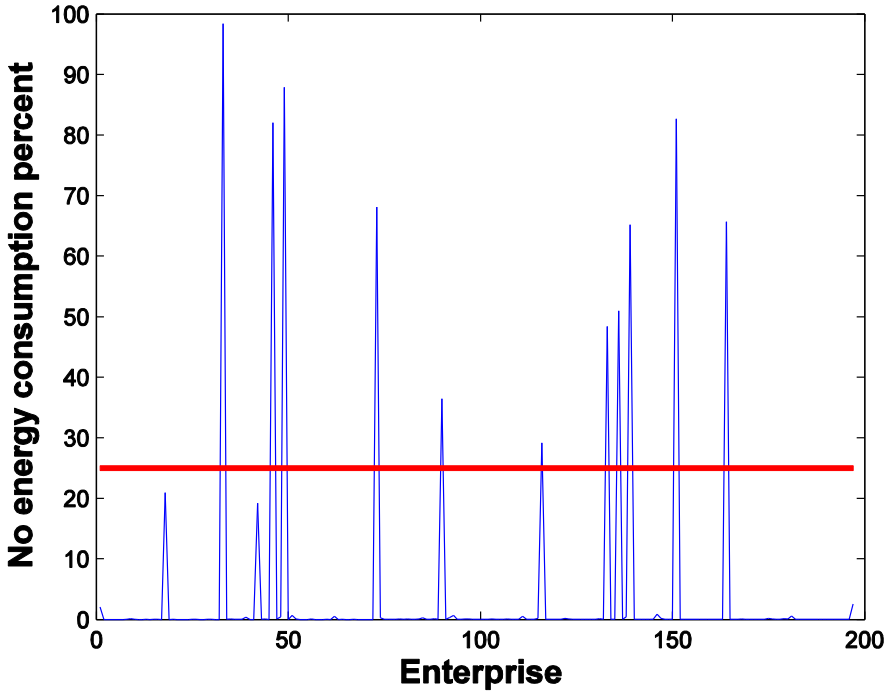
In order to make these predictions, we have half hour readings of electricity consumption for 197 Irish SME. We also have the answers of a survey with 92 questions for the majority of these enterprises, which provides additional information about customers.

The strategy we have followed to resolve this problem is:

- Analyse the smart meters data in order to identify faulty entries.
- Identify the open and closed hours for each SME: daily, half hourly and for weeks with a Bank Holiday.
- Group the enterprises by their energy consumption using the survey.
- Find the energy consumption rate for weeks with a public holiday.
- Build the profiles of energy usage for different time steps (e.g. seasonal, regular week, etc. predictions) using the smart meters data or the survey answers.

2. Data pre-processing

As a first step, we have identified which smart meters to drop from the study because of a large number of missing data of electricity consumption. We have decided to delete customers with more than 25% of no energy consumption entries. As we can see in the figure below, most of the enterprises have no faulty entries and if they have missing data, the majority of them have more than the 25%. As a consequence, we have taken this percentage as a threshold.

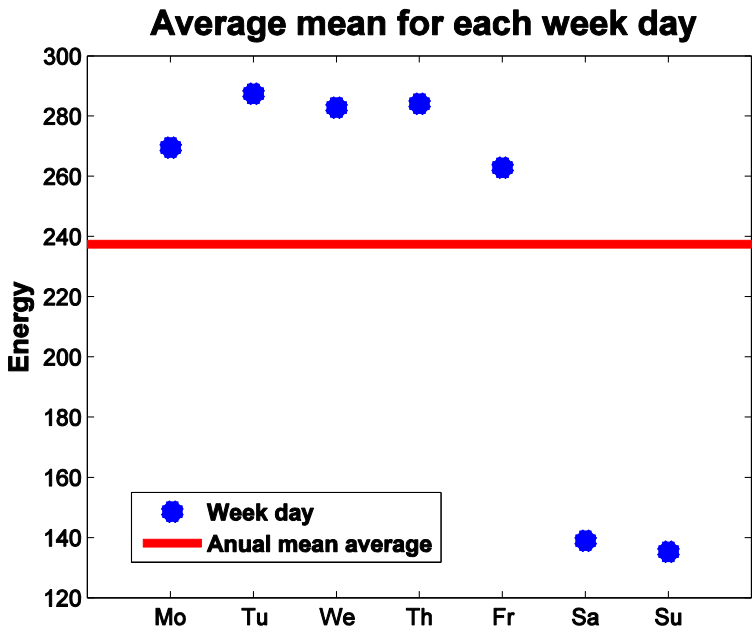


With this criterion we have removed from our study a total of 11 enterprises.

3. Open and closed hours

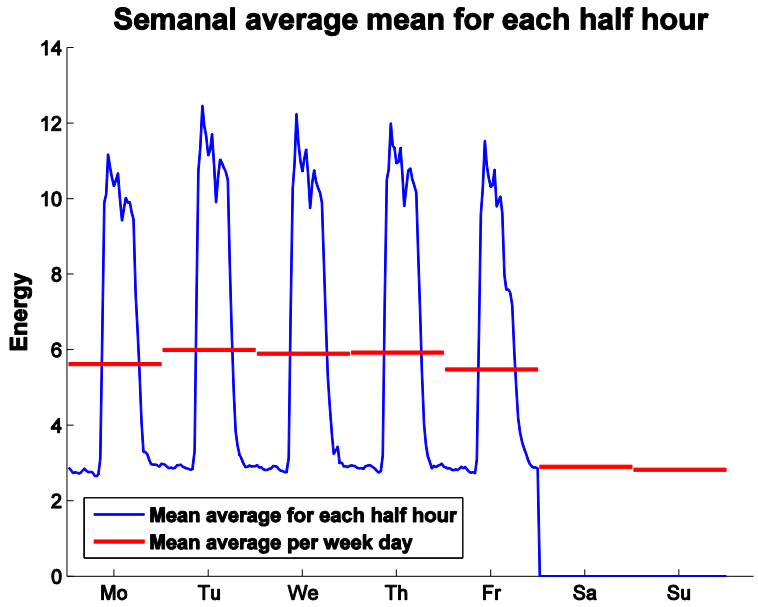
3.1. Estimation for weekdays

For every enterprise, we have compared the mean energy usage for each weekday with the overall daily mean for a year. If the mean on a particular weekday is over the overall daily mean, we set the day as open otherwise we set it closed. In the following figure, we can see what we have just explained for a random customer. The red line represents the annual mean and the blue dots the daily average.



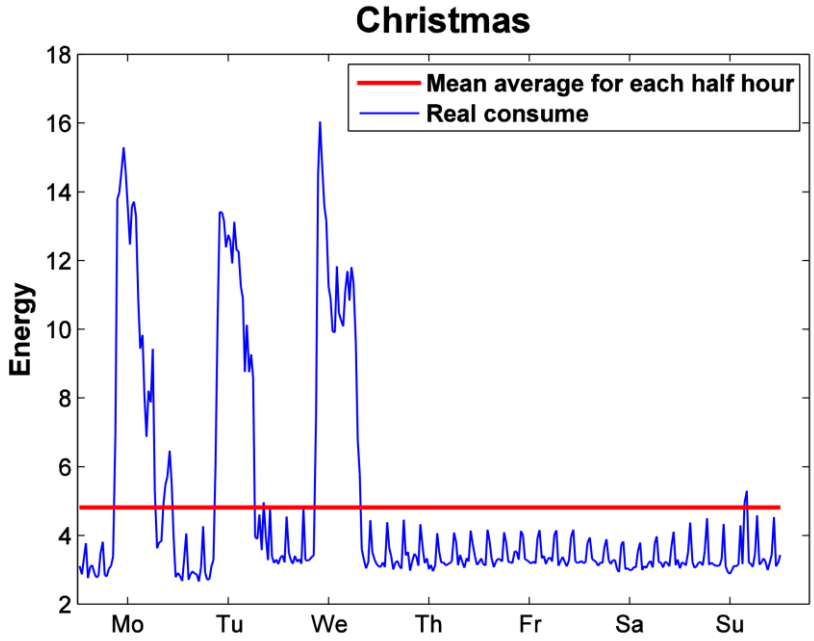
3.2. Estimation for half hours

Following the same idea as the previous point, we have computed the open and closed hours. We have compared the mean energy usage for half hour with the mean for each weekday. To avoid having open hours in closed days, we have set as zero the average of half hour for the closed day. This is the mean we are using as a threshold being lowered by low usage on closed days.
In the following figure we show an example of a random enterprise. So, if the blue line is over the red line we will determine the enterprise is open in that half hour; similarly, we will the half hour as closed if the blue line is under the red.



3.3. Estimation for weeks with a Bank Holiday

For each week with a Bank Holiday, we have compared the real consumption with the mean of energy usage for this week. We proceed as in the previous sections setting open hours as those half hours in which real consumption is over the mean average, and closed hours as those under the mean.



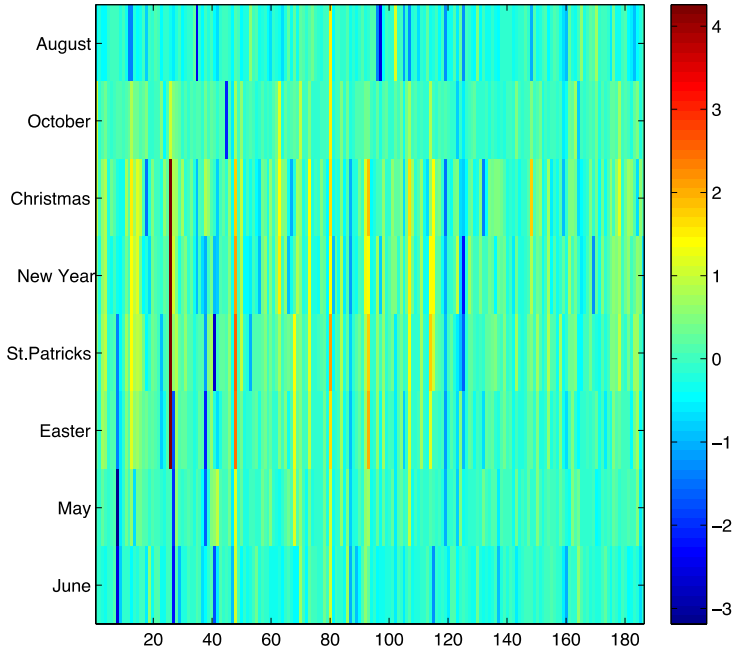
4. Relationship between a regular week and a week with a bank holiday

As a previous analysis, we want to know how holidays have a real effect on electricity consumption. To measure this, we have calculated the next rate for each enterprise i:

$$rate_i = \log\left(\frac{(energy\ average\ on\ a\ week\ with\ a\ bank\ holiday)_i}{(energy\ average\ on\ a\ normal\ week)_i}\right)$$

So, if $rate_i > 1$, the customer i works more on a week with a bank holiday than on a normal week, but if $rate_i < -1$, it works more on a normal week. Otherwise, the customer works similarly to a normal week.

In the following figure, this rate is shown for each enterprise, for each week with a public holiday. Warm colours represent higher energy used on a week with a bank holiday and cold colours represent smaller energy used on a week with a bank holiday.



Analysing the figure in detail, on the one hand we checked (with the survey) that the enterprises in warm colours concur with shops. On the other hand, we can observe that summer holidays don't affect the consumption considerably. However, winter holidays (Christmas, New Year, St. Patricks) do. In conclusion, we think we might have a seasonal effect.

5. Method to compute the profile

As our goal is to estimate the profile without a smart meter, we must calculate the open and closed average knowing the real daily average energy consumption. To do it, we fit a linear regression $U_{open} = ax + b$, where U_{open} is the estimation of open average and x the real daily usage. Once we have estimated the open average, we compute the estimation of closed average with

$$U_{closed} = \frac{E_{total} - E_{open}}{N_{closed}} \qquad (1)$$

where E_{total} is the total energy consumption of a year, E_{open} the open energy consumption and N_{closed} the total number of closed hours over a year.

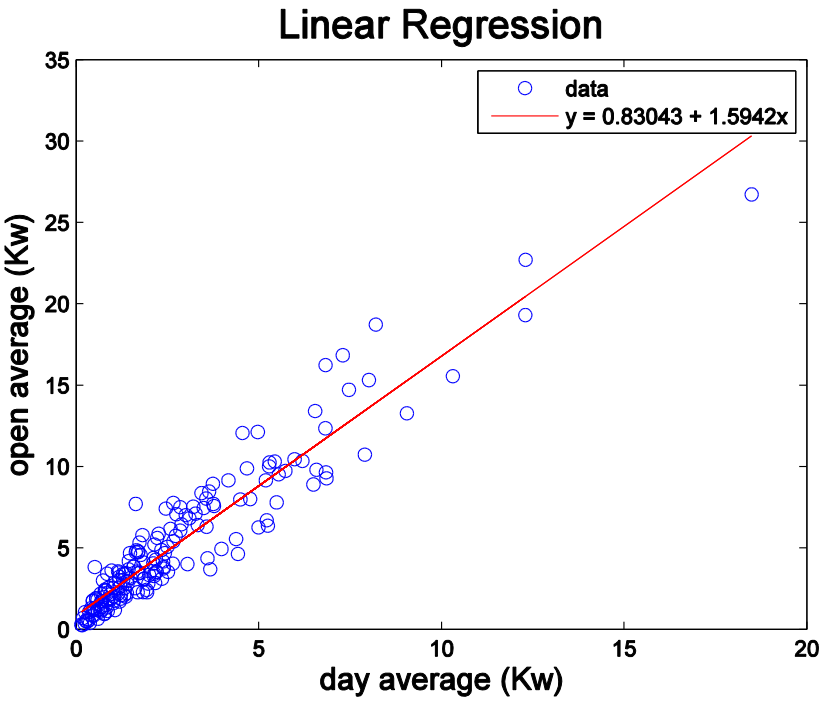
Knowing the estimation of open and closed consumption, U_{open}, U_{closed}, and the schedule of each enterprise, we can build the profile.

6. Energy consumption profiles

6.1. Seasonal estimation

As mentioned in section 4, we begin by estimating the electricity usage profile seasonally (winter, spring, summer and autumn). Splitting the data in 4 seasons we have fitted a linear regression -to each season- following the method explained in the previous section.

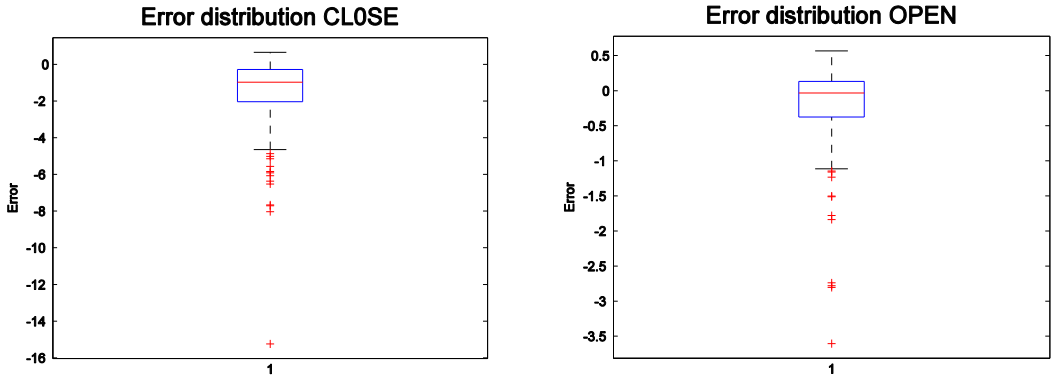
In the next figure we show the winter example, where it can be observed the average daily use vs. the average used on open days in blue dots and the fitted line in red. In this case, $U_{open} = 0.83043 + 1.5942x$ and then we can compute the estimation U_{closed} using equation (1).



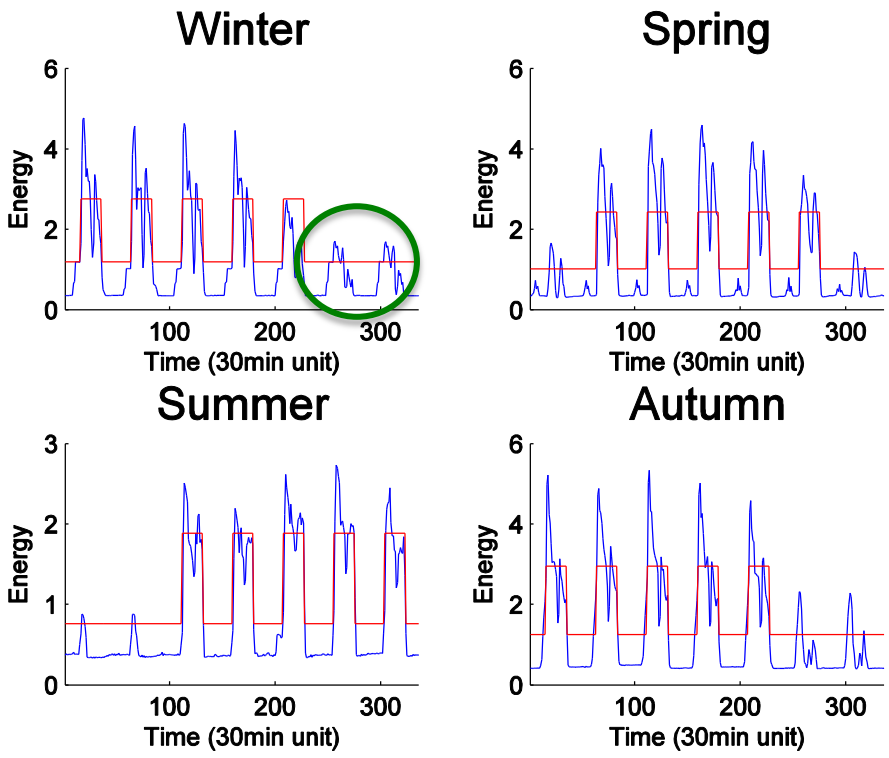
To check the goodness of our estimations we calculate the relative error for both, open and closed hours:

$$ERROR = \frac{U - \hat{U}}{U},$$

where U is the real (open or closed) value and Û the estimated value. In the following figures we show the boxplot for open and closed hours. We can observe many outliers when we try to estimate closed hours average whereas, for open hours the estimation is better. In spite of being the closed estimation worse than open estimation, it is not really important since open hour estimation is much more important.



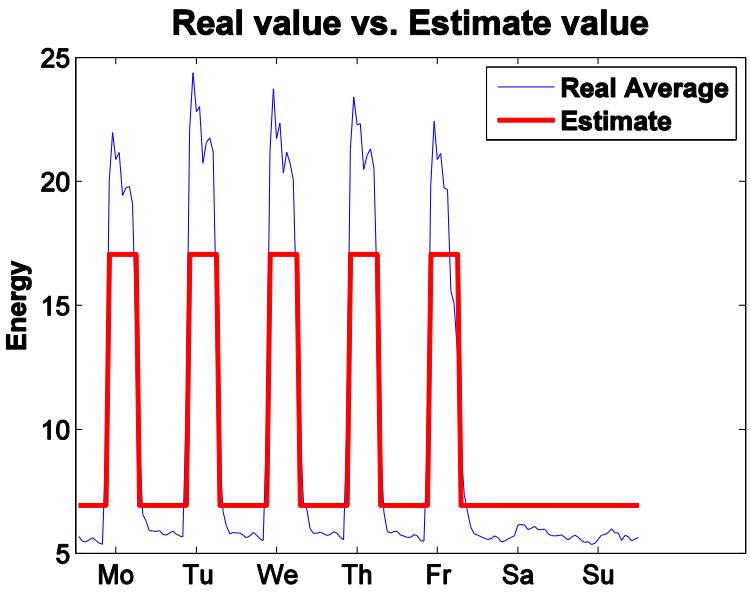
Finally, we plot the estimated profiles for a week of each season. The next figure shows the estimation (in red) together with the real data (blue) for a random customer.



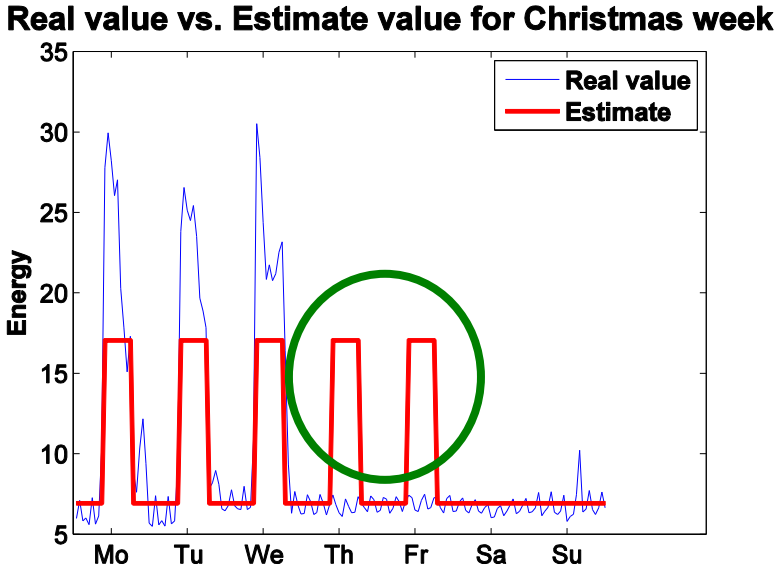
If we focus in the winter profile we can see that the closed hours are overestimated because the weekend timetable is not accurate as emphasised in the green circle. The underestimation of the open hours is also caused for the same reason. In the next step we should try to reduce the error by estimating the open and closed average usage using other time steps.

6.2. Hourly estimation

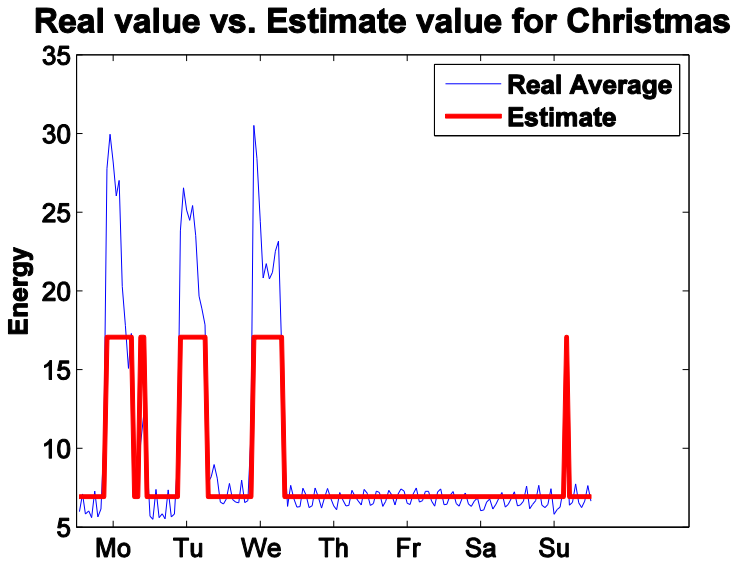
Seeing the seasonal estimation is not as good as we expected, we have tried to improve it computing the same open and closed hours estimation for the whole year. The profile for the same customer as before is shown in the next picture.



It can be observed the estimation has improved in general. However, the estimation is the same everyday but there are days with more consumption than others. So, we will analyse each weekday separately.
Moreover, if we plot the estimation together with the real value for a week with a bank holiday, we can observe that the estimation fails. In the following figure we show an example for the Christmas holiday week, in which the bank holiday is on Friday.



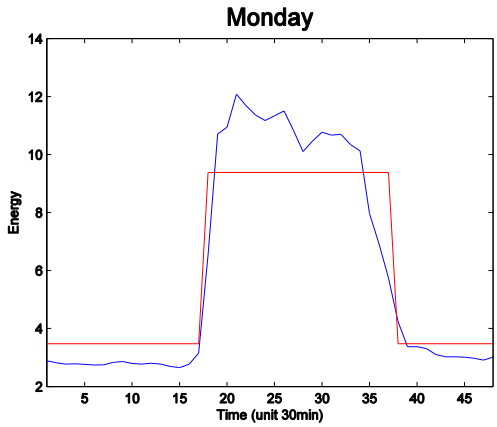
To improve this estimation, we have used the timetable specially computed for weeks with a bank holiday (section 3.3). The same example with its new estimation can be seen in the next picture.

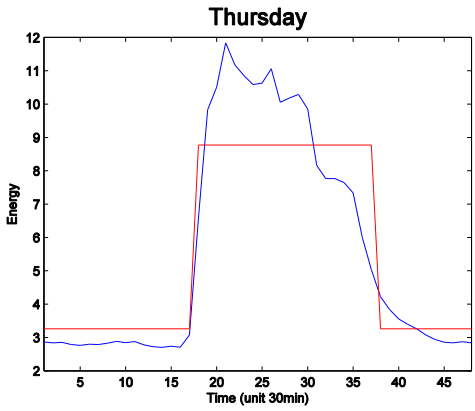


6.3. Weekday estimation

In this section we are going to solve the problem we mentioned at the beginning of the previous section. Now, following the same method we used for previous estimations, we are going to compute the profile for each weekday, including Saturdays and Sundays.

For instance, showing the estimation for Mondays and Thursdays, we can see that the profile is closer to the real value than in previous estimations. Because the real profile behaviour for Monday to Friday is quite similar, we only show these examples. Saturday and Sunday profiles, however, are unpredictable.

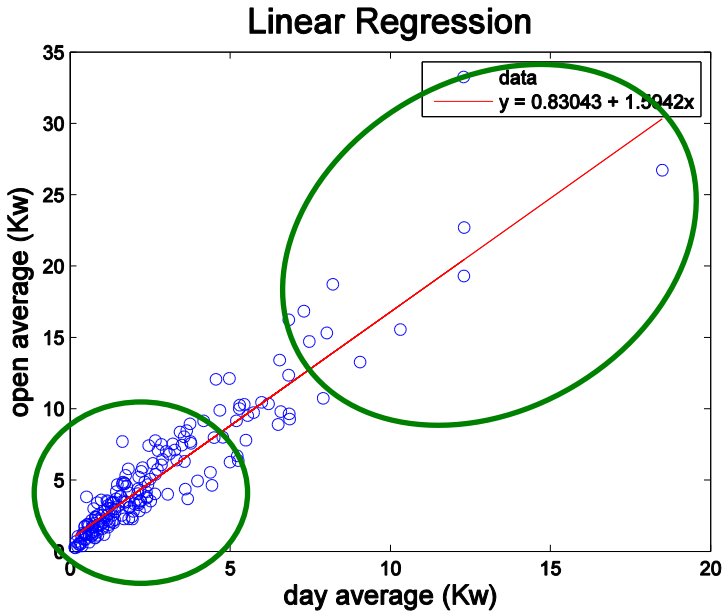




Although we have tried to improve the profile, we continue having overestimation for open hours and underestimation for closed hours (for weekdays) and the unpredictability for weekends. However, in a real world scene we would have the exact schedule for any enterprise. So, the results will improve considerably.

7. Predicting the electricity usage grouping enterprises

The linear regression treats big energy customers and small energy customers the same, as we can see in the picture.



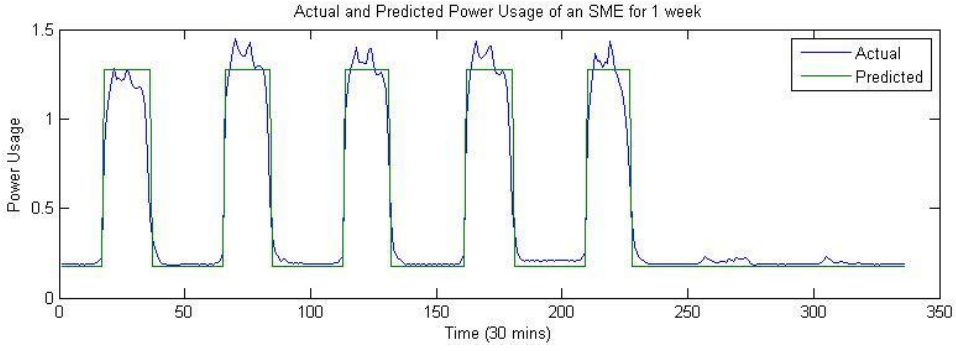
We are going to divide the businesses into smaller groups and compare like with like to improve the estimations, based on survey responses.

To group the enterprises, we are going to follow the next rules:

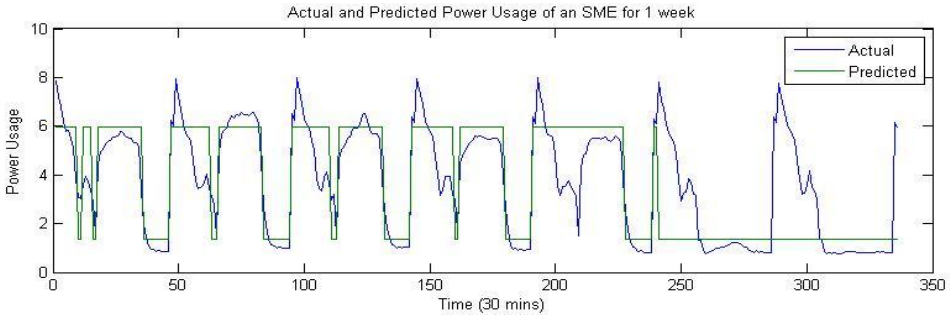
- We analyse the answers of the survey, deleting the unanswered or incorrectly answered (for example, the answer '999' when asked for a percentage).
- We stripped the survey down to only questions that had been answered by all SMEs (and appropriately).

The final database has 80 SMEs and a survey of 4 questions.

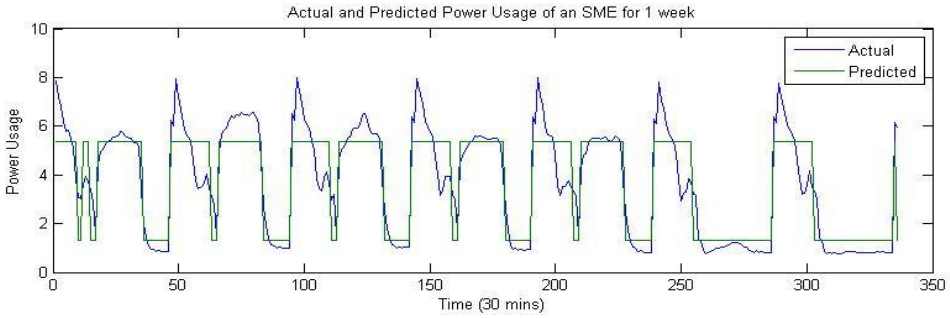
For one SME, we create a profile using the average of the 4 most similar SMEs, here is the result.



We can see the estimation is really good; but if we choose another group in which the enterprises are not so similar, the profile get worse as we can observe in the next figure.



Moreover, as we said previously, we would know the exact opening hours of the SME. For instance, in the picture above we have modelled this SME is closed all weekend, when in fact it is clear that they are open. If we amend our open and closed hours data (manually in our case) we see a much better approximation (improving the error from 9.4% to 3.3%), as we can see here.



To improve our estimates, we could include more relevant questions in the survey allowing us to recognise similar businesses with more accuracy. Additionally, peak times could be taken into considering on.

8. Conclusions

- We have been able to make an accurate profile for an SME without a smart meter.
- Monday to Friday estimations have been easy.
- Seasonality has little effect on energy usage (very different to domestic).
- Grouping customers improve the results a lot.

9. Further work

- Predicting when maximum usage occurs and how many kw for business without smart meters.
- Test the methods with accurate open and closed hours.