Modelling the weekly electricity consumption of small to medium enterprises

T. E. Lee[a], S. A. Haben[1], P. Grindrod[1]

[a]*The Mathematical Institute, University of Oxford, Andrew Wiles Building, Woodstock Road, OX2 6GG, UK.*

Abstract

Electricity consumption will evolve as more low carbon technology is adopted. To continue supplying electricity reliably into the future, Distribution Network Operators (DNOs) need to better understand low voltage customers. This can be gained from smart meter data, but DNOs may have limited access. We demonstrate a method to create an energy demand profile for a small to medium enterprise (SME) without a smart meter based only on operational hours, mean daily use (from quarterly readings), and smart meter data from other SMEs. We cluster the smart meter data using a simple Gaussian mixture model, thereby providing five groups of customers easily identified by their mean daily use. A new meter is assigned a cluster and an electricity profile created. For comparison, a profile is also created when the smart meter data is not clustered. The average difference between the actual and predicted operational/non-operational power is less than 0.2kWh, and clustering reduces the range around this difference. The methods presented here out perform the flat profile (akin to current methods). Additionally, we found no relationship between electricity consumption and non-energy traits such as type of business and number of employees; and SMEs electricity consumption is not significantly affected by weather.

Keywords: Low voltage, Smart meters, Clustering

1. Introduction

Small to medium enterprises (SMEs) contribute a large proportion of the total energy demand in the UK but are often overlooked in policy and research [3]. Therefore it is essential that their demand is accurately modelled to assist distribution network operators (DNOs) manage and plan the network. Such concerns are more immediate with the increase of low carbon technologies, such as electric vehicles and photovoltaics (PV), which will impact the low voltage (LV) network [9]. Currently DNOs use After Diversity Maximum Demand [6] to model maximum demand for SMEs. This approach does not account for time-of-day information, and is therefore becoming increasingly insufficient. For example, as more PV are installed, DNOs require the time of low demand.

The UK is in the preliminary stages of rolling-out smart meters, which measures energy demand of millions of customers every 30 minutes. In the UK, smart meter data is proprietary, and therefore possibly unavailable/expensive to DNOs. Nonetheless, they do have access to quarterly readings. As such, we ask if a DNO can use quarterly readings and publicly available information to estimate an SMEs electricity profile, which is regular and predictable (compared to domestic customers).

In this paper we develop our method using preprocessed smart meter data from the Irish smart meter trial [5]. We consider a years worth of smart meter data for 196 SMEs starting from midnight 14th July 2009. For each meter, we determine the average operational power and average non-operational power. Based on these attributes, we cluster the smart meter data using a simple Gaussian mixture model, thereby providing five groups of customers easily identified by their mean daily use.

From average values within a cluster, one can accurately estimate customers weekly energy demand profile using only their operating times (potentially public information) and their mean daily usage. Nonetheless, clustering is not necessary to create the profile, and indeed we compared clustering with not-clustering.

The advantage of the method presented in this report is two fold. Firstly, very little information is required to produce the estimates. Secondly, any new customers can be assigned to the current clusters, and therefore can be modelled knowing their mean daily demand and operational hours only. All the same, ideally non-energy attributes could be linked to energy consumption behaviour, thereby reducing the need for energy data.

Along with the smart meter data, there is a survey completed for 138 of the SMEs. This contains replies for a number of questions including, type of business, the number of employees, the age of the building, what weekend days are operational, etc. Surprisingly, we find no relationship between the survey responses and the clusters. This shows that energy information is necessary for estimating a SMEs profile, which is also the case for domestic customers []. However, unlike domestic customers [7], we find SMEs electricity consumption is not significantly affected by the weather.

We begin by describing how we determine operational and non-operational times from the smart meter data. In Section 3 we describe how we cluster the customers, and look for relationships between our five clusters and survey responses. In Section 4 we compare our estimates based on clustering and not clustering. We also investigate the effect of seasonality, and find it to be negligible, see Section 5. Finally, we summarise in Section 6.

2. Identifying operational hours

First we determine which days the business is open, and then, considering these days only, the hours that the business is open. We use a data driven approach, however this information may be publicly available.

2.1. Operational days

We first use the smart meter data to determine operational days. To do this we state that if the median daily usage for a particular day of the week $\tilde{x}_m^{DAY}$ is less than a chosen quantile of the overall daily use, q_m, we assume the business is closed on that day (for meter m). For example, suppose $\tilde{x}_m^{SUN} < q_m$, then we assume the business is closed on Sundays. To determine the actual quantile, we consider quantiles for probabilities from 0 to 1 (intervals of 0.01). For each quantile choice, the 'closed predictions' for Saturday and Sunday is compared to a survey response in which 138 out of the 196 customers stated which days the business is open on weekends. We use a common comparison measure, the F-score,

$$F = \frac{2PR}{P+R} \quad \text{where} \quad P = \frac{tp}{tp+fp} \quad \text{and} \quad R = \frac{tp}{tp+fn} \tag{1}$$

where tp is the true positive amount, fp is the false positive amount and fn is the false negative amount [8]. The 'closed predictions' and survey are in close agreement when the F-score is small, see Figure 1. The F-score is at a minimum at 0.13, however this is because both the true positive and false positive for 'Closed Sat & Sun' (black) is zero, see Figure 2. Therefore, the closest match to the survey responses is from the 0.4 quantile. Notice that even under this quantile choice the survey responses and predictions do not completely agree. This difference may be because the business is open, but the daily electricity consumption is lower than a typical open day, perhaps due to half days or less staff/machinery.

Figure 1: Using the F-score to compare the weekend 'closed predictions' with the survey responses. The predictions were calculated using cut off quantile choices between 0 and 1. The quantile value corresponding to a small F-score indicates a strong agreement between the prediction and survey.

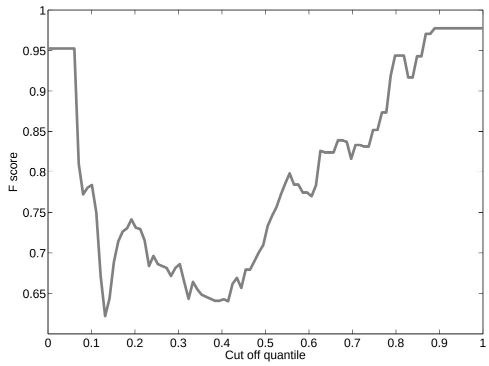
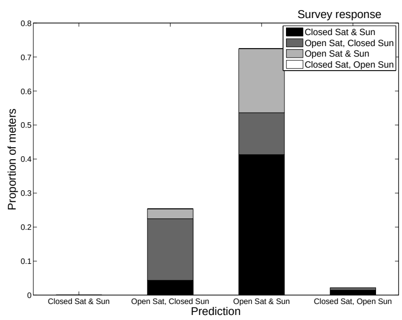
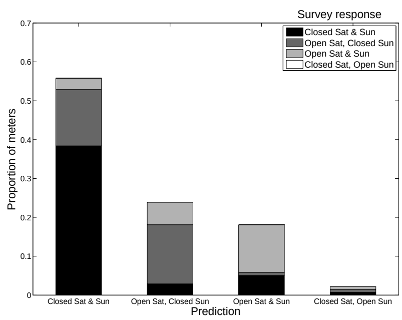


Figure 2: Comparing the weekend 'closed predictions' with survey responses. Prediction and survey responses would be in complete agreement if the first bar was completely black, the second dark grey, the third light grey, and the fourth empty

(a) Cut off quantile at 0.13.



(b) Cut off quantile at 0.4.



2.2. Operational hours

To determine whether a business is operational in a half hour we first remove ‘closed’ days (see previous subsection). Using the remaining ‘open’ days, the average power is calculated and compared with the average power for each half hour during the ‘open’ days. For example, if a meter is closed weekends, the average power for Monday to Friday is calculated and compared with the average power for each half hour during Monday to Friday. If the power for a half hour is larger than the power for the ‘open’ days, we consider the business to be operational during this half hour.

3. Clustering

The simplest choice for attributes to cluster is the daily mean and half hourly standard deviation. However, this did not lead to clusters which were easily differentiated by the daily mean alone, nor the survey responses. Alternatively we considered the normalised difference between day and night use (where the day is defined as 9am–6pm), the normalised difference between weekday and weekend use, and the half hourly standard deviation. But again, attributes did not lead to clusters easily defined by the daily mean, nor survey responses. We present the results from clustering based on operational power, non-operational power and half hourly standard deviation. This choice led to clusters being distinctly separated according to the daily mean.

We model our three attributes as a finite mixture model (FMM) of uncorrelated Gaussian distributions [4]. The parameters of the mixtures and the mixing proportions are found through an implementation of the Expectation-Maximisation algorithm, which finds the parameters that maximise the likelihood function of the model [4]. Traditionally the k-means algorithm has been used for clustering in power systems. However, this is simply a more restrictive version of the FMM [4]. We use the Matlab function `gmdistribution` to implement the algorithm.

As with many clustering algorithms, a disadvantage of the method is that the optimal number of clusters must be defined prior to implementing the clustering algorithm. There are no definitive ways of choosing the number of clusters, but there are some indicators and metrics that can be used to help inform the decision. We consider the Bayesian Information Criterion (BIC) which penalises the maximum likelihood function by the number of parameters that are used in the model [4]. In particular, the BIC penalises larger numbers of clusters. We use the BIC to choose a cluster size which obtains a good model of the observations without an excessive number of parameters. As shown in Figure 3, five clusters is a reasonable choice.

To identify whether a new meter can be allocated a cluster based on non-energy data, we compare the clusters to survey responses. Ideally there would be a distinct relationship; that is a large proportion of one colour would be present in only one bar in Figure 5, e.g., an agriculture SME belongs in Cluster 2. However, overwhelmingly, there is no relationship between the survey responses and clusters. This supports previous research which also found little correlation between energy consumption and household type [7].

Instead of non-energy data we find a distinct relationship between clusters and daily mean μ_m. When plotting clusters in terms of their usage and normalised standard deviation (Figure 4), the clusters present clear partitions based on daily mean. Therefore, an SME without smart meter data can be allocated to a cluster using the quarterly readings alone.

4. Predicting electricity use

For each meter m, $m = 1, 2, \ldots, 196$, we assume that only the daily mean μ_m and operational hours are known. We estimate the operational power $e_{m,O}$ and non-operational power $e_{m,N}$ for meter m using its daily mean μ_m and either (i) the mean operational power and mean non-operational power of the remaining 195 meters, or (ii) the mean operational power and mean

Figure 3: Clustering when the attributes are average operational power, average non-operational power, and normalised standard deviation. The BIC indicates five clusters is most suitable.

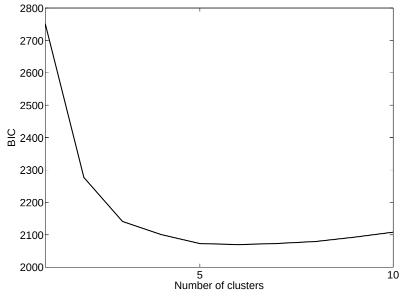
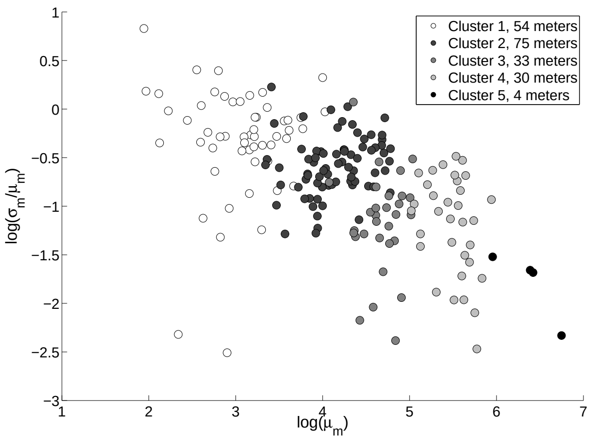


Figure 4: The five clusters arranged by their daily mean μ_m and normalised standard deviation σ_m/μ_m. Clusters are numbered by daily mean μ_m: Cluster 1 has the lowest daily mean and Cluster 5 has the highest daily mean. (Domestic households have upper bounds of $\log(\sigma_m/\mu_m) \approx 0.65$ and $\log(\mu_m) \approx 1$, see [2].)



non-operational power of the meters remaining in the cluster for meter m. By simply applying the estimated average operational power and non-operational power to the operational hours, one may create a coarse electricity profile.

We denote these initial estimates for $e_{m,O}$, $e_{m,N}$ by $\hat{e}_{m,O}$ and $\hat{e}_{m,N}$. From the initial estimates, and the operational hours, the total predicted energy during an average week, T_m, for meter m, is

$$T_m = H_{m,O}\hat{e}_{m,O} + H_{m,N}\hat{e}_{m,N}, \tag{2}$$

where $H_{m,O}$ is the number of operational hours and $H_{m,N}$ is the number of non-operational hours in a week. Ideally, the estimate (2) matches the weekly energy use for the meter, $T_m = 7\mu_m$. Therefore, we adjust the profile by setting

$$e_{m,O} = \frac{7\mu_m}{T_m}\hat{e}_{m,O} \quad \text{and} \quad e_{m,N} = \frac{7\mu_m}{T_m}\hat{e}_{m,N}. \tag{3}$$

Using these adjusted power values for meter m, and the operational hours (see §2), we compose the predicted profile. As expected, the adjustment from $T/7\mu$ is larger for (i) - the data is not clustered (a maximum adjustment of 16.43 compared with a maximum adjustment of 3.98). This confirms that the other members of a cluster have similar weekly usages.

Figure 6 compares this predicted average weekly profile with the actual average weekly profile for a meter using non-clustered data and clustered data. The peak and trough behaviour is captured better when using the data from the cluster (cluster 1, 54 meters) compared to the whole data set (195 meters). This is similar for other customers.

To test the accuracy of these two methods (non-clustered and clustered), we compare them with a flat estimate, which is the daily mean μ_m divided by 48 half hours ($e_{m,O} = e_{m,N} = \mu_m/24$). Two error measures are calculated, one for the operational power and one for the non-operational power,

$$E_{m,O} = \bar{e}_{m,O} - e_{m,O}, \tag{4}$$

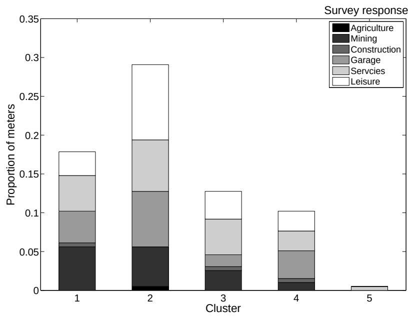
$$E_{m,N} = \bar{e}_{m,N} - e_{m,N}, \tag{5}$$

where $\bar{e}_{m,O}$ is the actual average operational power, and $\bar{e}_{m,N}$ is the actual average non-operational power.

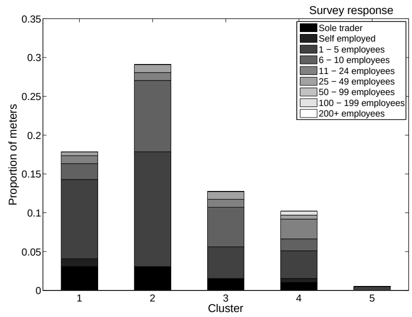
Both clustering and not clustering the data outperforms the flat estimate, see Figures 7. The adjustment stage, equation (3), ensures that the non-clustered and clustered approaches are competitive to each other. Clustering the data first provides greater accuracy (the medians are two orders of magnitude smaller), on the assumption that the cluster corresponding to meter m is correctly identified. This is an fair assumption since the clusters are clearly grouped by their daily mean μ_m, Figure 4.

Figure 5: Breakdown of the five clusters according to responses from survey questions. A relationship would be apparent if all of one colour was in one bar. Full survey questions are in the Appendix.

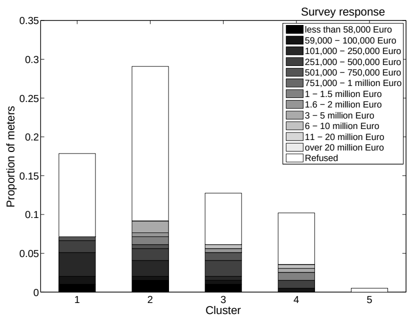
(a) What is the organisation's business?



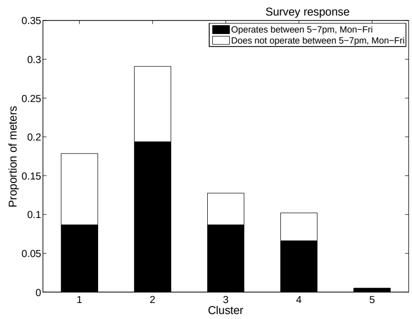
(b) How many employees are there in your business?



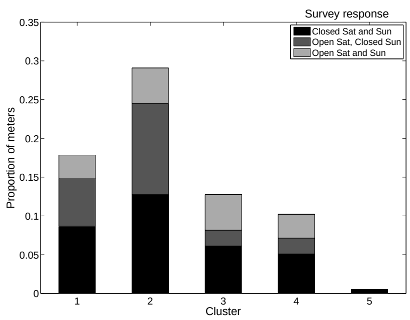
(c) What was your approximate annual turnover during the last financial year?



(d) Do hours of operation include 5pm to 7pm from from Monday to Friday?



(e) Does the premises operate during the weekend?



(f) Approximately how old is your premises?

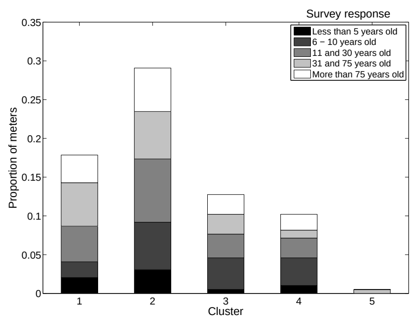
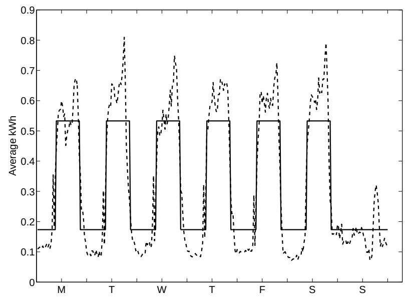


Figure 6: The predicted average week (solid line) and the actual average week for meter $m = 1$. The data was clustered according to average operational power, average non-operational power and normalised standard deviation.

(a) Data not clustered data.



(b) Data clustered.

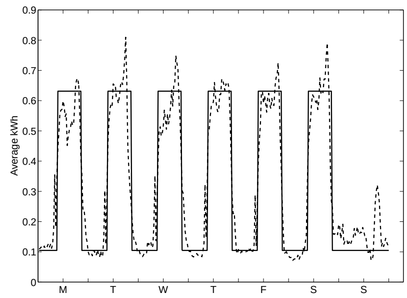
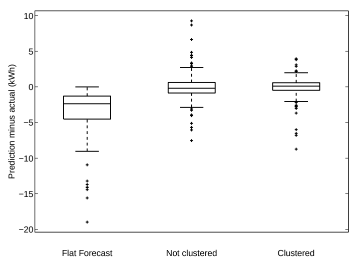
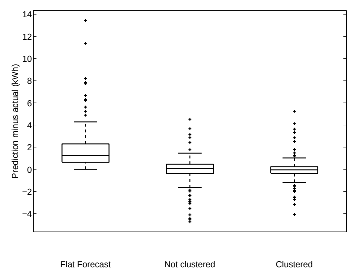


Figure 7: Comparing the error from a flat estimate, our prediction when not clustering the data, and our prediction when clustering the data.

(a) Average operational power. From left to right, the medians are -2.368kWh, -0.196kWh and 0.099kWh.



(b) Average non-operational power. From left to right, the medians are 1.236kWh, 0.084kWh and -0.046kWh.



5. The effect of seasonality

From the half hourly use over the year, aggregated across all meters, seasonal behaviour is apparent, see Figure 8. As such, we investigate the effect of temperature (Celsius), humidity, cloud cover proportion and wind chill factor. Wind chill factor is a measurement which incorporates temperature and wind speed to provide a 'feels like' measurement [1]. The corresponding heat index (which combines temperature and humidity) is not suitable for the Irish data as the temperatures are too low. Since the location of the SMEs within Ireland are unknown we take an average of the weather readings from Cork, Dublin and Shannon.

There are clearly two different types of usage, easily defined by grouping half hours into ones which occur 9am–6pm, Monday to Friday (excluding bank holidays), and others. The electricity consumption during half hours that do not occur 9am–6pm, Monday to Friday are not affected by the weather variables, see Figure 9. In fact only temperature and, by proxy, wind chill factor, during 9am–6pm, Monday to Friday appear to affect electricity consumption. However, the corresponding linear R^2 measure is only 0.4 for both temperature and wind chill factor. Therefore, electricity consumption of SMEs is weakly affected by weather. To check that statement is accurate, we applied our methods to the only 13 weeks (14th July 2009 – 13th October 2009), thereby reducing the effect of seasonality. We avoided the Christmas period in the 13 weeks since it may need to be treated separately, see Figure 8.

When using 13 weeks, the BIC is at a minimum for eight groups, see Figure 10a. However, as a smaller number of groups is preferable, and the BIC for five is only marginally larger, we take five groups.

When using data from the complete year, a clear division between clusters and daily mean μ_m is apparent, see Figure 4. This relationship is not apparent from using 13 weeks of data since the clusters, when arranged by daily mean, appear to overlap considerably, see Figure 10b. (This overlapping is also present when eight groups are used.) Therefore, it would be challenging to allocate a new meter to a cluster based on daily mean alone.

Setting this aside (by assuming meter *m* can be easily allocated a cluster), we used the process in Section 4 to asses whether 13 weeks of data provides a more accurate prediction of operational and non-operational power. We discovered the difference to be negligible, see Figures 7 and 11. The greatest difference was operational power (for clustered data), where the medians differ by only 0.093kWh. Consequently, from this data set, it is recommended that seasonality effects are ignored and data from a full year is used.

Figure 8: The half hourly use aggregated over all meters for a year

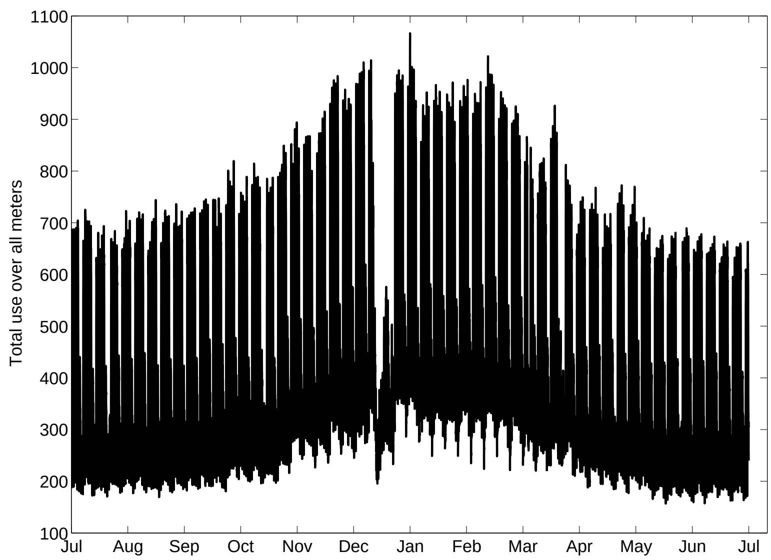
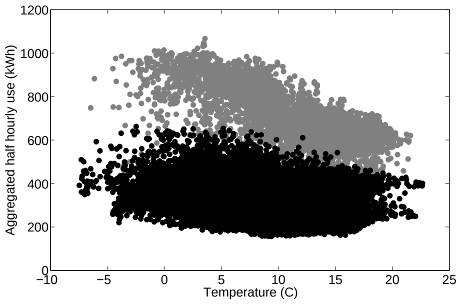
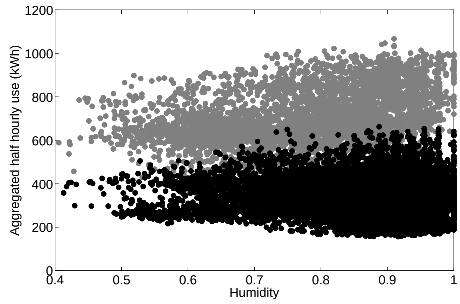


Figure 9: The effect of weather variates. Each point represents the half hour use aggregated over all meters. Half hours which fall between 9am and 6pm, Monday to Friday (excluding bank holidays) are in grey. Other half hours are in black.

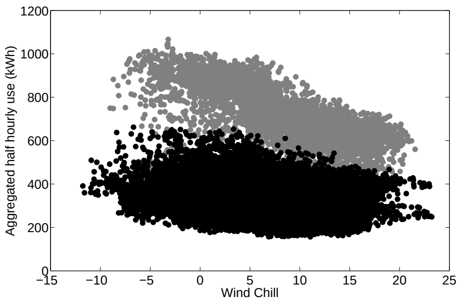
(a) Temperature.



(b) Humidity.



(c) Wind chill factor.



(d) Cloud cover proportion.

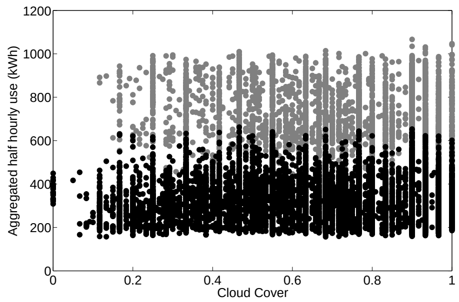


Figure 10: Using only 13 weeks worth of data. Clustering when the attributes are average operational power, average non-operational power, and normalised standard deviation. Clusters are numbered by daily mean μ_m: Cluster 1 has the lowest daily mean and Cluster 5 has the highest daily mean.

(a) The BIC indicates five clusters is suitable.

(b) The five clusters arranged by their daily mean μ_m and normalised standard deviation σ_m/μ_m.

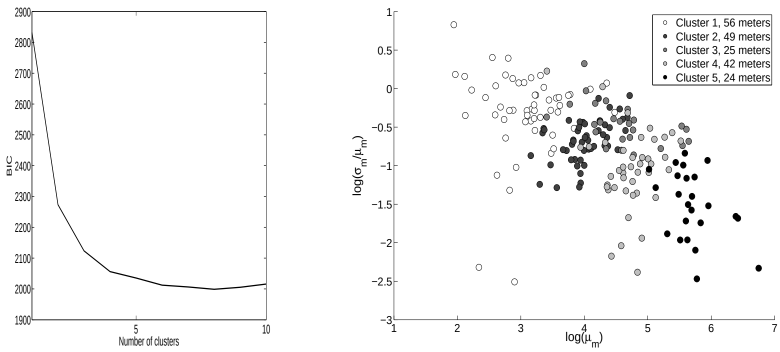
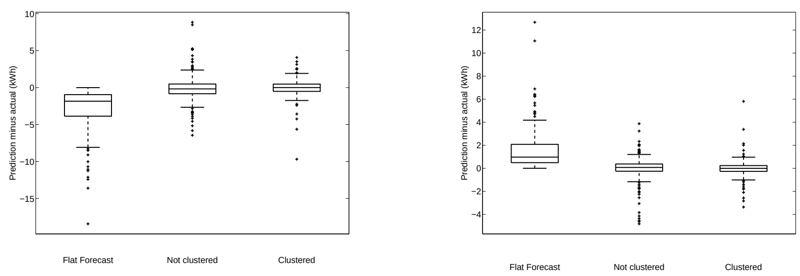


Figure 11: Using only 13 weeks worth of data. Comparing the error from a flat estimate, our prediction when not clustering the data, and our prediction when clustering the data.

(a) Average operational power. From left to right, the medians are -1.843kWh, -0.179kWh and -0.006kWh.

(b) Average non-operational power. From left to right, the medians are 0.973kWh, 0.076kWh and -0.004kWh.



6. Conclusion

With merely the operational hours (publicly available information) and the daily mean (available from quarterly readings), we have created a close approximation of a SME customers weekly energy demand profile. The prediction significantly outperforms the flat estimate, which is akin to current methods. This approximation can be further improved by clustering the data before making a prediction. We used operational power, non-operational power and standard deviation to cluster the meters into five categories. The operational and non-operational power are related to the daily mean. Consequently, customers without smart meters can be readily placed in a cluster when only their quarterly readings are available. This is highly beneficial to DNOs since smart meter data may be unavailable or expensive.

SMEs electricity consumption is not greatly affected by weather. Moreover, taking into account the effect of weather by considering 13 weeks of data (and not a year) produces clusters which are not easily defined by the daily mean. Perhaps a larger data set, with location information, would indicate more sensitivity to weather variables.

Counter-intuitively, there is no apparent correlation between electricity consumption and non-electricity features such as type of business and number of employees. However, this is in line with earlier work, [7].

Acknowledgment

The authors thank SSEPD for support via the New Thames Valley Vision Project (SSET203 New Thames Valley Vision), funded through the Low Carbon Network Fund.

References

[1] Feinberg, E. A., & Genethliou, D. (2005). Load forecasting. In *Applied mathematics for restructured electric power systems*, 269–285. Springer US.

[2] Haben, S. (2013) Categorisation of data: Technical report. *University of Reading.*

[3] Kannan, R., & Boie, W. (2003). Energy management practices in SME: case study of a bakery in Germany. Energy Conversion and Management, 44(6), 945–959.

[4] McLachlan, G., & Peel, D. (2004). *Finite mixture models.* John Wiley & Sons.

[5] McLoughlin, F., Duffy, A., & Conlon, M. (2012). Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. *Energy and Buildings*, **48**, 240–248.

[6] McQueen, D. H., Hyland, P. R., & Watson, S. J. (2004). Monte Carlo simulation of residential electricity demand for forecasting maximum demand on distribution networks. *Power Systems, IEEE Transactions*, **19(3)**, 1685–1689.

[7] Morley, J. & Hazas. M. The significance of difference: Understanding variation in household energy consumption. (2011) *ECEEE Summer study*

[8] Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, **2(1)**, 37–63.

[9] Putrus, G. A., Suwanapingkarl, P., Johnston, D., Bentley, E. C., & Narayana, M. (2009). Impact of electric vehicles on power distribution networks. *Vehicle Power and Propulsion Conference, VPPC'09. IEEE*, 827–831.

Appendix A. Survey questions in full

- What is the organisation's business?

    1. Agriculture, forestry and fishing.

2. Industry mining and quarrying Manufacturing Electricity, gas, steam and air conditioning supply Water supply, sewerage, waste management and re mediation activities.
3. Construction.
4. Wholesale and retail trade; repair of motor vehicles and motorcycles.
5. Business and Professional Services Information and communication Financial and insurance activities Real estate activities Professional, scientific and technical activities Administrative and support service activities Public administration and defence; compulsory social security Education Human health and social work activities.
6. Other Transportation and storage Accommodation and food service activities Leisure hotels Arts, entertainment and recreation Other service activities Activities of households as employers; undifferentiated goods- and services producing activities of households for own use Activities of extra-territorial organisations and bodies.

- How many employees are there in your business?

1. Sole trader.
2. Self employed.
3. 1 - 5 employees.
4. 6 - 10 employees.
5. 11 - 24 employees.
6. 25 - 49 employees.
7. 50 - 99 employees.
8. 100 - 199 employees.
9. 200+ employees.

- What was your approximate annual turnover during the last financial year?

    1. less than 58,000 Euro.
    2. 59,000 Euro - 100,000 Euro.
    3. 101,000 Euro - 250,000 Euro.
    4. 251,000 Euro - 500,000 Euro.
    5. 501,000 Euro - 750,000 Euro.
    6. 751,000 Euro - 1 million Euro.
    7. 1 million Euro - 1.5 million Euro.
    8. 1.6 million Euro - 2 million Euro.
    9. 3 million Euro - 5 million Euro.
    10. 6 million Euro - 10 million Euro.
    11. 11 million Euro - 20 million Euro.
    12. over 20 million Euro.
    13. Refused

- Do hours of operation include 5pm to 7pm from from Monday to Friday?

    1. Yes.
    2. No.

- Does the premises operate during the weekend?

    1. Does not operate at the weekend.
    2. Saturday only.
    3. Saturday and Sunday.
    4. Sunday only.

- Approximately how old is your premises?
    1. Less than 5 years old.
    2. Between 6 10 years old.
    3. Between 11 and 30 years old.
    4. Between 31 and 75 years old.
    5. More than 75 years old.