Modelling the electricity consumption of small to medium enterprises

Presented at The 18th European Conference on Mathematics for Industry conference, Taormina Italy June 9–14th 2014

T. E. Lee, S. A. Haben & P. Grindrod

Abstract Estimating the demand on the low voltage network is essential for the distribution network operator (DNO), who is interested in managing and planning the network. Such concerns are particularly relevant as the UK moves towards a low carbon economy, and the electrification of heating and transport. Furthermore, small to medium enterprises (SMEs) contribute a significant proportion to network demand but are often overlooked. The smart meter roll out will provide greater visibility of the network, but such data may not be readily available to the DNOs. The question arises whether useful information about customer demand can be discerned from limited access to smart meter data?

We analyse smart meter data from 196 SMEs so that one may create an energy demand profile based on information which is available without a smart meter. The profile itself comprises of simply two estimates, one for operational power and another for non-operational power. We further improve the profile by clustering the SMEs using a simple Gaussian mixture model. In both cases, the average difference between the actual and predicted operational/non-operational power is less than 0.15kWh, and clustering reduces the range around this difference. The methods presented here out perform the flat profile (akin to current methods).

1 Introduction

Small to medium enterprises contribute a large proportion of the total energy demand in the UK but are often overlooked in research [2]. Therefore it is essential that their demand is accurately modelled so that distribution network operators (DNOs) can 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 [8]. Currently DNOs use After Diversity Maximum Demand [5] to model maximum demand for SMEs. This approach does not account for

T. E. Lee
Mathematical Institute, University of Oxford, UK e-mail: tamsin.lee@maths.ox.ac.uk

Key words: small to medium enterprises, low voltage, smart meter data, clustering

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 to DNOs. Nonetheless, they do have access to quarterly readings. As such, we ask if a DNO can estimate an SMEs electricity profile from quarterly readings and publicly available information. Since SMEs behaviour is regular and predictable (compared to domestic customers), one should be able to accurately recreate their demand profiles with this limited information.

In this paper we use preprocessed smart meter data from the Irish smart meter trial [4]. We consider a years worth of smart meter data for 196 SMEs starting from midnight 14th July 2009. From this data we demonstrate how one could accurately estimate customers weekly energy demand profile using only knowledge of their operating times (potentially public information) and their mean daily usage (potentially available from their quarterly meter readings available from the supplier).

We further improve this estimate by using a basic clustering of operational and non-operational power. 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 only knowing their mean daily demand and operational hours.

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. We consider the relationship between the survey responses and the clusters.

We begin in by describing how we determine operational and non-operational times from the smart meter data. In §3 we describe how we cluster the customers, and in §4 we compare our estimates based on clustering and not clustering. Finally, we summarise in §5.

2 Identifying operational hours

To determine a businesses operational hours we use a data driven approach. However, potentially this information is publicly available.

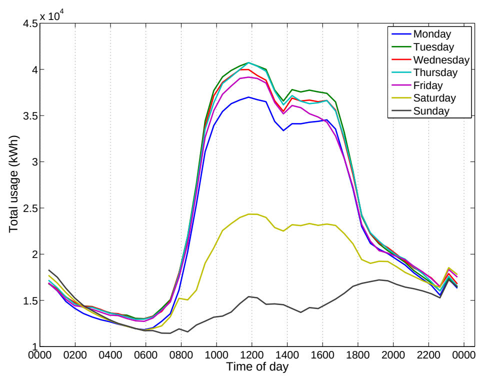
2.1 Operational days

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 example, suppose $\tilde{x}_m^{SUN} < q_m$ (for meter m), then we assume the business is closed on Sundays. To determine the quantile, we consider quantiles for probabilities from 0 to 1 (intervals of 0.01). For each quantile choice, the 'closed outcomes' for Saturday and Sunday is compared to a survey response in which customers state which days the business is open on weekends. From the F-score [7], we take the 0.4 quantile since it offers the closest match to the survey responses. There does not appear to be a distinct pattern for the spread of incorrectly predicted meters.

2.2 Operational hours

From the aggregation plot of all customers and days in Figure 1, a typical business day splits into four natural segments: (i) midnight - 9am, (ii) 9am to 6pm, (iii) 6pm to 10pm, (iv) 10pm to midnight. To determine whether a business is operational in a segment we first remove 'closed' days (see §2.1). Using the remaining 'open' days, the average power is calculated and compared to the average power in each of the four segments. If the average power for a segment is larger than the average power over all the 'open' days, we consider the business to be operational during this segment.

Fig. 1: The aggregated half hourly use from all 196 meters for particular days of the week.



3 Clustering

In order to group customers with similar attributes we use clustering based on their operational and non-operational power, and the half hourly standard deviation (as a measure of variation). We can then use these cluster groups to improve our estimate (see §4).

We model our three attributes as a finite mixture model (FMM) of uncorrelated Gaussian distributions (reference). The parameters of the mixtures and the mixing proportions are found easy through an implementation of the Expectation-Maximisation algorithm, which finds the parameters that maximise the likelihood function of the model [3]. Traditionally the k-means algorithm has been used for clustering in power systems. However, this is simply a less versatile model than the FMM [3]. We use the Matlab function `gmdistribution` to implement the algorithm.

As with many clustering algorithms, a disadvantage of the method is that the number of clusters must be defined before the clustering algorithm is implemented. 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 essentially penalises the maximum likelihood function by the number of parameters that are used in the model [3]. In particular, the BIC penalises larger numbers of clusters. By considering how the BIC changes with the number of clusters, we use it to choose a cluster size which obtains a good model of the observations without an excessive number of parameters. As shown in Figure 2a, five clusters is a reasonable choice.

Figure 2b shows our clusters in terms of their usage and normalised standard deviation. Since the groups form clear partitions of daily mean, μ_m, an SME without smart meter data can be placed in a cluster using the quarterly readings alone. It was found when comparing the clusters that there is no relationship between the survey responses (such as number of employees, annual turnover, age of building etc.) and the clusters (therefore the daily mean). This supports previous research which also found little correlation between energy consumption and household type [6].

4 Predicting electricity use

We estimate the average operational power $\hat{e}_{m,O}$ and average non-operational power $\hat{e}_{m,N}$ (for meter m) for each meter $m = 1, 2, \ldots 196$ using the mean for the corresponding variables from the remaining 195 meters. Therefore, since we know the operational hours for meter m, we estimate the total predicted energy during an average week, T_m, for meter m, via

$$T_m = H_{m,O} \sum_{i=1}^{H_{m,O}} \hat{e}_{m,O} + H_{m,N} \sum_{i=1}^{H_{m,N}} \hat{e}_{m,N}, \tag{1}$$

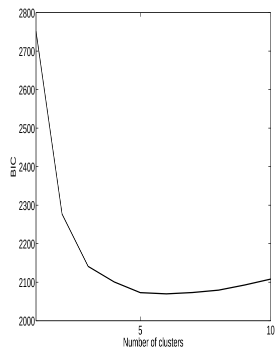
where $H_{m,O}$ is the number of operational hours and $H_{m,N}$ is the number of non-operational hours in a week. Ideally, the estimate (1) matches the weekly energy use for the meter, $T_m = 7\mu_m$ (obtained for quarterly reading data). Therefore, we adjust the profile by setting

$$e_{m,O} = \frac{7\mu_m}{T_m}\hat{e}_{m,O} \quad \text{and} \quad e_{m,N} = \frac{7\mu_m}{T_m}\hat{e}_{m,N}, \tag{2}$$

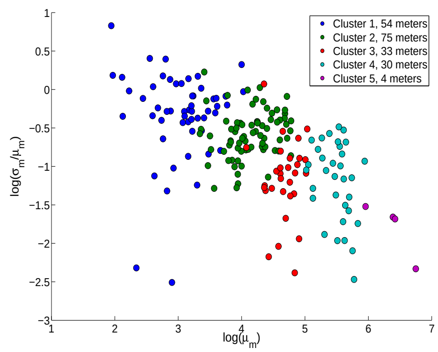
where $e_{m,O}$ is the adjusted average operational power, and $e_{m,N}$ is the adjusted average non-operational power. Using these adjusted power values for

Fig. 2: 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 most suitable.



(b) The five clusters arranged by their daily mean μ_m and normalised standard deviation σ_m/μ_m. (Domestic households have upper bounds of $\log(\sigma_m/\mu_m) \approx 0.65$ and $\log(\mu_m) \approx 1$, see [1].)



meter m, and the operational hours (see §2), we compose the predicted profile. This process is carried out for all meters. We repeat this process for the clustered data set, using the mean values for the current meter's cluster. As expected, the adjustment from $T/7\mu$ is larger when 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 3 compares this predicted average weekly profile with the actual average weekly profile for meter number 1021 using non-clustered data and clustered data. The peak 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. Two error measures are calculated, one for the operational power and one for the non-operational power,

$$E_{m,O} = \bar{e}_{m,O} - e_{m,O}, \tag{3}$$

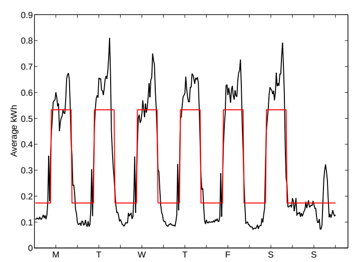
$$E_{m,N} = \bar{e}_{m,N} - e_{m,N}, \tag{4}$$

where $\bar{e}_{m,O}$ is the actual average operational power, and $\bar{e}_{m,N}$ is the actual average non-operational power. We consider the prediction coming from the whole data set, from the clustered data set, and from a simple flat prediction ($e_{m,O} = e_{m,N} = \mu_m/24$). Such an estimate comes from only knowing quarterly information and thus we have no time-of-use information.

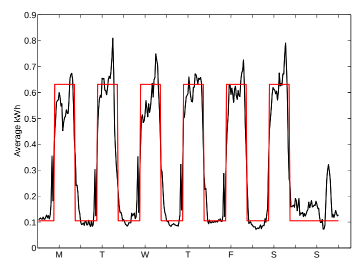
Both clustering and not clustering the data outperforms the flat estimate, see Figures 4. It is likely that the adjustment stage, equation (2), ensures that the non-clustered and clustered approaches are competitive to each other. To remove the effect of seasonality, we applied our methods to the first three months of the data set. The difference in results from using three months to one year is negligible, and hence not presented here.

Fig. 3: The predicted average week (red) and the actual average week for meter 1021. 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.



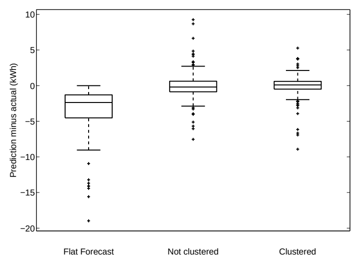
5 Conclusion

This paper identified when a business is open based upon their electricity use. With merely the operational hours (publicly available information) and the daily mean (available from quarterly readings), we have created a good 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.

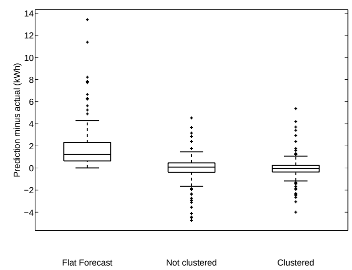
Counter-intuitively, there is no apparent correlation between opening hours and energy use. Nor have we found any correlation between data attributes (such as opening hours, daily mean, and standard deviation) and

Fig. 4: 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 1.13kWh, 0.08kWh and -0.05kWh.



(b) Average non-operational power. From left to right, the medians are -2.15kWh, -0.13kWh and -0.06kWh.



non-electricity features of the data (such as number of employees and the type of business). However, this is in line with earlier work, [6].

Acknowledgements 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. Haben, S. (2013) Categorisation of data: Technical report. *University of Reading.*
2. 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.
3. McLachlan, G., & Peel, D. (2004). *Finite mixture models.* John Wiley & Sons.
4. 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.
5. 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.
6. Morley, J. & Hazas. M. The significance of difference: Understanding variation in household energy consumption. (2011) *ECEEE Summer study*
7. 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.
8. 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.