Analysis and clustering of residential customers energy behavioral demand using smart meter data

Article

Accepted Version

Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data

Stephen Haben[*], Colin Singleton[**], and Peter Grindrod[*]

[*]Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK, e-mail: haben@maths.ox.ac.uk
[**]CountingLab Limited, Reading, RG6 6AX, UK, email: colin@countinglab.co.uk

November 27, 2015

Abstract

Clustering methods are increasingly being applied to residential smart meter data, providing a number of important opportunities for distribution network operators (DNOs) to manage and plan the low voltage networks. Clustering has a number of potential advantages for DNOs including, identifying suitable candidates for demand response and improving energy profile modelling. However, due to the high stochasticity and irregularity of household level demand, detailed analytics are required to define appropriate attributes to cluster. In this paper we present in-depth analysis of customer smart meter data to better understand peak demand and major sources of variability in their behaviour. We find four key time periods in which the data should be analysed and use this to form relevant attributes for our clustering. We present a finite mixture model based clustering where we discover 10 distinct behaviour groups describing customers based on their demand and their variability. Finally, using an existing bootstrapping technique we show that the clustering is reliable. To the authors knowledge this is the first time in the power systems literature that the sample robustness of the clustering has been tested.

1 Introduction

In power systems, clustering methods have traditionally been applied to energy data of high/medium voltage customers or on larger aggregations of customers [8], [7]. In recent years, research has increasingly considered clustering at the household level based on half hourly energy demand recorded by smart meters

[14]. Such methods are driven by the need for energy suppliers and distribution networks operators (DNOs) to better understand how residential customers use their energy and their effect on the low voltage (LV) networks. This issue is increasingly relevant as energy demand changes dramatically with the uptake of new technologies such as electric vehicles, photovoltaics, combined heat and power. In addition, new large energy demand technologies, not even thought of, could also be introduced, as shown by virtual currency mining. The UK roll-out of smart meters is expected to be complete by 2020 and will provide greater visibility of all residential customers energy use. However, since such data is proprietary it is not certain if DNOs will have complete access to such data and, even if it can be purchased, it is likely to be expensive [12]. Hence the value of the data must be properly assessed before infrastructure is put in place to manage it.

Smart meter data is being analysed through publicly available data sets such as the Irish smart meter trial and other research projects [5], [16], [17]. Clustering customers based on attributes derived from smart meter data creates better understanding of the different types of energy behavioral groups. The focus of this paper is to derive and cluster suitable attributes from smart meter data which can assist a DNO with two main applications for better LV network modeling and management. The primary application that motivates our approach is to help a DNO identify suitable customers for energy management solutions such as demand side response (DSR) and through storage devices. These applications have already been considered recently by a number of authors, see for example [5], [18], [10]. By choosing the correct attributes a DNO can use the clustering to identify suitable customer groups for demand reduction solutions and hence help to reduce network demand and volatility. For example, customers with heavy but regular demand in the evening time period could be ideal candidates for peak demand reduction through storage devices, whereas those with irregular demand may be more suitable for DSR. The secondary application is for identifying links between energy behavioural usage and publicaly available information (such as house-type and socio-demographics) [15], [3]. Such links can be used to improve modeling of residential customer demand and reduce necessary monitoring on the LV network. However, besides these two, the methodology can also be used for further applications of smart meter based clustering which we present in Section 2.

Choosing the correct attributes is potentially the most important aspect of a successful clustering. This is particularly challenging when considering extremely volatile household level demand which is much less regular than higher voltage demands [16]. The number of attributes should be optimized to: ensure the data distribution over parameter space is dense, reduce computational costs, and ensure the results can be easily interpreted [23]. In addition, if the number of attributes is not optimized then the clustering is less likely to be representative and fit for purpose. A main contribution of this paper is a detailed analysis of a large amount of domestic smart meter data, especially with regard to better understanding of peak demands and sources of variability. This helps us to identify and minimise the important attributes to be used in the clustering.

In addition, we discover four key time periods within which data should be analysed since they describe the most frequent largest demand behaviour.

There have been a variety of clustering methods which have been applied in the power systems literature [7], [3], [24]. In this paper we cluster our attributes by considering it as a finite mixture model (FMM) of Gaussian multivariate distributions, we justify the choice of method in Section 3 [19]. FMM are rarely applied to cluster smart meter data despite their versatility compared to the more common k-means method. We check the reliability of the outputs of the FMM clustering using a bootstrapping technique [13]. To the authors knowledge, checks of clustering reliability in the power systems literature have not been considered. We follow the bootstrapping method as described in [11]. Checking the sample robustness of the clustering is a vital part of ensuring that the final clustering is robust with respect to the sample used. This work has been developed in collaboration with Scottish and Southern energy power distribution, one of the DNOs in the UK, as part of the New Thames Valley Vision (NTVV) project.

In section 2 we perform a literature review of the recent clustering being applied to household level energy demand. In section 3 we introduce the theory behind the finite mixture model and the bootstrapping method for our final clustering. We derive the attributes selection in section 4 and apply the clustering and bootstrapping in section 5. In section 6 we summarise the paper and outline future work.

2 Literature Review

Historically, electricity data for a household is only available from real or estimated quarterly cumulatives which reveal no intraday information of a households electric usage behaviour. As such, much research has tried to infer energy demand characteristics from other sources such as socio-demographics and household properties (house type, size etc.). Although there exist correlations between some characteristics and attributes, such as maximum demand, they are weak and do not describe intra-day behaviours [20], [26, p.74]. There are also several studies which show there is a large degree of variation in demands which cannot be accounted for despite similar dwelling types, size and occupants (see [21] and references therein.). Other work has also shown that there are poor links between energy behaviour and typical socio-demographic classifications [15]. This suggests that to understand customer energy usage behaviour requires analysing actual consumption data.

Clustering at the individual domestic customer level has many potential uses for energy companies. In particular, clustering can ensure that experiments include a representative sample of the population (by requiring a sufficient number of customers in each group); allows an experimenter to adjust results to reflect biases in their sample; identify which characteristics correlate with energy behavioral use [3], [20]; help create more suitable tariffs [14]; and more accurate customer profiles [23], [25]; identify which customers are viable for energy saving

solutions such as demand response [5], [2], [10], [18]; identify which combination of customer types can cause network constraint violations; improving forecast methodologies [22], [6]; and compare different groups and trends in behaviour.

There are a large number of clustering methods which have been applied to household level data including, k-means and k-mediods [14], [23], [10], [4], finite mixture models [24], [25], principle component analysis, [1], self organizing maps [3] and spectral clustering [2]. Perhaps more important than the methods used in the clustering is the attributes which are clustered. These split into two main categories, those which try to cluster the full time series (for example 48 half hours a day) [14] and those which cluster specific features such as size of peak demand or distribution parameters [23], [10]. It is often preferable to cluster as few attributes as possible since this reduces computational expense and makes the clustering less sensitive to missing values [23]. This also avoids the curse of dimensionality. For example, if clustering hourly data over a day, and assuming a simple case where each hour can only take two values, a high and low, then there are potentially $2^{24} \approx 16.8$m different time series, which requires a huge number of clusters to isolate the distinct behaviours. Similarly, choosing a reduced number of features requires that they describe as much information as possible to be useful. This is particularly challenging for the volatile residential demand usage where a large number of behaviours can potentially exist.

3 Finite Mixture Modelling

Finite mixture models (FMMs) are a commonly used technique in clustering data attributes and have recently been used in clustering smart meter data [19], [25]. There are a number of practical advantages for clustering using FMMs which we discuss in this Section. A main advantage of FMMs over many other non-hierarchical clustering methods, such as k-means, is the ability to model a mixture of both continuous and categorical data. Categorical data can be particularly useful for including identifiers of low carbon technologies, for example electric vehicles which, although not implemented in this research, may be desirable in future clusterings. The FMM is set in a statistical framework and describes the distributions and correlations between attributes by automatically choosing the optimal weights for each of the input parameters for each cluster. This is in contrast to the k-means which is essentially a FMM but where the weights (axis scales) are fixed a priori. Additionally for FMMs there is no need to pre-define a distance measure and it allows variability for each attribute on a cluster-by-cluster level. A final advantage of the method is that the Bayesian information criterion (BIC), a common statistical criterion for selecting the optimal number of clusters, can be calculated as a by-product of the likelihood framework of the FMM for no extra cost. Disadvantages of the FMMs are that the number of clusters must be chosen a priori, the algorithm can converge to local optima and the method is sensitive to the input attributes chosen. Given the pros and cons, FMMs seem a suitable choice for our clustering task.

We describe the main properties of FMMs in this Section, for further details

see [19]. In FMMs we assume measurements are described by a finite mixture of distributions, typically from a single parametric family. The problem is often addressed using the expectation-maximization (EM) algorithm which considers the so-called *incomplete problem* [19]. In this setting the observed data is modelled as incomplete since information on which cluster an observations belongs to is unknown. A log likelihood function is formed which must be maximized to find the correct mixing proportions and parameters for the distributions.

The EM algorithm iteratively find a local maximum of the system by alternating between an E-step and M-step. In the E-step we calculate the expected log-likelihood, given the current values, to update the posterior probabilities of the component label vectors, where the probability of the j^{th} observation being in the i^{th} cluster is denoted by $\tau_{i,j}$. In the M-step, using the posterior probabilities, the expected log-likelihood function is maximized to update the mixing proportions and the distribution parameters. Once convergence has been achieved a soft or hard clustering can be produced. In the soft clustering each observation is assigned to every cluster with a certain probability determined by the final posterior probabilities, $\tau_{i,j}$, of the component labels. In the hard clustering the observation is assigned to the cluster with the largest posterior probability. In this paper we assume the underlying distributions are independent Gaussians, this has the added advantage that the E and M steps are particularly easy to calculate [19]. Caution must be displayed when implementing the EM algorithm. The EM algorithm converges quickly, but not necessarily to a global maximum, thus the algorithm must be run repeatedly, with different initial conditions, to ensure global convergence is achieved.

In non-hierarchical clustering there is always an ambiguity choosing the optimal number of clusters. A compromise must be made between having sufficiently many clusters to describe the majority of the variability, but few enough to avoid large dimensionality and ensure they are manageable. There are many clustering performance measures that have been developed [9]. For our work we consider the Bayesian information criterion (BIC) due to its statistical relevance and its ease of calculation within the log-likelihood framework of the FMM [19]. Essentially, the BIC penalizes the likelihood in proportion to the number of parameters in the model. An alternative to the BIC is the Akaike information criterion but we choose the BIC as it penalizes the number of parameters more strongly and, for operational reasons, we do not want too many clusters. Observing how the BIC improves as we increase the number of clusters allows us to justify the appropriate number of groups to use in the final clustering.

3.1 Bootstrapping

After producing the final clusters, the natural question arises as to how reliable are the outputs? This has not been considered in the clustering methods in the powers systems literature but is important for testing the robustness of the results. A reliable segmentation means that, for any individual household, we are quite certain which cluster the household belongs to. There are numerous reasons why a clustering might be unreliable. The technique used might be poor

(inappropriate for the data); the wrong attributes may be chosen; the nature of the data one is using might not lend itself to clustering; or there just might not be enough data for the clustering to be reliable.

Bootstrapping methods are common tests for measuring the accuracy of parameter estimates (for example, the variance or mean) of the underlying, unknown, true distribution, by sampling from an approximate distribution [13]. Bootstrapping is particularly useful when there is a small amount of empirical data, as in our case with only 3622 meters available. Traditionally bootstrapping is repeated hundreds or thousands of times for various alternative realizations of the data. Thus one covers a large number of the possibilities that could have occurred (although the number of possibilities is extremely large), enough to estimate uncertainties around the measurements and statistics of interest. The reliability of the bootstrap method depends upon the number of times the process is repeated (which we are free to choose) and also how well the underlying distribution can be approximated from the real data. We note that this method is dissimilar to cross-validation methods which serve to validate the model rather than estimating the uncertainty of our classification.

For our clustering we resample, with replacement, the attributes from $N = 3622$ customers from the data and then recluster, we repeat this $M = 10,000$ times. Following the work in [11] we calculate particular variables to indicate the robustness of our clustering. We first consider a measure of classification uncertainty for customer j given by

$$e_j = \min_i (1 - \tau_{i,j}), \tag{1}$$

where $\tau_{i,j}$ are the final posterior probabilities found for the current bootstrap. The classification uncertainty is the sum of the probabilities of all the classes that customer j does not belong. If classification is certain e_j is close to zero. If uncertain then e_j will be close to $(g-1)/g$, (i.e. random) where g is the number of groups in the clustering. We keep track of e_j for each customer for each bootstrap. As an aggregate measure of the uncertainty of the overall cluster model we consider

$$E = 1 - EN(\tau)/(N \log(g)), \tag{2}$$

where

$$EN(\tau) = -\sum_{j=1}^{N} \sum_{k=1}^{g} \tau_{k,j} \log(\tau_{k,j}). \tag{3}$$

The relative entropy, E, provides an estimate as to how well separated the classes are, with E close to one for well separated classes and E close to zero for poorly separated classes. We can obtain an estimate of E from the original data with the EM-algorithm but the bootstrap can also be used to provide an estimate of how robust the clustering is to sampling errors.

Finally we consider the mean and variance of each parameter within each class. In an ideal situation these values will be the very similar in each bootstrap but due to errors through relabelling we simply use the means (ignoring the variances) and for each parameter, report the nearest class mean value to the

one obtained with the true (original) sample. This may underestimate the uncertainty in the mean parameter value for that class but at least should give some indication. We then average over all these "nearest means" to obtain a rough estimate of the sampling uncertainty in the definition of the cluster mean.

4 Data Analysis and Attribute Selection

Choosing the appropriate attributes for the application is essential if we are to produce informative groups from our clustering. Our main application is to identify households who are appropriate for demand reduction through demand side response or the implementation of storage devices. Hence, we are interested in discovering clusters describing the different types of peak demands and major variabilities and seasonalities in the data. In order to reduce the dimensionality, in this work we focus on clustering on a limited number of attributes. We discover four key time periods in which the data should be analysed. This enables us to choose a lower temporal resolution version of the smart meter data to best describe customer peak demand behaviour.

The data used in the clustering is from the publicly available Irish smart meter trial which consists of half hourly energy readings from smart meter data for just under 4000 residential customers from around Ireland [17]. We considered the first year of data (starting midnight 14^{th} July 2009) and after removing customers with spurious and missing data reduced the set to 3622 customers. In the trial the customers are assigned to different tariffs but we do not distinguish them in the clustering we present in Section 5. Fig. 1 shows the normalized variation of the half hourly demand against the mean for each customer. The figure highlights two important features which must be considered before we create our attributes. Firstly the variability in the mean daily demands indicates that it is important to normalise features which describe the intra-day demand since otherwise the final clusters may simply describe the mean demand and not the distribution of demand over the day [7], [14]. Secondly the wide spread in normalized standard deviation shows the importance, when describing average attributes, to include a measure of the variability to get a fuller understanding of the customer behavioral use. Without such a measure we have a limited understanding of how representative the mean cluster value is of its members.

Due to the natural volatility of residential customers, modelling peak demand is challenging. For most customers (2700) a large peak usage of 5kWh over any half hour, occurs at least once throughout the year but these occur less than 5% of all half hours in the dataset. A much more managable, but still important, task is to understand the occurance of frequently occuring daily peak demands. Fig. 2 shows a count according to time of day of when different size peaks occur. The counts are broken down according to seasons (seasons are determined by standard equinox and solstice dates for Ireland) and show that different size peaks occur at different times of the day depending on the season.

The plots show consistent time intervals during the day when the largest demands occur. In addition, the largest demands are quite seasonally driven and

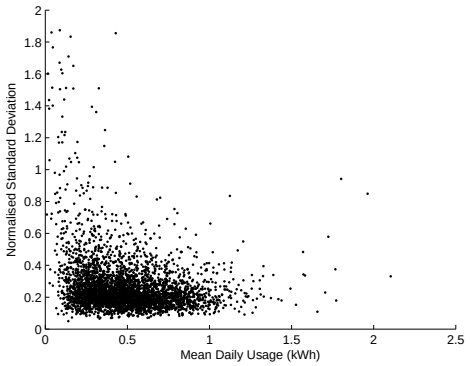

Figure 1: Normalized standard deviation of daily demand versus mean daily demand for each residential customer.

occur at particular time intervals. For example, in Fig. 2 (a) and (b), when considering smaller demands, the distributions are focused during the evening time period (about 3.30pm to 10.30pm) and also the daytime (about 9am-3.30pm). In addition, the daytime peaks are weighted towards weekend days (not shown) as expected. In contrast, the largest peaks, as shown by Fig. 2 (c) and (d), show that the breakfast time period (about 6.30am-9am) and overnight time period (about 10.30pm to 6.30am) become more prominent. In particular the overnight time period is likely to be due to high energy overnight storage heaters which coincides with the fact that the majority of these peaks are from the Autumn and Winter seasons. However, we note that the the largest demands, as shown in Fig. 2 (d) consists of, circa 1000 individual customer half hours, which is a relatively small fraction of the 63m customer-half-hours considered and small compared with the number of customer-half hours exceeding 0.9kWh as shown in Fig. 2 (a) which is of an order of magnitude of around 10^5.

Fig. 3 shows the distribution of peaks for a larger number of demands from 0.9kWh to 11kWh in 0.2 kWh intervals. The plot confirms Fig. 2 with the largest distributions of peak demand occurring in four time periods as indicated by the four local maxima in the distributions. The mean proportion is shown for clarity and again indicates that there are some common time periods where larger demands tend to occur. Motivated by this analysis we define four key time periods, chosen by the approximate local minimums between the modal peak times as shown in Fig. 3. The four chosen time periods are

- **Time period 1**: 10.30pm-6.30am - The Overnight period
- **Time period 2**: 6.30am-9.00am - The Breakfast period
- **Time period 3**: 9.00am-3.30pm - The Daytime period
- **Time period 4**: 3.30pm-10.30pm - The Evening period

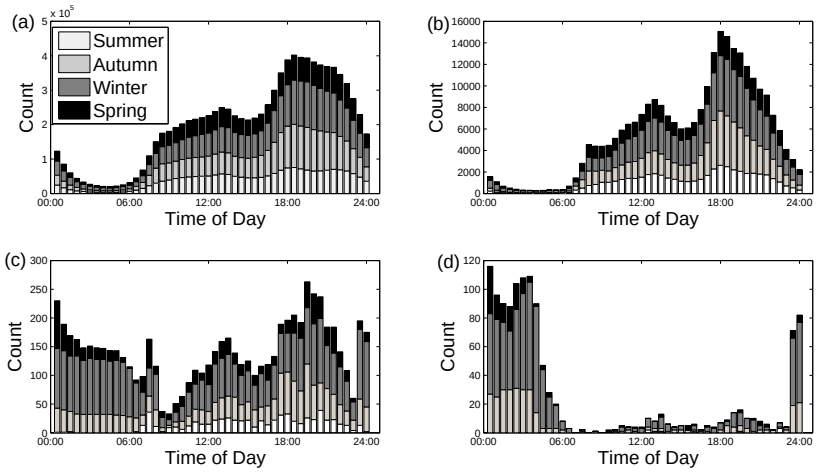

Figure 2: Count of number of half hours in the data set which exceed consumptions of (a) 0.9kWh (b) 4.1 kWh, (c) 8.1 kWh and (d) 10.7 kWh with respect to time of day. The counts are broken down according to Season.

Perhaps unsurprisingly the chosen time periods coincide with typical periods of household activity. These time periods allow us to reduce the dimensionality of our attributes.

To investigate the daily and seasonal effects on demand, we consider the aggregate demand of all customers in each time period throughout the year. Fig. 4 (a)-(d) shows the total aggregated usage for each day of the trial for each time period. The seasonal effect is most prominent in the evening time period (Fig. 4 (d)) with smaller effects of seasonality visible in the breakfast and overnight time periods (Fig. 4 (b) and 4 (a) respectively). The increasing demand is likely due to electric heating and lighting. The lower seasonal effect in the daytime period (Fig. 4 (c)) is likely due to reduced occupancy. In addition, weekend effects are also clearly visible in Fig. 4 (a)-(c). Peaks in demand in the overnight period occur at the weekend (Friday and Saturday night) due to late nights whereas dips in demand occur in the breakfast period on the Saturday and Sunday (most likely due to occupants waking up at later periods). The dips in demand during the breakfast time (Fig. 4 (b)) period correlate with increased demand on the Saturday and Sunday in the daytime period. There is not much difference in the weekends for the evening period since meal times are likely to be similar day-to-day. Large changes in demand due to seasonality and weekly (weekend vs. weekday) effects occur differently depending on the time periods thus we split this variation accordingly when defining our attributes.

Finally, we note that there are some unusual peaks occurring on specific days. Large peaks occur on Christmas and New Years eve/day and there are

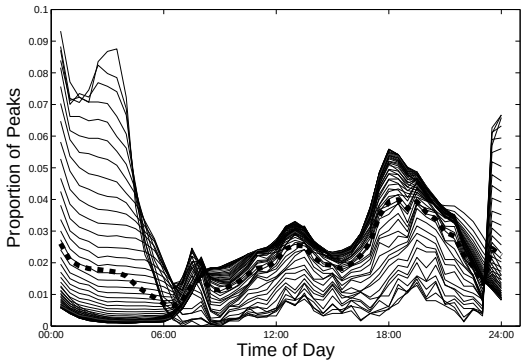

Figure 3: Distribution of customer half hour demands greater than various magnitudes with respect to time of day. The mean proportion is shown as the bold dashed line.

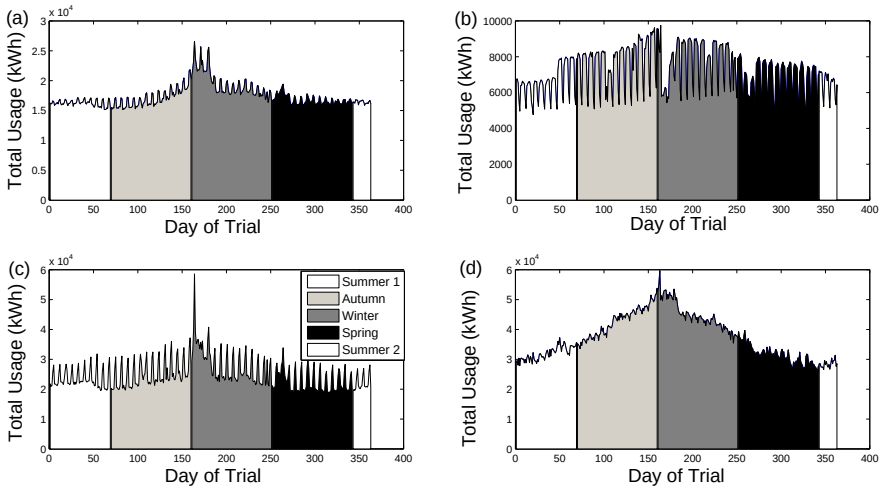

Figure 4: Sum of usage of all customer for each day in time (a) period 1 (Overnight), (b) period 2 (Breakfast), (c) period 3 (Daytime) and (d) period 4 (Evening).

increases in demand during the Easter holidays (4^{th} April 2010). There is a particularly large demand on the 9^{th} and 10^{th} of January 2010. This could be related to the extreme Winter weather conditions which occurred on this day[1]. The special days: Christmas eve and day, New years eve and day and the 9/10 January are removed from our analysis to ensure that we are modelling the typical behaviour of customers. Such days would be perhaps better modelled on a case-by-case basis. There is also variability which may not be accounted for by the seasonality and type of day. For this reason we also must include a general measure of the variability and irregularity of each customer.

To define our attributes we introduce some notation. For a particular customer we define P_i to be the mean power in each time period $i = 1, 2, 3, 4$ over the entire years worth of data, with corresponding standard deviation σ_i. We let $\hat{P}$ be the daily mean power for this customer over the entire years worth of data. In addition, we define the mean power over the Summer and Winter seasons in each time period as P_i^S and P_i^W respectively. Finally we also require the mean weekend and weekday power in each time period over the entire years worth of data, which we notate as P_i^{WE} and P_i^{WD} respectively. Using the above notation we define the following attributes for each customer

- **Attributes 1 to 4**: The relative average power in each time period over the entire year, $P_i^R = P_i/\hat{P}$, $i = 1, 2, 3, 4$.
- **Attribute 5**: Mean relative standard deviation over the entire year given by $\hat{\sigma} = \frac{1}{4}\sum_{i=1}^{4} \sigma_i/P_i$.
- **Attribute 6**: A seasonal score given by $S = \sum_{i=1}^{4} \frac{|P_i^W - P_i^S|}{P_i}$.
- **Attribute 7**: A weekend vs weekday difference score given by $W = \sum_{i=1}^{4} \frac{|P_i^{WD} - P_i^{WE}|}{P_i}$.

Thus for each customer we have defined seven attributes which give a summary description of their yearly behaviour. We normalise the power since otherwise the clustering is simply a function of the mean daily power [7], [14]. This is not desirable since we are interested in the time of use behaviour. We calculate the seasonality in each time period before summing since, as shown in Fig. 2, the demand increased in some time periods whereas it reduces in others and thus by taking an absolute sum we can get a measure of the overall seasonal effect. In addition, we refrained from using a seasonality for each time period to minimise the number of attributes. Similarly we took the absolute sum of the weekend versus weekday difference.

Normalizing the seasonal score and weekend vs weekday difference allows us to consider relative change rather than absolute change. Without normalization low average demand customers would be perceived to have a greater change in their behaviour compared to heavier demand customers. Finally there are variations in demand which are not a result of the type of day or the seasonality and

[1] `http://en.wikipedia.org/wiki/Winter_of_2009-10_in_Great_Britain_and_Ireland`

therefore we included the relative standard deviation to calculate the variability of a customers behaviour. Large difference in demand due to seasonality and day type can cause an increase in standard deviation but there can also be a large standard deviation without much seasonal and day differences. We considered the relative standard deviation in each time period but they were strongly correlated and hence we took the mean. We note that only weak correlations exist between the final attributes and hence we can assume this in our finite mixture model.

We considered other attributes to cluster. We attempted splitting the day into 3 and 4 equal length time periods but we found this increased the variability in each time period since, as shown in Fig. 3, this effectively split times of similar behaviour into separate groups. In addition, we also considered clustering the average daily time series profile (i.e. 48 attributes) but this caused two main problems. Firstly, the large dimensionality and stochasticity of household level demand meant that the centroids of the final clusters didn't accurately represent the group members very well and, similar to results in [14] and [20], several clusters produced mean profiles which were not easily distinguishable. In addition, to include variability measures would require increasing the dimensionality even further.

5 A finite mixture based clustering

For each customer we calculated the seven attributes as defined in Section 4. We do not consider the overall mean daily power so that we can focus on the time of demand rather than the total amount. This also provides the opportunity of a hierarchical structure for our clustering where initial customers are macro-classified according to their overall energy usage, say into types A, B, C etc. and then also clustered according to the attributes presented here into 1s, 2s, 3s, etc. so that customers can be classified according to a combination of the two, e.g. C3s. Since the correlations are weak between our chosen attributes we model the multivariate Gaussians in our FMM with uncorrelated covariance matrices. This also ensures a more stable convergence of the EM algorithm.

5.1 Clustering Results

For different numbers of clusters the algorithm is run 1000 times with different initial conditions to ensure that the global maximum likelihood is found. To determine the number of clusters we considered the BIC for the different number of clusters as shown in Fig. 5 (a). As the plot shows, increasing the numbers of clusters from about one to five brings a large reduction in the BIC but limited reductions beyond ten clusters. We found that ten clusters also provided a manageable number of behaviours with clearly distinguishable features for ease of use by a DNO. Due to this and the diminished returns in the BIC we fixed out choice at ten clusters.

None of the ten clusters has sparse membership with at least 170 customers

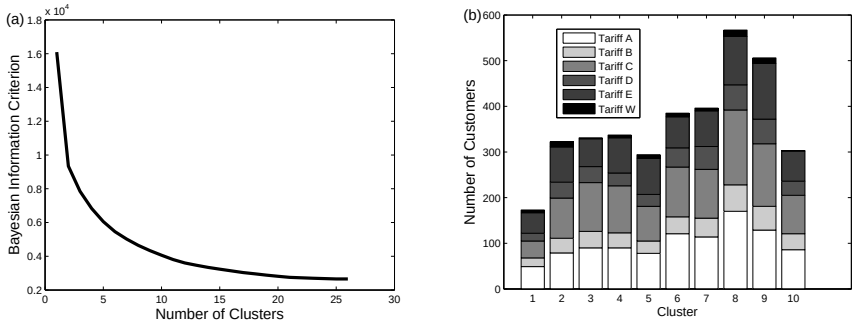

Figure 5: (a) BIC for different size clusters and (b) Numbers of customers per group in final clustering.

in each cluster as shown in Fig. 5 (b). Each bar is broken down in terms of the number of customers in each tariff from the Irish smart meter trial (see [17] for more details). Since no cluster is dominated by any tariff type then perhaps it can be suggested that no significantly new behaviours are created by different tariff types.

The mean attributes of the final clusters are shown in Table 1 ordered in terms of evening relative power. As shown from the second to fifth columns of Table 1 many of the clusters can be distinguished by the mean overnight, breakfast, daytime and evening relative power values. Fig. 6 shows the normalized mean daily profile from clusters $2, 4, 7$ and 10 which show profiles with large average daytime, overnight, breakfast and evening demands respectively. The clusters thus identify customers with heavy usage and high volatility at different time periods who could create violations of the low voltage network constraints. Some clusters have very similar mean profiles, however these customers are distinguishable by other attributes such as their seasonality (as with cluster 5 and 6).

The clusters can be used to identify groups of customers who are suitable for various demand reduction initiatives [18]. For example, from Table 1 we can see that cluster 8 customers have high evening demand with low variability (as seen from the STD, seasonal and Weekend difference attributes). Such customers could have their peak load reduced through the implementation of storage devices. Cluster 5 customers also have low variability in general yet have quite high seasonal difference in demand. Hence, such customers could be offered non-electric or more efficient hearing alternatives in order to reduce their Winter demand.

The clusters can also be useful for identifying customer behaviour types from non-energy based characteristics. We considered smart meter data from a separate trial called the New Thames Valley Vision project[2]. This data consists

[2]See `http://www.thamesvalleyvision.co.uk/` for more details.

Table 1: The mean value of each of the 7 attributes for each cluster.

	Attributes						
Cluster	Overnight	Breakfast	Daytime	Evening	Mean STD	Seasonal	Week diff
1	0.72	1.11	1.07	1.21	1.18	3.47	1.26
2	0.44	0.80	1.45	1.29	0.53	1.05	0.76
3	0.75	0.78	1.06	1.31	0.42	0.81	0.44
4	0.90	0.75	0.85	1.34	0.60	1.61	0.63
5	0.58	0.78	1.17	1.40	0.76	2.37	0.61
6	0.55	0.74	1.17	1.44	0.48	0.93	0.65
7	0.56	1.24	0.94	1.48	0.59	1.22	1.43
8	0.67	0.68	0.93	1.56	0.49	0.98	0.74
9	0.44	0.70	1.10	1.65	0.53	1.01	0.96
10	0.53	0.65	0.80	1.85	0.64	1.41	1.07

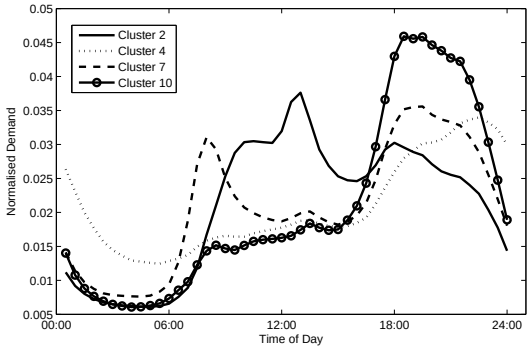

Figure 6: Mean Daily profiles for a selection of the final clusters.

of half-hourly smart meter data from 235 residential volunteers. We considered a years worth of data beginning 27^{th} February 2013 and assigned them to the clusters (by calculating the posterior probabilities $\tau_{i,j}$) using their derived attributes. For the NTVV customers we also had access to the profile class information and thus found that 70% of customers with overnight storage heaters were members of cluster 1. This is unsurprising given the large relative seasonality score for this group. We note, that if we cluster the entire mean half hourly profile the final cluster profiles are more similar and we miss certain features, for example, the cluster consisting of large overnight demand. By construction, the clusters naturally identify customers who have heavy demand during particular time periods of the day.

There is a small degree of subjectivity on the precise endpoints selected for the time periods. To test how sensitive the final clusters are to small changes

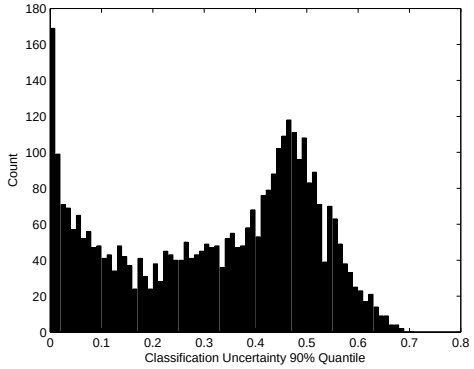

Figure 7: The 90% quantile of the classification uncertainty from the bootstrap samples for each customer.

in the boundary we implemented the clustering algorithm eight times (each endpoint changed by a half hour earlier or later) to the same attributes but with different time periods. In each case the average change in cluster centres was less than 6%. Hence the centres are similar when similar time periods are chosen and not too sensitive to changes in the endpoints.

5.2 Bootstrapping

We create M = 10,000 bootstrap samples and consider the uncertainty measures defined in Section 3.1 to test the reliability of the clustering. The entire process took less than 2 hours on a basic desktop computer (single core processor at 2.30GHz and 4GB RAM memory) and only 2 of the bootstraps did not converge which, since this is such a low number, were ignored. We first consider the classification uncertainty (1). This indicates how large our classification uncertainty could be if the clusters had been defined differently. In Fig. 7 we report on the worst 10% for each customer. The plot shows that there is a number of customers without very certain clusters (close to 0.5) but there are an equally large number of customers who are very certain (close to zero). However, since this is the 90% quantile the average classification uncertainty is much better than what is shown in the plot. The customers who tend to be uncertain most likely lie on the boundaries between multiple cluster centres. This is confirmed by the fact that a large proportion of customers have a classification uncertainty of just less than 0.5 meaning that they are as likely to lie in all other clusters as lie in the most probable cluster.

Next we consider the Entropy given by equation (2). The entropy for the original sample was $E = 0.7972$ and the mean over all bootstrap samples is $E = 0.8056$ with standard deviation 0.0063. The similarity of entropy values amongst all samples suggests that the entropy is not affected by sampling. Also,

crucially, all values are close to 1 suggesting that the clusters are well separated and thus well defined.

Finally we considered the mean values over all bootstrap of each attribute for each cluster. As mentioned in Section 3.1 due to potential mislabeling in each clustered bootstrap we first matched the clusters centres in each bootstrap to the closest (with respect to the Euclidean distance) centre in the original clustering presented in Table 1. We found that these mean clustered bootstrap samples matched the original mean centres given in Table 1 with all attribute means agreeing to the first decimal place in all cases. In addition when considering the standard deviation of each attribute in each cluster over all bootstraps none were larger than 10% of the mean values given. This is unsuprising since considering the classification uncertainty is small most customers had a definite cluster and thus a stable cluster centre from one bootstrap sample to the next. The bootstrapping results confirm that the majority of customers have been classified with certainty. Hence, having assigned customers in the initial clustering we can be confident that they belong to their respective clusters described by a particular mean and covariance of our chosen attributes.

6 Summary

Large quantities of information about how customers use their energy is becoming available through the uptake of smart meters. Clustering the most important attributes of customers is a very common method for better understanding the different residential energy behaviours that exist and has many applications. In this paper we present a methodology for extracting, classifying and then verifying the reliability of the final clustering. We began with detailed analysis of smart meter data to identify some of the most important characteristics. In particular we analysed different size demands and their distribution as a function of time-of-day and season. We identified four key time periods which described different peak demand behaviour, coinciding with common intervals of the day: overnight, breakfast, daytime and evening. We also found that demand in the different time periods changed as a function of seasonality and days of the week, thus identifying two major sources of variation. In addition, we also included a mean normalized standard deviation of the demand as a measure of the irregularity of a customer. The time periods not only helped us to model peak time behaviours and their variation but also allowed us to reduce the number of attributes in our clustering implementation. We presented a clustering of our chosen attributes into ten groups using a finite mixture of Gaussian distributions. Such a method is commonly used in clustering but has not been explored in great detail within the power systems literature despite the advantages over more common methods. We showed that the final clusters identified many important behaviours of the customers. As well as time periods of greatest demand, we identified those customers with the largest variability according to seasonality and weekend versus weekday differences. Understanding such changes in a customers seasonal behaviour can aid network operators

in longer term planning of the networks. Once we have assigned customers to a cluster we would like to evaluate how confident we are that the customers belongs to the group they were assigned. That way we can be more confident that the cluster centers are representative of the members. Such classification uncertainty measures are not often included in the clustering power systems literature. Hence, we assessed the sample robustness of our clustering using three measures applied to several thousand bootstrapped samples. Following a bootstrapping method presented in [11] we considered three measures of sample robustness in our clustering We verified that, assuming the underlying distribution of the data is well approximated from the real data, the final clustering is reliable with a high degree of certainty of which customers belonged to which cluster.

Natural extensions of the work is to incorporate the modelling of significant low carbon technologies into the clustering. Further work is especially needed to test the limits to which clustering can reduce the need for expensive monitoring. Previous work has shown the potential of linking demand attributes to public household data [20]. Such, even weak, correlations can be utilized, together with other information, to more accurately model household level demand. Finally, many more smart meter-based trials are being produced in order to better understand the LV network. Clustering, such as the one presented in this paper can be used to ensure representative customers have been monitored in such trials. Such methodology can potentially reduce costs and ensure that results are statistically significant.

Acknowledgments

We wish to thank Scottish and Southern Energy Power Distribution (SSEPD) for their support via the New Thames Valley Vision Project (SSET203 New Thames Valley Vision), funded through the Low Carbon Network Fund.

References

[1] J. M. Abreu, F. P. Camara, and P. Ferrao. Using pattern recognition to identify habitual behavior in residential electricity consumption. *Energy and Buildings*, 49:479–487, 2012.

[2] A. Albert and R. Rajagopal. Smart meter driven segmentation: What your consumption says about you. *IEEE Trans. Power Syst.*, 28:4019–4030, 2013.

[3] C. Beckel, L. Sadamori, T. Staake, and S. Santini. Revealing household characteristics from smart meter data. *Energy*, 78:397–410, 2014.

[4] I. Benitez, A. Quijano, J-L. Diez, and I. Delgado. Dynamic clustering segmentation applied to load profiles of energy consumption from spanish customers. *Electrical Power and Energy Sys.*, 55:437–448, 2014.

[5] H. Cao, C. Beckel, and T. Staake. Are domestic load profiles stable over time? an attempt to identify target households for demand side management campaigns. In *39th Annual Conference of IEEE Industrial Electronics Society (IECON 2013)*, pages 75–86, Vienna, Austria, 2013.

[6] M. Chaouch. Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves. *IEEE Trans. Smart Grid*, 5:411–419, 2014.

[7] G. Chicco. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, 42:68–80, 2012.

[8] G. Chicco and F. Piglione R. Napoli. Comparisons among clustering techniques for electricity customer classification. *IEEE Trans. Power Syst.*, 21:933–940, 2006.

[9] I. Dent, T. Craig, U. Aickelin, and T. Rodden. An approach for assessing clustering of households by electricity usage. In *UKCI 2012, 12th Annual Workshop Comp. Intelligence*, Heriot-Watt, UK, 2012.

[10] I. Dent, T. Craig, U. Aickelin, and T. Rodden. Finding the creatures of habit; clustering households based on their flexibility in using electricity. In *Digital Futures 2012*, Aberdeen, UK, 2012.

[11] J.G. Dias and J.K. Vermunt. Bootstrap methods for measuring classification uncertainty in latent class models, a. in rizzi and m. vichi (eds.). In *COMPSTAT2006. Proc. Comp. Stats, Heidelberg: Physica/Springer-Verlag*, pages 31–41, 2006.

[12] EA Technology. Assessing the impact of low carbon technologies on great britain's power distribution networks (last accessed june 2014), 2014.

[13] B. Efron. Bootstrap methods: Another look at the jackknife. *Ann. Stats.*, 7:1–26, 1979.

[14] C. Flath, D. Nicolay, T. Conte, C. van Dinther, and L Filipova-Neumann. Cluster analysis of smart metering data - an implementation in practice. *Bus. and Syst. Eng.*, 4:31–39, 2012.

[15] S. Haben, M. Rowe, D. V. Greetham, P. Grindrod, W. Holderbaum, B. Potter, and C. Singleton. Mathematical solutions for electricity networks in a low carbon future. In *CIRED 22nd International Conference on Electricity Distribution*, Stockholm, Sweden, 2013.

[16] S. Haben, J. A. Ward, D. V. Greetham, P. Grindrod, and C. Singleton. A new error measure for forecasts of household-level, high resolution electrical energy consumption. *Int. J. of Forecasting*, 30:246–256, 2014.

[17] Irish Social Science Data Archive. CER Smart Metering Project, 2012.

[18] J. Kwac, J. Flora, and R. Rajagopal. Household energy consumption segmentation using hourly data. *IEEE Trans. Smart Grid*, 5:420–430, 2014.

[19] G. McLachlan and D. Peel. *Finite Mixture Models*. Wiley Series in Probability and Statistics, 2000.

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

[21] J. Morley and M. Hazas. The significance of difference: Understanding variation in household energy consumption. In *ECEEE 2011 Summer School*, pages 2037–2046, Stockholm, Sweden, 2011.

[22] F. L. Quilumba, W. Lee, H. Huang, D. Y. Wang, and R.L. Szabados. Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities. *IEEE Trans. Smart Grid*, 6:911–918, 2014.

[23] T. Rasanen and M. Kolehmainen. Feature-based clustering for electricity use time series data. *Adaptive and Natural Computing Algorithms: Lecture Notes in Computer Science*, 5495:401–412, 2009.

[24] B. Stephen and S. J. Galloway. Domestic load characterization through smart meter advance stratification. *IEEE Trans. Smart Grid*, 3:1571–1572, 2012.

[25] B. Stephen, A. J. Mutanen, S. J. Galloway, G. Burt, and P. Jentausta. Enhanced load profiling for residential network customers. *IEEE Trans. Power Delivery*, 29:88–96, 2014.

[26] J-P Zimmermann, M. Evans, J. Griggs, N. King, L. Harding, P. Roberts, and C. Evans. Household electricity survey: A study of domestic electrical product usage. Intertek Report R66141, 2012.