This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Multi-Resolution Load Profile Clustering for Smart Metering Data

Ran Li, Furong Li, *Senior Member, IEEE*, and Nathan D. Smith

***Abstract*—This paper proposes a novel multi-resolution clustering (MRC) method that for the first time classifies end customers directly from massive, volatile and uncertain smart metering data. It will firstly extract spectral features of load profiles by multi-resolution analysis (MRA), and then cluster and classify these features in the spectral-domain instead of time-domain. The key advantage is that the proposed method will allow a dynamic load profiling to be flexibly re-constructed from each spectral level. MRC addresses three key limitations in time-series based load profiling: i) large sample size: sample size is reduced by a novel two-stage clustering, which firstly clusters each customers' massive daily profiles into several Gaussian mixture models (GMMs) and then clusters all GMMs; ii) volatility: it avoids the interferences between different load features (e.g. magnitude, overall trend, spikes) by decomposing them onto different resolution levels, and then clustering separately; iii) uncertainties: as the GMM can give a probabilistic cluster membership instead of a deterministic one, an additive classification model based on the posterior probability is proposed to reflect the uncertainty between days. The proposed method is implemented on 6369 smart metered customers from Ireland, and compared with the load profiles used by the U.K. industry and traditional K-means clustering. The results show that the developed MRC outperformed the traditional methods in its ability in profiling load for big, volatile and uncertain smart metering data.**

***Index Terms*—Big data, clustering, customer classification, GMM, load profiles, multi-resolution analysis, spectral analysis, X-means, smart grid, smart meter.**

I. Introduction

THE increasing number of low carbon techniques (LCTs), e.g., electric vehicles, photovoltaic and heat pumps, will bring pressure to distribution networks in terms of thermal and voltage constraints [1], [2]. One solution to mitigate the pressure is demand side response (DSR), where customers vary demand to support distribution network operators in addressing network problems [3]. The realization of DSR is commonly based on the flexibility of customers' electricity consumption, which requires characterization of customers' load. In order to characterize the diverse and massive end customers, load profiling is traditionally used to classify similar customers and create *typical load profiles* (TLPs).

Manuscript received May 03, 2015; revised September 05, 2015 and January 06, 2016; accepted February 27, 2016. Paper no. TPWRS-00618-2015.

The authors are with the Department of Electronic and Electrical Engineering, University of Bath, Bath BA2 7AY, U.K. (e-mail: rl272@bath.ac.uk; f.li@bath.ac.uk; n.d.smith.93@hotmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2016.2536781

The popularization of smart meters brings the opportunity for more accurate load profiling. However, they also present major challenges to the traditional load profiling techniques. Smart metering data are massive, volatile and uncertain, leading to significant differences between TLPs and individual smart meter readings. Thus, DSR strategies based on the TLP may not guide individual customers to effectively support energy and network needs [4], [5]. Worse, it could further aggravate energy or network problems.

Technically, directly applying previous load profiling methods on the smart metering data could have three major limitations as follows.

i) Large sample size: massive number of customers and days will increase the computational burden, especially for techniques requiring distance matrices such as hierarchical clustering [6], [7]. Traditional data reduction techniques [8] also have two main drawbacks (e.g., principal component analysis): a) they only reduce the number of variables of each sample, but not the sample size; b) they may discard some similar sample points (e.g., common evening peaks), which cannot be recovered, thus causing detailed information loss.

ii) Volatility: An essential prerequisite for traditional load profiling methods (clustering analysis in the time domain) is the sufficient similarity in load patterns. However, load profile of individual smart meters can be extremely volatile. Directly applying previous methods would result in either massive groups or huge variances within groups. A sudden spike or a tiny time shift (e.g., communication delay) may lead to completely different clustering results. Different factors, such as magnitudes, overall trends and spikes will interfere with each other during the clustering. Therefore, instead of individual customers, previous research usually focuses on "large customers" or "aggregated customers" whose load profiles are more smooth and regular [6]–[11].

iii) Uncertainty: A customer could belong to different clusters on different days due to the variances between days. Instead of individual days, previous research usually focuses on "typical days" or "month/season average" which artificially eliminates the uncertainty between days [7], [9], [11], [12]. The averaged load profiles [10] can be very different from individual ones, if they form non-convex sets.

This paper proposes a novel multi-resolution clustering (MRC) method to address these three key limitations. It is the first time to develop load profiles in the spectral domain, where volatile smart metering data are decomposed into regular spectral coefficients. Different features (e.g., overall trend and spikes) are separated onto different spectral levels and clustered separately without interferences. Further, features on different spectral levels will show different uncertainties over a period. Thus, it develops load profiling to a more granular level: individual customer and individual day. Sub-typical load profiles (Sub-TLPs) will firstly be developed on each spectral level, and then aggregated to form the final TLP. By this way, different permutations of sub-TLPs provide a more flexible load profiling with much less computation. By $\sum_{i=1}^{I} n_i$ levels of computation, MRC can express equivalent to $\prod_{i=1}^{I} n_i$ levels of load profiling, where n_i is the number of TLPs on the ith decomposition level. The method includes three main steps; each addresses one of the limitations as follows.

i) Decomposition: MRA extracts spectral features of load profiles, which will be used as input data for clustering. A small number of typical spectral features can accurately reconstruct massive load profiles in time domain.

ii) Clustering: Clustering based on MRA not only breaks down the computational burden, but also isolates different load features (e.g., magnitudes, overall trends and spikes) on different resolution levels. On each level, a sub-TLP is developed to represent the common load features.

iii) Reconstruction: Gaussian mixture model (GMM) is adopted to pre-cluster each customer over days in order to reduce sample size. An additive classification model based on the posterior probability is proposed to reflect the uncertainty between days. It can allocate the sampled load profile to the most-likely sub-TLP on each level. As different combinations of sub-TLPs can provide substantial flexibility, the proposed method will improve the load profiling accuracy whilst reducing the computational burden.

The rest of the paper is constructed as follows. Section II introduces the MRA, GMM and X-means techniques. Section III demonstrates the limitations of traditional clustering analysis in the time domain. Section IV proposes the MRC method and classification model. Section V implements the method on the Irish smart metered data. Results are demonstrated and compared with other clustering methods in Section VI. Conclusions are drawn in Section VII.

II. Multi-Resolution Analysis and Clustering Techniques

A. MRA

For power system, wavelet-based MRA has mainly been studied for short-term load forecasting (STLF) at system level [13]–[16]. It can decompose load profiles into components, each with different scales and shifts over time [13], [16]–[18].

As shown in Fig. 1, the load profile is decomposed by high-pass and low-pass filters. The coefficients of the filters are determined by the choice of mother wavelet. The down-sampling process breaks down original load profiles into lower resolution components. A higher level of decomposition process will generate lower resolution components. The large-scale components are called "approximations" (A) while small-scale components are called "details" (D). Through up-sampling and reconstruction filters, approximation component *A1* and detail component *D1* can be obtained. By this way, the original load profile can be represented by multi-resolution analysis (MRA) shown by (1):

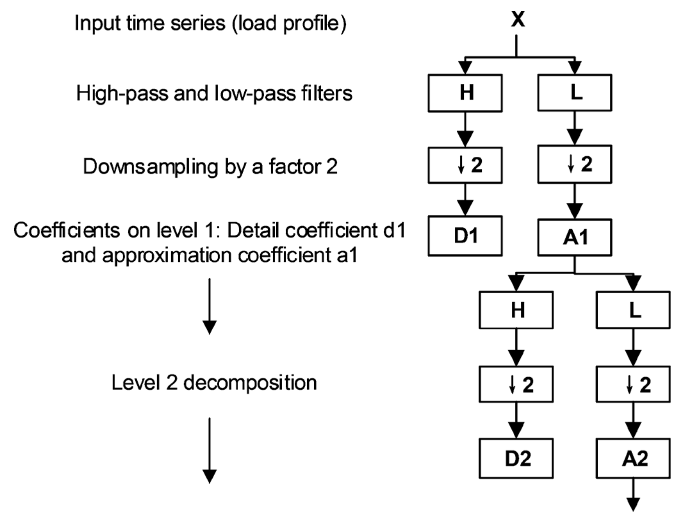

Fig. 1. Multi-resolution analysis by DWT.

TABLE I
Sample Sizes Across Different Decomposition Levels

Decomposition level	Number of coefficients	Interval Duration
Original data	48 samples per day	Half hour
Level 1	24	1 hour
Level 2	12	2 hours
Level 3	6	4 hours
Level 4	3	8 hours

$$X = D1 + A1 = D1 + D2 + A2 = \ldots = \sum_{j=1}^{J} D_j + A_J \quad (1)$$

where X is the reconstructed load profile; A_j and D_j are the approximation and detail components at level j, J is the total levels of decomposition.

This paper chooses a decomposition level of 3. This is chosen based on the sample rate of data and previous experiments [13]. As shown in Table I, a 3 level decomposition would give reasonable scale in components. However, the scale at level 4 is too large to represent most electricity usage.

Haar is chosen as the mother wavelet because it is likely to be coherent with the nature of the individual customer's load profile as the square wave can better portray the turn on-off of domestic appliances. Fig. 2 gives an example of using DWT to decompose an individual customer's load profile. Curves from top to bottom are original load profile, D1, D2, D3 and A respectively.

B. GMM

A mixture model develops a mixture probability density function (PDF) for observations. The mixture PDF is described as a

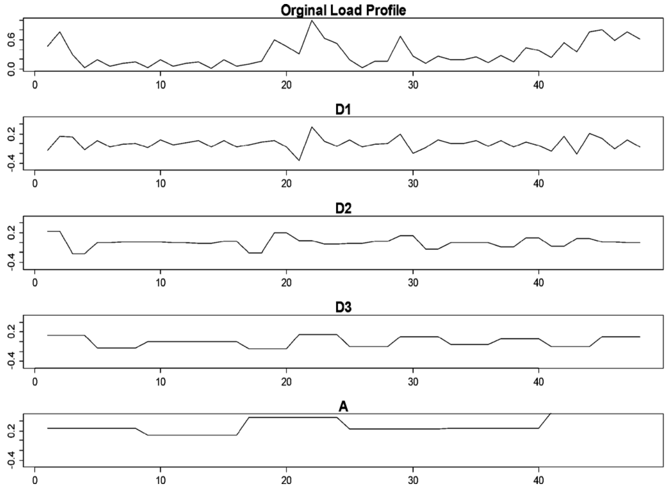

Fig. 2. Load profile decomposition by DWT (coefficients multiplied by basis wavelets).

weighted sum of a finite number of known components. It has been applied to areas such as clustering and probability distribution. In [7], GMM is used to represent the probability distribution function of the load at a bus. In this paper, it is used to cluster one customer (multiple days) into several within-customer typical components (WTCs) [19].

Normalization of the data is necessary to develop clusters that can reflect load pattern rather than magnitude. The j^{th} daily load profile $\boldsymbol{p_j}$ is normalized according to its daily peak $\max\{\boldsymbol{p_j}\}$ as in (2). In our case, $\boldsymbol{p_j}$is a vector with 48 variables.

$$\boldsymbol{s_j} = \frac{\boldsymbol{p_j}}{\max\{\boldsymbol{p_j}\}} \tag{2}$$

Suppose $\boldsymbol{s_j}$ is the jth load profile of a sample customer, and $\boldsymbol{s_j^{(l)}}$ is its multi-resolution analysis (MRA) component at the lth level. The dimension of $\boldsymbol{s_j^{(l)}}$ is $N(l)$. $f(\boldsymbol{s_j^{(l)}};\ \Psi)$ denotes the PDF of dependent variable $\boldsymbol{s_j^{(l)}}$. Vector Ψ denotes all the unknown parameters in the mixture model including weights and parameters of components, as $\Psi = \{\lambda_k, \theta_k\}_{k=1}^{K}$. A K-component finite mixture PDF can be written as (3), assuming component densities $f_k(\boldsymbol{s_j^{(l)}}\ ; \boldsymbol{\theta}_k)$ are specified to belong to some parametric family,

$$f(\boldsymbol{s_j^{(l)}}; \Psi) = \sum_{k=1}^{K} \lambda_k\, f_k\left(\boldsymbol{s_j^{(l)}}; \boldsymbol{\theta k}\right), \tag{3}$$

where K is the number of mixture components, and λ_k is the weight of the kth component. As the total probability of each component density as well as the mixture density equals to 1, the summation of all weights must be unity as in (4),

$$\sum_{k=1}^{K} \lambda_k\ = 1;\quad 0 \le \lambda_k \le 1\,. \tag{4}$$

In a Gaussian mixture model (GMM), $\boldsymbol{\theta}_k$ can be represented by: $\boldsymbol{\theta_k} = (\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})$ $(k = 1, 2, \ldots K)$ for the multivariate case, where $\boldsymbol{\mu_k}$ is a vector and $\boldsymbol{\Sigma_k}$ is a matrix. The component PDF in (3) can be written as a normal distribution in (5), shown at the bottom of the page.

The parameters λ_k, μ_k and Σ_k can be estimated by Expectation Maximization (EM). The solution is introduced in Appendix and detailed algorithm can be found in [20].

C. X-Means

X-Means clustering is very similar to conventional K-means clustering. Instead of using a pre-defined number of clusters K, X-means will search in a range of different values of K, and determine the optimum K based on a model selection criterion such as Bayesian Information Criterion (BIC). The detailed algorithm is introduced in [21]. The strategy is as following. Firstly, starting from initial K, the K-means clustering is applied to give deterministic centroids and its BIC value is calculated. Secondly, a new centroid is introduced by splitting some existing centroids into two. This is achieved by a local K-means $(K = 2)$ within the centroid subset. Whether the split is meaningful is determined by local BIC improvement. The process is iterated till K meets its upper bound. The selection of K is determined by their BIC scores. The BIC brings a penalty term for the number of parameters in the model in order to assess the trade-off between likelihood and number of clusters. It is expressed in (6), shown at the bottom of the page, to find the optimum number of clusters [22], where $K \times N(l)$ is the number of parameters, for K clusters, and $N(l)$ is the dimension of $\boldsymbol{s_j^{(l)}}$. J is the sample size.

III. TRADITIONAL CLUSTERING TECHNIQUES IN THE TIME DOMAIN

The general steps of traditional load profiling are shown in Fig. 3. The input data of clustering techniques are usually time-series load profiles. The techniques aim to group M load profiles

$$f_k\left(\boldsymbol{s_j^{(l)}}; \theta_k\right) = \frac{\det(\Sigma_k)^{-\frac{1}{2}}}{(2\pi)^{\frac{N}{2}}}\ \exp\left[\frac{-(\boldsymbol{s_j^{(l)}} - \boldsymbol{\mu_k})^{\mathrm{T}}\boldsymbol{\Sigma_k}^{-1}(\boldsymbol{s_j^{(l)}} - \boldsymbol{\mu_k})}{2}\right]. \tag{5}$$

$$L_{BIC}\left(\boldsymbol{s_1^{(l)}} \ldots \boldsymbol{s_J^{(l)}}\right) = -2\ln L\left(\boldsymbol{s_1^{(l)}} \ldots \boldsymbol{s_J^{(l)}}\right) + J\ln(K \times N(l)), \tag{6}$$

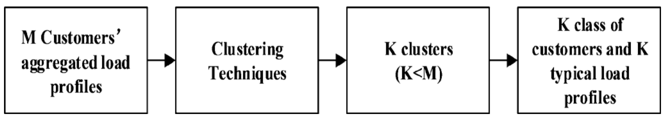

Fig. 3. Conventional load profile clustering process.

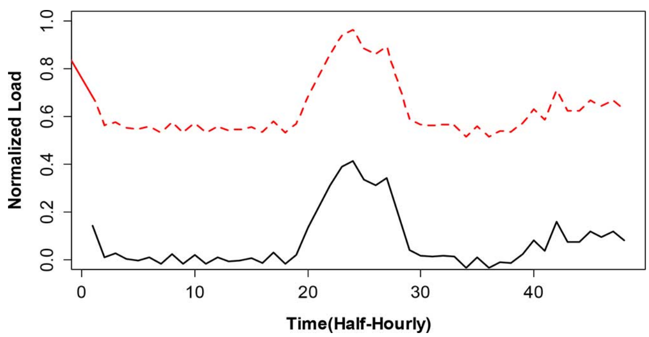

Fig. 4. Problems with time-series clustering: magnitude difference within clusters (two load profiles from the same cluster with similar shape but different magnitudes).

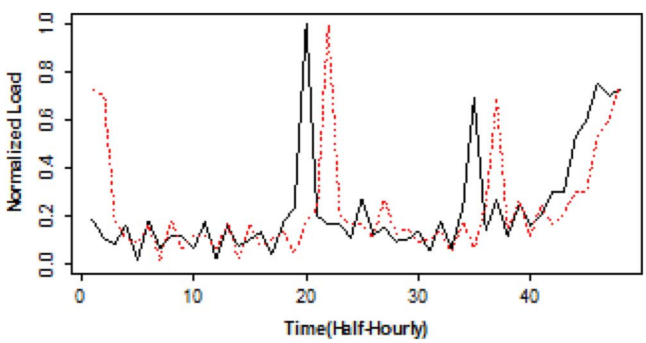

Fig. 5. Problems with time-series clustering: time difference in spikes (two similar load profiles are assigned to different clusters due to time delay).

into K clusters. The partition is based on metrics between load profiles (e.g., Euclidean distance).

The time-series clustering shows limitations in handling magnitude, volatility and uncertainty, which have been explained in the Introduction. Fig. 4 depicts an example of the magnitude difference within groups. They are from real load data clustered by K-means. As the input load data are normalized according to its daily peak in the clustering process, the clustering process is entirely based on the similarity of shape. Two load profiles in Fig. 4 are clustered into the same group due to similar load shapes. However, when considering the factor of load magnitude (e.g., normalized by maximum load over study period as in Fig. 4), there is a substantial difference between their original magnitudes. Therefore, the TLP of this cluster can only represents load shape but not magnitudes.

Due to the nature of time-series metrics, volatility has a severe impact on clustering. For instance, two very similar load profiles are clustered into different groups by time-series clustering in Fig. 5. Due to the slight delay and significant volatility of two load profiles, the peak of one load profile keeps meeting the trough of the other one at each sample point, which dramatically increases the distance between two load profiles.

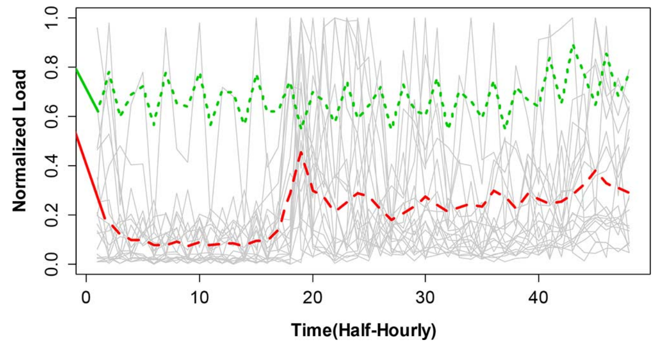

Fig. 6. Problems with time-series clustering: uncertainties between days (a customer's load profiles on different days (grey), with substantial variation (green and red load profiles)).

The key challenge of developing load profiles of individual customers on individual days is the uncertainty of the customer between days, which means the same customer may have very different load profiles between days. Most studies are specified for a season or day type (e.g., weekday or weekend).

However, if the daily load profiles form a non-convex set, the averaged load profiles may be out of the set, which will not only lose detailed information, but may also cause mis-classification. In Fig. 6, the grey lines are load profiles of the same customer on different days. A clear variation can be seen between days. There are two different TLPs (red and green) over those days, making classification more difficult.

IV. Proposed Multi-Resolution Clustering

A novel MRC method is proposed in this paper to cluster daily load profiles in the spectral domain. Multi-resolution analysis based on the wavelet transform can decompose a volatile and irregular load profile into a smooth large scale component describing the underlying shape, and several small scale components describing volatilities.

The methodology of MRC is illustrated by the flow chart in Fig. 7. It consists of four main stages:

i) Decomposition: A volatile load profile can be decomposed into several components including magnitude and several spectral components: approximation (A) component, and detail (from *D1* to *DI*) components. Thus different features are separated onto various spectrum components, which prevent interference between underlying trend and volatile elements. The basic idea of MRC is to implement clustering analysis on each of the spectrum components separately, and develop a typical component (TC) for each cluster at each level (e.g., *A-1* indicates cluster 1 on the Approximation level). Accordingly, in a classification process, a customer's daily load profile will have a cluster membership on every spectrum component level.

ii) Clustering over days: a more detailed flowchart of two-stage clustering is shown in Fig. 8. It uses GMM to pre-cluster each customer through multiple days so that one customer's daily profiles can be represented by several typical models. Then X-means clustering will be performed only on these models. By the two-stage clustering, large amounts of input data are reduced into

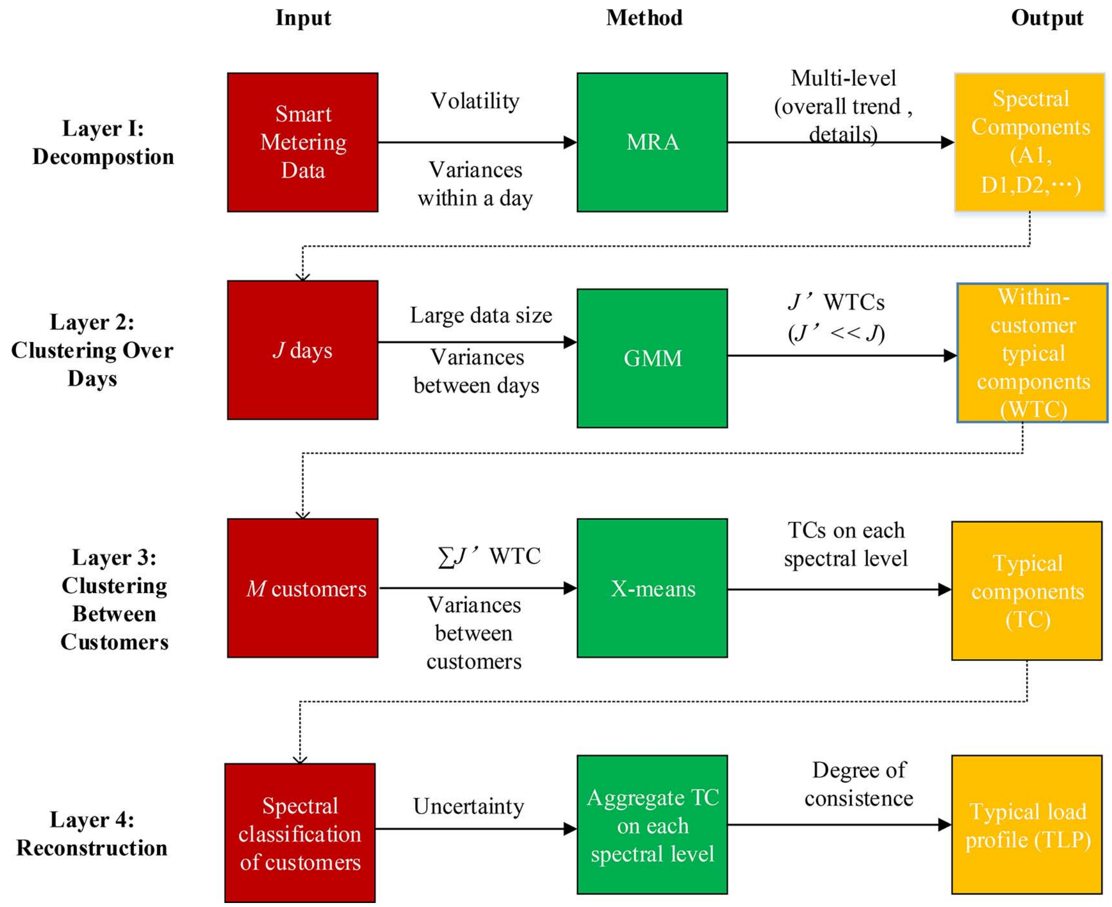

Fig. 7. Overall methodology of multi-resolution clustering.

a small number of PDFs while original information is maintained in the models. For example, in layer 2, for level (l) of components (e.g., component A), instead of clustering all customers all days, a pre-clustering is conducted on each customer. For the mth customer with J days' load profiles $\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}, j = 1 \ldots J$, the GMM is used to cluster them into K_m models. Each model will be represented by a within-customer typical component (WTC), which is defined as the daily load profile with lowest uncertainty (highest posterior probability) in the cluster. The uncertainty is defined as 1 minus the maximum posterior probability. For the number of WTCs, i.e., number of clusters, starting with $K_m = 1$, the aim is to find the least number of K_m while keeping the uncertainty of every day below threshold β. It ensures every daily component is sufficiently close to the model centre. Thus, J daily components can be reduced and represented by K_m within-customer WTCs as shown in (7)

$$WTC_k^m = \arg\max_{\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}} (P(k|\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})})),\ k = 1, 2, \ldots K_m, \quad (7)$$

where WTC_k^m is the WTC of cluster k, and $P(k|\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})})$ is the posterior probability of cluster k given sample $\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}$, which will be explained in V.

iii) Clustering between customers: in this stage, the WTCs of each customer are used as the input data of clustering. A total number of $\sum_{m=1}^{M} K_m$ WTCs are clustered by X-Means clustering. The outputs are the centres of each cluster, which are regarded as typical components (TCs). At this stage, the input data have already been processed with data reduction, features isolation and uncertainties modelling; therefore, most clustering techniques could theoretically deliver decent clustering performance. However, as the number of W can still be large especially for small-scale components, the main considerations of choosing clustering techniques at this stage are: i) low computation complexity while handling large sample size; ii) requiring no pre-knowledge on the number of clusters as the range could be very large. X-Means clustering, as an extended K-Means, is chosen because it inherits the simplicity of K-Means while automatically searching for an optimal number of clusters based on Bayesian Information Criterion (BIC).

iv) Reconstruction: The TLP is developed by aggregating assigned TCs as shown in (8),

$$TLP = TC_{mag} \times \left(TC_A + \sum_{i=1}^{I} TC_{Di} \right), \quad (8)$$

where I is the total number of detail levels. The synthesis of TCs from MRA (A to DI levels) provides the shape of the load

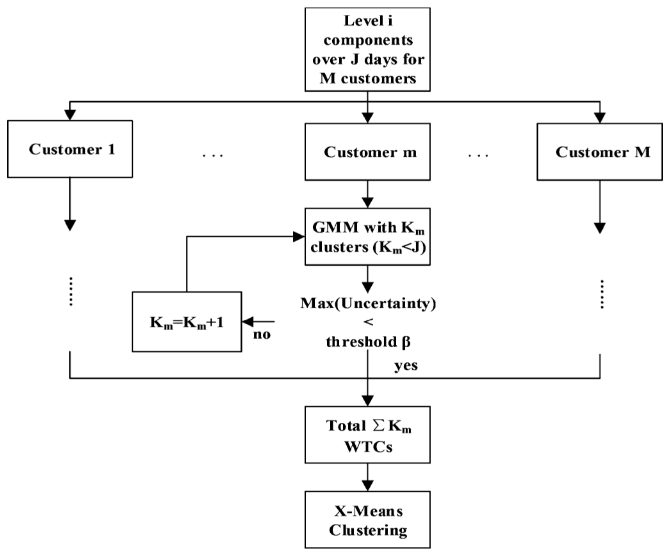

Fig. 8. Two-stage GMM and X-means clustering implemented in MRC.

profile, and the magnitude TC will scale up the shape to the typical loading level.

The obvious advantage is that each clustering will focus on one feature without interference from the others. Other improvements are as follows. i) Input data size is substantially reduced by getting rid of those spectral coefficients which are close to zero so as to reduce the computation of clustering. However, after the clustering process, the original coefficients are used to reconstruct the TCs. ii) Load magnitude and shape are separately clustered, but jointly integrated in TLPs. iii) The problem caused by volatility can be resolved as overall trends and spikes are separated. iv) Flexibility for the number of clusters: In traditional clustering methods, due to the uncertainties, the load profiles between days can be very different, which requires a huge number of clusters to express. MRC provides an opportunity to have different cluster numbers on different levels. For example, a few clusters may be sufficient for approximation (A) level as the overall trends of customers are likely to be similar and stable over days while detail levels may require more clusters to distinguish random spikes.

There are generally three types of classical clustering analysis: connectivity-based, centroid-based and distribution-based. In the analysis of smart metering data, our proposed spectral-based clustering is superior to classical methods in the following ways:

i) Connected-based clustering such as hierarchical and nearest neighbor classifiers require distance matrix. However, large sample size of smart meters will create immense computational burden on this. In our case, 6369 customers daily load profiles over one year produces over 2.2 million observations, which needs a huge memory to build up the distance matrix.
ii) Smart metering data are volatile and irregular. For centroid-based method such as K-means, it inevitably leads to large variances within cluster or excessive number of centroids (clusters).
iii) GMM is embedded in our method to compress multiple days' load profiles into several models. However, distribution-based method on its own may encounter similar issues which centroid-based ones face.

V. Classification

The classification process is to determine the cluster membership of a customer. For new customers, sometimes their smart metering data are limited and thus the classification is performed based on their eco-social information. In this paper, under the smart meter scenario, classification is based on historical sampled load data.

As the GMM pre-clustering stage groups a customer's load profiles over days into several Gaussian models (the centres are the WTCs), the second stage is actually to cluster these WTCs (centroid of the models). After X-means clustering, all these models are clustered into Q new clusters. Each cluster $\omega_q (q = 1 \ldots Q)$ can be treated as a new mixture model, made up of different Gaussian distributions (WTCs) from the first stage. The parameters of individual PDFs will not change but the weight in the new mixture model requires a re-normalization to ensure the unity sum. For cluster ω_q, containing $n(q)$ WTCs (Gaussian models), the new weights are calculated as (9):

$$\lambda_k^q = \frac{\lambda_k}{\sum\limits_{r=1}^{n(q)} \lambda_r}, \; k = 1, 2 \ldots n(q) \, , \tag{9}$$

where λ_k is the previous weight; λ_k^q is the new weight of each component in cluster ω_q.

As each cluster ω_q is expressed as a mixture model now with new weightsλ_k^q, the unlabelled sampled data $\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}$ can be classified based on its posterior probability. Assuming equal prior probabilities for each cluster, the likelihood of $\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}$ belonging to cluster $\omega_q$is the weighted sum of the likelihoods of $\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}$ belonging to each WTC in cluster ω_q as shown in (10):

$$p\left(\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})} | \omega_q\right) = \sum_{k=1}^{n(q)} \lambda_k^q \, p_k \left(\boldsymbol{s}_{\mathbf{j}}^{(\boldsymbol{l})} | \omega_q\right) \, , \tag{10}$$

where $pk(\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})} | \omega_q)$ is a Gaussian function as in (5). The posterior probability can be obtained by (11):

$$P\left(\omega_q | \boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}\right) = \frac{p\left(\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})} | \omega_q\right)}{\sum\limits_{r=1}^{Q} p\left(\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})} | \omega r\right)}, \tag{11}$$

where $P(\omega_q | \boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})})$ is the posterior probability of cluster ω_q given sample $\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}$.

In summary, the sample load data of a new unlabelled customer will firstly be decomposed by the same procedure. Each daily component will be assessed by the posterior probability of each cluster ω_q. It will be allocated to the one with highest posterior probability as in (12),

$$\hat{\omega}(\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}) = \underset{\omega_q}{\arg} \max(P(\omega_q | \boldsymbol{s}_{\mathbf{j}}^{(\boldsymbol{l})})) \, , \tag{12}$$

TABLE II
NUMBER OF CLUSTERS OF EACH GROUP AND DECOMPOSITION LEVEL

	A	*D1*	*D2*	*D3*
Residential	20	11	8	10
SME	19	8	13	8
Other	15	11	8	9

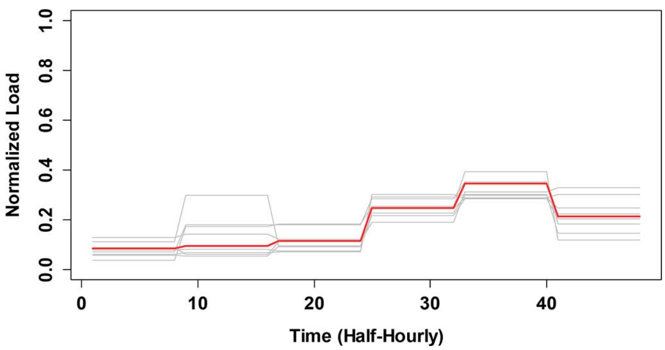

Fig. 9. TC (red) and members (grey) of cluster 1 on A level (normalized daily load components).

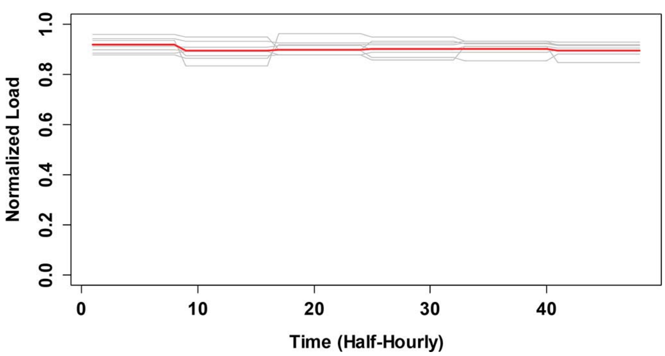

Fig. 10. TC (red) and members (grey) of cluster 2 on A level (normalized daily load components).

where $\hat{\omega}(\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})})$ is the final classification of $\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}$.

In our case, the classification would be on 5 levels, i.e., magnitude (M), (A), $(D1)$, $(D2)$, and $(D3)$: $\hat{\omega}(\boldsymbol{s}_{\boldsymbol{j}}^{(M)}) - \hat{\omega}(\boldsymbol{s}_{\boldsymbol{j}}^{(A)}) - \hat{\omega}(\boldsymbol{s}_{\boldsymbol{j}}^{(D1)}) - \hat{\omega}(\boldsymbol{s}_{\boldsymbol{j}}^{(D2)}) - \hat{\omega}(\boldsymbol{s}_{\boldsymbol{j}}^{(D3)})$. If a customer's is classified as 4-2-7-10-3 (60%), it indicates the magnitude is assigned to cluster 4 of magnitude TC, component A is assigned to cluster 2 (A) TC and so on. The probability 60% is calculated as the number of days belonging to 4-2-7-10-3 over the total number of sample days. It indicates the degree of consistency of classification over days.

VI. RESULTS

A. Typical Components

All 6369 customers are firstly macro-classified into three groups: i) residential, ii) small and medium enterprise (SME), and iii) others. The method is applied to each group separately. For each group, there are 4 decomposition levels and several clusters as shown in Table II.

Figs. 9 to 14 show some of the typical components (TCs) as examples. The TCs are from residential customers at the A level, $D2$ level and $D1$ level.

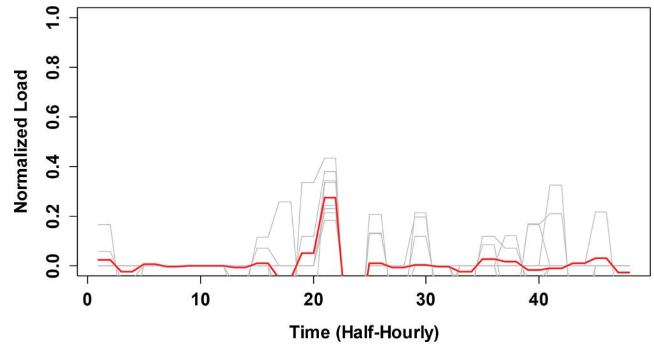

Fig. 11. TC (red) and members (grey) of cluster 1 on $D2$ level (normalized daily load components).

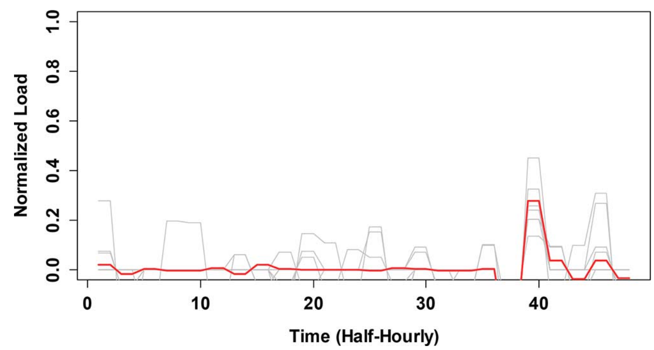

Fig. 12. TC (red) and members (grey) of cluster 2 on $D2$ level (normalized daily load components).

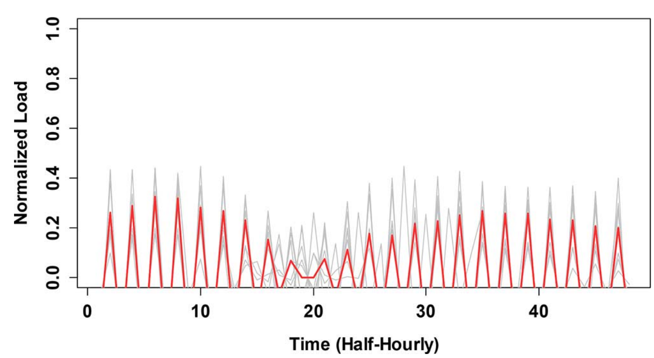

Fig. 13. TC (red) and members (grey) of cluster 1 on $D1$ level (normalized daily load components).

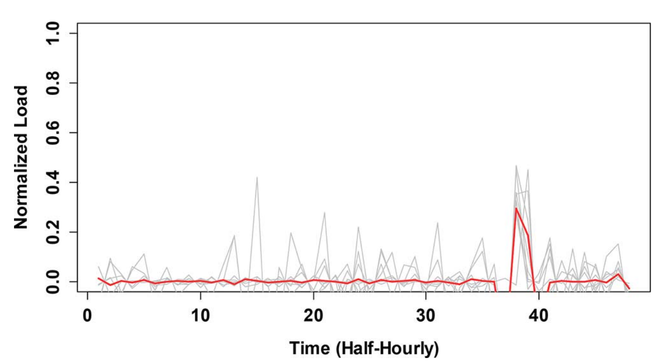

Fig. 14. TC (red) and members (grey) of cluster 2 on $D1$ level (normalized daily load components).

Figs. 9 and 10 are two clusters from the A level. Grey lines are some sampled daily components from member customers. The red line is the TC of the cluster, which is derived from the average within the cluster. It is seen that the TC in Fig. 9 shows a typical working class household with low loading level in the day time and peak in the evening. The TC in Fig. 10 however, has a high loading level consistently through a day. It represents a household with full-day occupancy.

Figs. 11 and 12 are two clusters from the *D2* level. The scale is getting smaller and there is more variation within the cluster. The components at this level are likely to represent more random activities and short-interval usage (e.g., kettles). Fig. 11 sees few activities from 1 a.m. to 8 a.m. (sleeping time) and frequent activities from 9 a.m. The cluster in Fig. 11 has similar patterns while activities are more concentrated around 8 p.m. with a consistent peak.

Figs. 13 and 14 are two clusters from the *D1* level, which has the smallest scale. They contain more unpredictable spikes which are possibly caused by the turn-on of some appliances. The cluster in Fig. 13 has very consistent periodical spikes especially during night and noon, which are probably from stand-by white goods such as refrigerators. The cluster in Fig. 14 shows no periodical loads, but there are congested high peaks in the morning and evening due to human activities.

B. Comparisons

The comparison is between three types of load profiles: MRC TLPs, time-series K-means clustering and TLPs used by the U.K. industry. It is conducted by assessing the similarities between the three types of TLPs and smart metering data on random days.

In the U.K., due to the absence of smart meters on every customer in the market, small customers (below 100 kW maximum demands) are pre-classified into 8 classes for electricity settlement. Each class of customers are represented by a TLP, which has been widely used by the U.K. industry for decades. The classification is generally based on the nature of customer, such as residential, commercial and industrial. Residential and commercial customers are further classified by tariff types while industrial customers are further differentiated by load factors.

The same smart metering data used in our proposed method are also processed by K-means clustering in the time domain. The clustering is based on individual load profiles. Each customer each day is assigned with a deterministic cluster and TLP.

The classification is based on some sampled data from the investigated customer through the method in IV. As an example, customer 1076 over a month is classified as in Table III. In summary, customer 1076 belongs to i) class 1 in the industry typical load profiling; ii) cluster 6 in time-series K-means clustering $(K = 30)$; iii) 7-5-7-6 in MRC (total number of clusters: 10-5-8-7). We deliberately choose $K = 30$ for K-means, which is the sum of the total cluster number in MRC $(10+5+8+7)$, to give a fair comparison between the two methods. Because the industry TLPs represent the average magnitude over millions of customers, and the K-means is based on normalized shape, it is difficult to compare different load profiling methods in terms of magnitudes. The comparison on this paper is based on normalized load.

TABLE III
CLASSIFICATION OF SAMPLED CUSTOMER BY DIFFERENT LOAD PROFILING METHODS

Customer	Date	UK TLP	K-means	MRC (*A-D1-D2-D3*)
1076	26/Aug/2009	1	6	7-5-6-7

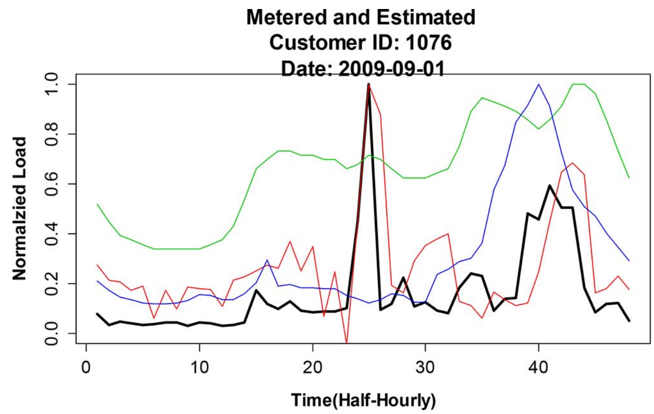

Fig. 15. Comparison between smart metering data of customer 1076 on 26/08/2009 (black) and three load profiling methods: UK TLP (green), K-means TLP (blue), MRC TLP (red).

TABLE IV
COMPARISON BETWEEN SMART METERING LOAD PROFILES AND THREE LOAD PROFILING METHODS (SAMPLE Size = 2994)

Load Profiling Methods	MME (per unit)	MAPE (per unit)	PTE (hour)
UK TLP	0.82	0.45	6.4
K-means (K=30)	0.66	0.23	4.0
MRC(10-5-8-7)	0.58	0.17	2.8

Still, taking customer 1076 as an example, Fig. 15 shows the comparison between smart metering data and three load profiles. The black line is the real metered data of the customer on the day. The green line is the TLP class 1 from industry. It is unable to express the real daily load profile. The blue line is the TLP from cluster 6 in time-series K-means clustering. Although following the base load well, it cannot capture the spike at noon. The proposed MRC TLP, depicted by the red line, expresses the daily energy usage more accurately in terms of both overall trend and spikes.

The load profiles used by the power industry are based on seasonal average load profiles within a group of pre-defined customer class. It is only an approximation of group customers over the long term. Moreover, there are only 2 classes describing residential customers, i.e., 2 clusters. Time-series analysis, due to issues such as volatility and uncertainty, is more feasible for analysing average load profiles over time. However, the proposed MRC successfully addresses these issues by separating different load features on different decomposition levels.

An extensive assessment of the three load profiling methods is conducted over 2994 smart-metering load profiles. The representativeness of the load profiling methods are evaluated by the following indices: Maximum Magnitude Error (MME), Mean Absolute Percentage Error (MAPE), Peak Time Error (PTE). They are widely used in the relative field so as to give a fair comparison [23]–[25]. The comparison in Table IV shows clear improvements of MRC over the other two methods. PTE has a huge decrease to 2.8 hours. It shows the capability of MRC on capturing the volatilities. MME and MAPE also decrease compared to K-means. The reason is that different permutations of sub-TLPs provide a more flexible load profiling with much less computation. By $\sum_{i=1}^{I} ni$ levels of computation, MRC can express equivalent to $\prod_{i=1}^{I} ni$levels of load profiling, where n_i is the number of TLPs on the ith decomposition level, with a total of I levels.

VII. Conclusion

This paper proposes a novel multi-resolution clustering method. It successfully develops a spectral-domain load profiling for smart metering data. It is specifically designed for smart metering data to overcome the problems for traditional time-series analysis.

i) Multi-resolution analysis based on wavelet analysis decomposes load profiles into different features, addressing the volatility in smart metering data.
ii) A two-stage MRC technique is proposed to cope with uncertainty and reduce clustering input size.
iii) Different permutations of sub-TLPs provide a more flexible load profiling with much less computation.

The results show significant improvements over traditional time-series analysis, providing more accurate load profiling at a more granular level. The proposed MRC offers a promising approach for using smart metering data to develop a smart grid and enhance power system efficiency. Future work will focus on relating customer information and external factors to load profiling. The research will aim to find the customers' behaviour modes behind load features. This research can then be used for more accurate demand forecasting, supporting efficient use of DSR and enhancing settlement (supplier) efficiency.

Appendix
Expectation Maximization

For a given J observations, the log-likelihood function can be expressed by:

$$\ln L\left(\boldsymbol{s}_{\boldsymbol{1}}^{(\boldsymbol{l})} \ldots \boldsymbol{s}_{\boldsymbol{J}}^{(\boldsymbol{l})}\right) = \sum_{j=1}^{J} \ln\left(f\left(\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}; \Psi\right)\right). \tag{13}$$

Parameters Ψ can be estimated by maximizing (13) using the EM algorithm. In the E stage, the parameters are estimated, and then each observation can be assigned to each mixture component k by the highest posterior probability (i.e., Bayes rule). The posterior probability of observation j belonging to mixture component k is given by (14),

$$P\left(k|\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}\right) = \frac{\lambda_k fk\left(\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}; \theta_k\right)}{\sum_{r=1}^{K} \lambda r fr\left(\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}; \theta r\right)}. \tag{14}$$

In the M stage, the parameters are re-estimated by maximizing (13) again under new posterior probabilities. λ_k, μ_k and Σ_k are obtained by (15), (16) and (17) respectively.

$$\overline{\lambda_k} = \frac{1}{J}\sum_{j=1}^{J} P\left(k|\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}\right) \tag{15}$$

$$\mu_k = \frac{\sum_{j=1}^{J} \boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})} \times P(k|\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})})}{\sum_{j=1}^{J} P(k|\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})})} \tag{16}$$

$$\Sigma_k = \frac{\sum_{j=1}^{J} P\left(k|\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}\right)\left(\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})} - \boldsymbol{\mu_k}\right)\left(\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})} - \boldsymbol{\mu_k}\right)^{\mathrm{T}}}{\sum_{j=1}^{J} P\left(k|\boldsymbol{s}_{\boldsymbol{j}}^{(\boldsymbol{l})}\right)} \tag{17}$$

The EM algorithm can maximize (13) by following iteration steps [20]:

1. Before the first iteration, s = 0, initializing the number of clusters K (J/K > I), and a starting partition. This can be achieved by random partition or clustering techniques.
2. Given any partition, parameters can be estimated by (15)–(17) to maximize (13) under.
3. Once parameters from s iteration are estimated, each observation can be re-assigned to each cluster by new posterior probabilities.
4. Repeat step 2 and 3 until converge:$\left|\max(\ln L)^{s+1} - \max(\ln L)^{s}\right|$ is smaller than stopping criterion.

References

[1] "2050 Pathways Analysis," DECC. London, U.K., Jul. 2010.
[2] "UK Renewable Energy Roadmap," DECC. London, U.K., Jul. 2011.
[3] "Estimated impacts of energy and climate change policies on energy prices and bills," D. o. E. C. Change, Ed., ed 2010.
[4] W. Zhimin, G. Chenghong, L. Furong, P. Bale, and S. Hongbin, "Active demand response using shared energy storage for household energy management," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1888–1897, Dec. 2013.
[5] S. Valero, M. Ortiz, C. Senabre, C. Alvarez, F. J. G. Franco, and A. Gabaldon, "Methods for customer and demand response policies selection in new electricity markets," *IET Gener., Transm., Distrib.*, vol. 1, pp. 104–110, 2007.
[6] D. Müllner, "Fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python," *J. Statist. Softw.*, vol. 53, no. 9, pp. 1–18, 2013.
[7] G. J. Tsekouras, N. D. Hatziargyriou, and E. N. Dialynas, "Two-stage pattern recognition of load curves for classification of electricity customers," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1120–1128, Aug. 2007.
[8] G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 933–940, May 2006.
[9] Z. Tiefeng, Z. Guangquan, L. Jie, F. Xiaopu, and Y. Wanchun, "A new index and classification approach for load pattern analysis of large electricity customers," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 153–160, Feb. 2012.
[10] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "Allocation of the load profiles to consumers using probabilistic neural networks," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 548–555, May 2005.
[11] S. V. Verdu, M. O. Garcia, C. Senabre, A. G. Marin, and F. J. G. Franco, "Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1672–1682, Nov. 2006.

[12] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, "Load pattern-based classification of electricity customers," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 1232–1239, May 2004.

[13] A. J. R. Reis and A. P. A. da Silva, "Feature extraction via multiresolution analysis for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 189–198, Feb. 2005.

[14] Z. A. Bashir and M. E. El-Hawary, "Applying wavelets to short-term load forecasting using PSO-based neural networks," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 20–27, Feb. 2009.

[15] A. S. Pandey, D. Singh, and S. K. Sinha, "Intelligent hybrid wavelet models for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1266–1273, Aug. 2010.

[16] C. Ying *et al.*, "Short-term load forecasting: Similar day-based wavelet neural networks," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 322–330, Feb. 2010.

[17] A. M. M. Kociołek, M. Strzelecki, and P. Szczypiński, "Discrete wavelet transform-Derived features for digital image texture analysis," in *Proc. Int. Conf. Signals and Electronic Systems*, Lodz, Poland, 2001, pp. 163–168.

[18] M. P. Tcheou *et al.*, "The compression of electric signal waveforms for smart grids: State of the art and future trends," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 291–302, Jan. 2014.

[19] G. McLachlan and D. Peel, *Finite Mixture Models*, 2004 [Online]. Available: http://www.wiley.com.

[20] W. S. DeSarbo and W. L. Cron, "A maximum likelihood methodology for clusterwise linear regression," *J. Classification*, vol. 5, pp. 249–282, 1988.

[21] K. Imamura, N. Kubo, and H. Hashimoto, "Automatic moving object extraction using x-means clustering," in *Proc. 2010 Picture Coding Symp. (PCS)*, 2010, pp. 246–249.

[22] P. Chaussé, "Computing generalized method of moments and generalized empirical likelihood with R," *J. Statist. Softw.*, vol. 34, no. 1, pp. 1–35, 2010.

[23] Z. Shiyin and K. Tam, "A frequency domain approach to characterize and analyze load profiles," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 857–865, May 2012.

[24] B. F. Hobbs, S. Jitprapaikulsarn, S. Konda, V. Chankong, K. A. Loparo, and D. J. Maratukulam, "Analysis of the value for unit commitment of improved load forecasts," *IEEE Trans. Power Syst.*, vol. 14, no. 4, pp. 1342–1348, Nov. 1999.

[25] P. Mandal, T. Senjyu, N. Urasaki, T. Funabashi, and A. K. Srivastava, "A novel approach to forecast electricity price for PJM using neural network and similar days method," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2058–2065, Nov. 2007.

Ran Li received his B.Eng. degree in electrical power engineering from University of Bath, U.K., and North China Electric Power University, Beijing, China, in 2011. He received the Ph.D. degree from University of Bath, in 2014 and became a lecturer in Bath in 2015. His major interest is in the area of big data in power system, deep learning and power economics.

Furong Li (SM'09) was born in Shannxi province, China. She received the B.Eng. degree in electrical engineering from Hohai University, Nanjing, China, in 1990 and the Ph.D. degree from Liverpool John Moores University, Liverpool, U.K., in 1997. She is a professor in the Power and Energy Systems Group, University of Bath. Her major research interest is in the area of power system planning, analysis, and power system economics.

Nathan D. Smith is a lecturer in the Department of Electronic and Electrical Engineering, University of Bath, with interests which include pattern processing/classification, inverse problems, and data assimilation.