ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2018)
67, *Part* 4, *pp.* 897–915

The functional latent block model for the co-clustering of electricity consumption curves

Charles Bouveyron,

Institut National de Recherche en Informatique et en Automatique Sophia Antipolis and Université Côte d'Azur, Nice, France

Laurent Bozzi,

Électricité de France, Paris-Saclay, France

Julien Jacques

Université de Lyon, France

and François-Xavier Jollois

Université Paris-Descartes, France

[Received June 2017. Revised December 2017]

Summary. As a consequence of recent policies for smart meter development, electricity operators nowadays can collect data on electricity consumption widely and with a high frequency. This is in particular so in France where the leading electricity company Électricité de France will be able soon to record the consumption of its 27 million clients remotely every 30 min. We propose in this work a new co-clustering methodology, based on the functional latent block model (LBM), which enables us to build 'summaries' of these large consumption data through co-clustering. The functional LBM extends the usual LBM to the functional case by assuming that the curves of one block live in a low dimensional functional subspace. Thus, the functional LBM can model and cluster large data sets with high frequency curves. A stochastic expectation–maximization–Gibbs algorithm is proposed for model inference. An integrated information likelihood criterion is also derived to address the problem of choosing the number of row and column groups. Numerical experiments on simulated and original Linky data show the usefulness of the methodology proposed.

Keywords: Electricity consumption curves; Latent block model; Model-based co-clustering; Multivariate functional data

1. Introduction

Nowadays, electricity meters are mostly electromechanical meters. They measure consumption and require a technician if a change in power or an outage occurs. Linky is a communicating meter, which means that it can receive and send data without the need for the physical presence of a technician. Installed in end consumer properties and linked to a supervision centre, it is in constant interaction with the electricity network. After the installation of 300 000 Linky smart meters between 2009 and 2011 in the area of Lyon and Tours (France), the French authorities

Address for correspondence: Charles Bouveyron, Laboratoire J. A. Dieudonné, Université Nice Sophia Antipolis, Parc Valrose, 06108 Nice Cedex 02, France.
E-mail: charles.bouveyron@math.cnrs.fr

 0035–9254/18/67897

decided to generalize these meters throughout the territory. By 2021, 35 million meters should be replaced in French households by Linky meters, allowing electricity operators to record electricity consumption remotely. For an operator like Électricité de France (EDF) with 27 million residential dwellings, these new smart meters represent a great opportunity to gather customer consumption data and therefore to improve client knowledge. Indeed, until now, customer data have been recorded only every 6 months, whereas, with the smart meter, the data can be taken up to every second. In practice, EDF plans to access the data every half-hour, which means 17472 measures per year for each of the 27 million customers. Nevertheless, this data flood may also be a drawback since they represent a mass of data to store and manage. For this, it will be necessary to build meaningful 'summaries' of these data, and one way to achieve that is though co-clustering.

Clustering of the time series corresponding to customer consumptions, also called *functional data* (Ramsay and Silverman, 2005), can be performed by using functional data clustering techniques (see Jacques and Preda (2014) for a survey). With such approaches, a whole set of customers can be summarized as a small number of clusters. Nevertheless, the interpretation of these clusters, using for instance the mean consumption curves, is difficult because of the long period of observation (several months or even years). To provide a synthetic summary of the consumption data, it has been decided to cut the period of observation into small units of times. Taking into account EDF's expectations in term of interpretation, a daily unit has been selected: thus, 365 daily curves consumption are observed per year for each of the 27 million customers. If for each cluster the interpretation of one mean daily consumption curve is feasible (and meaningful for EDF), it is still not possible to interpret them for all days. There is also a need to summarize the days of observations as a small number of clusters. Consequently, the analysis of the data provided by the Linky meters needs to build both clusters of customers and clusters of days of observation. From a statistical point of view, we are faced with a problem of clustering both the individuals (customers) and the features (days of observation), which is know in the literature as a *co-clustering* problem.

In the context of data recorded in a table where rows index individuals and columns index features, co-clustering techniques aim to cluster individuals and features simultaneously as homogeneous sets. Thus, the large data matrix can be summarized by a reduced number of blocks of data (or co-clusters). If the earliest (and most cited) method is probably due to Hartigan (1972), model-based approaches have recently proved their efficiency either for continuous, binary, categorical or contingency data (Govaert and Nadif, 2013; Jacques and Biernacki, 2017). Those approaches rely on the latent block model (LBM) (Govaert and Nadif, 2013), which tackles combinatorial issues by assuming local independence, i.e. all the random variables representing the cells of the data table are independent once the row and column partitions have been fixed.

The originality of the present work is that the objects which must be co-clustered are functional data (electricity consumption curves). To the best of our knowledge, the only work dedicated to the co-clustering of such data is Ben Slimen *et al.* (2016), which proposed a co-clustering for functional data based on a two-steps approach: first, functional principal component analysis (PCA) (Ramsay and Silverman, 2005) is carried out on the whole set of curves; second a Gaussian LBM is applied to the first functional PCA scores. As model-based approaches have recently improved the two-steps approches in the clustering context (see Jacques and Preda (2014)), we propose in this work a functional LBM to improve the modelling capacity of the model developed in Ben Slimen *et al.* (2016). The advantages of the functional LBM that is developed in this work are the following: first, the whole set of functional PCA scores are modelled and not just the first as in Ben Slimen *et al.* (2016); second, the parameterization remains parsimonious since

the data are assumed to live in block-specific functional subspaces; finally, the functional PCAs are carried out block by block, enabling detection of detailed phenomena in the data structure. In addition, although clustering techniques have already been considered for the analysis of electricity consumption curves (Abreu *et al.*, 2012; Keyno *et al.*, 2009; Tsekouras *et al.*, 2007; Melzi *et al.*, 2017), this work proposes the first use of a functional co-clustering technique to provide both a segmentation of households (rows) and days (columns) in interpretable groups for an electricity operator such as EDF.

The paper is organized as follows. Section 2 introduces the functional LBM as well as its inference algorithm. Numerical experiments illustrate the interest and the behaviour of the proposed co-clustering strategy in Section 3. Then, Section 4 presents the co-clustering analysis of the Linky meter data. Some concluding remarks are provided in Section 5.

2. The functional latent block model

After introducing the notation, the functional LBM is defined. Then its inference is investigated through a stochastic expectation–maximization (SEM) algorithm embedding Gibbs sampling. The section ends with the definition of model selection criteria to choose the number of co-clusters.

2.1. The data

We assume that the data set is composed of a matrix of n individuals (rows or samples) of p curves (columns or functional features): $\mathbf{x} = (x_{ij}(t))_{1 \leqslant i \leqslant n, 1 \leqslant j \leqslant p}$ where $t \in [0, T]$ corresponds to the time in a day of observation ($T = 24$ for the Linky data). In practice, the functional expressions of the observed curves are not known and we have access to only the discrete observations at a finite set of ordered times. As explained in Aguilera *et al.* (2011), it is therefore necessary first to reconstruct the functional form of the data from their discrete observations. A common way to do this is to assume that curves belong to a finite dimensional space spanned by a basis of functions (see for example Ramsay and Silverman (2005)). We also assume that each observed curve x_{ij} ($1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant p$) can be expressed as a linear combination of basis functions $\{\phi_h\}_{h=1,\ldots,m}$:

$$x_{ij}(t) = \sum_{h=1}^{m} a_{ijh}\, \phi_h(t), \qquad t \in [0, T].$$

The basis expansion coefficients $\mathbf{a}_{ij} = (a_{ijh})_h$ of each curve x_{ij} can be estimated by least square smoothing (Ramsay and Silverman, 2005). With this assumption, each curve x_{ij} will be represented by its basis expansion coefficient vector $\mathbf{a}_{ij}$. Let $\mathbf{a} = (\mathbf{a}_{ij})_{ij}$ be the whole data set to co-cluster.

2.2. The model

The LBM (Govaert and Nadif, 2013) is certainly the most popular model for co-clustering. It assumes that the two random variables $\mathbf{z} = (z_{ik})_{1 \leqslant i \leqslant n, 1 \leqslant k \leqslant K}$ and $\mathbf{w} = (w_{jl})_{1 \leqslant j \leqslant p, 1 \leqslant l \leqslant L}$, indicating respectively the row and column partitions, are independent and that, conditionally on $\mathbf{z}$ and $\mathbf{w}$, the $n \times p$ random variables $\mathbf{x}$ are also independent. Note that a standard binary partition is used for both $\mathbf{z}$ and $\mathbf{w}$, i.e. $z_{ik} = 1$ if observation i belongs to the row cluster k and $z_{ik} = 0$ otherwise. Adapting it to functional data, we define the functional LBM:

$$p(\mathbf{a}; \theta) = \sum_{\mathbf{z} \in Z} \sum_{\mathbf{w} \in W} p(\mathbf{z}; \theta)\, p(\mathbf{w}; \theta)\, p(\mathbf{a} | \mathbf{z}, \mathbf{w}; \theta) \tag{1}$$

where (the straightforward ranges for i, j, k and l are omitted hereafter)

(a) Z is the set of all possible partitions of rows into K groups and W is the set of partitions of the columns into L groups,
(b) $p(\mathbf{z};\theta)=\Pi_{ik}\alpha_k^{z_{ik}}$ and $p(\mathbf{w};\theta)=\Pi_{jl}\beta_l^{w_{jl}}$ where α_k and β_l are the row and column mixing proportions, belonging to $[0,1]$ and summing to 1, and
(c) $p(\mathbf{a}|\mathbf{z},\mathbf{w};\theta)=\Pi_{ijkl}\,p(\mathbf{a}_{ij};\theta_{kl})^{z_{ik}w_{jl}}$ is the probability density of the basis expansion coefficients $\mathbf{a}_{ij}$.

The model that we assume for the density $p(\mathbf{a}|\mathbf{z},\mathbf{w};\theta)$ is the model that is used for each cluster in the parsimonious model FunHDDC (Bouveyron and Jacques, 2011). We therefore assume that, for each block, there is a low dimensional latent subspace in which the curves can be adequately described. Following this model, $p(\cdot;\theta_{kl})$ is an m-variate Gaussian density with mean $U_{kl}\mu_{kl}$ and variance $U_{kl}\Sigma_{kl}U_{kl}^{\mathrm{T}}+\Xi_{kl}$:

$$p(\mathbf{a}_{ij};\theta_{kl})=\mathcal{N}(\mathbf{a}_{ij};U_{kl}\mu_{kl},U_{kl}\Sigma_{kl}U_{kl}^{\mathrm{T}}+\Xi_{kl}),$$

where

(a) U_{kl} is the $m\times d$ matrix ($d<m$) defined such that the orthogonal $m\times m$ matrix describing the linear transformation between the original space of the $\mathbf{a}_{ij}$ and the low dimensional latent space can be decomposed into $Q_{kl}=[U_{kl},V_{kl}]$ with V_{kl} of size $m\times(m-d)$ with $U_{kl}^{\mathrm{T}}U_{kl}=I_d$, $V_{kl}^{\mathrm{T}}V_{kl}=I_{m-d}$ and $U_{kl}^{\mathrm{T}}V_{kl}=0$,
(b) μ_{kl} and Σ_{kl} are the mean and variance of the projection of the basis expansion coefficients of the curves belonging to block kl in the low dimensional subspace, with $\Sigma_{kl}=\mathrm{diag}(\sigma_{kl1},\ldots,\sigma_{kld})$ a diagonal matrix,
(c) Ξ_{kl} the noise covariance matrix of size $m\times m$, assumed to be such that $\Delta_{kl}=Q_{kl}^{\mathrm{T}}(U_{kl}\Sigma_{kl}U_{kl}^{\mathrm{T}}+\Xi_{kl})Q_{kl}$ can be written as

$$\Delta_{kl}=\begin{pmatrix} s_{kl1} & & 0 & & \\ & \ddots & & \mathbf{0} & \\ 0 & & s_{kld} & & \\ & & & b_{kl} & & 0 \\ & \mathbf{0} & & & \ddots & \\ & & & & & \ddots \\ & & & 0 & & b_{kl} \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} d \\ \left.\vphantom{\begin{matrix}a\\a\\a\\a\\a\\a\end{matrix}}\right\} m-d \end{matrix}$$

with $s_{klj}>b_{kl}$ for all $j=1,\ldots,d$, and
(d) $\theta_{kl}=(\mu_{kl},\Sigma_{kl},U_{kl},\sigma_{kl}^2)$ are the model parameters of block kl,

The whole set of mixture parameters is finally denoted by $\theta=(\alpha_k,\beta_l,\theta_{kl})_{1\leqslant k\leqslant K,1\leqslant l\leqslant L}$. At this point, it is possible to state the links with and differences from the work of Ben Slimen *et al.* (2016). In contrast with the model that is described here, the model that was proposed by Ben Slimen *et al.* (2016) assumes that the data for all groups live in a common functional subspace and therefore it operates a unique dimension reduction through functional PCA. In addition, the clustering of the curves is done in Ben Slimen *et al.* (2016) in the reduced space with only the first functional PCA scores, whereas all dimensions are used with different weighting in the functional LBM that we propose here.

2.3. Model inference with stochastic expectation–maximization–Gibbs algorithm

The aim is to estimate θ by maximizing the observed log-likelihood

$$l(\theta;\mathbf{a}) = \sum_{\mathbf{z},\mathbf{w}} \ln\{p(\mathbf{a};\theta)\}. \tag{2}$$

For computational reasons, and because the EM algorithm is not tractable in the co-clustering case (see Govaert and Nadif (2013)), we thus opt for one of its stochastic versions, called the *stochastic expectation–maximization (SEM)–Gibbs* (Keribin *et al.*, 2010) *algorithm*. The main idea of this algorithm is, in a so-called SE step, to generate the unobserved row and column partitions ($\mathbf{z}$,$\mathbf{w}$) without having to compute their joint distribution (which is computationally intractable), thanks to a Gibbs sampling.

Starting from an initial value $\theta^{(0)}$ for the parameter set and an initial partition $\mathbf{w}^{(0)}$ for the unobserved column grouping, the qth iteration of the partial SEM–Gibbs algorithm alternates the following SE and maximization steps.

(a) *SE step*: execute a small number of iterations of the two following steps (Gibbs sampling):

(i) generate the row partition $z_i^{(q+1)} = (z_{i1}^{(q+1)}, \ldots, z_{iK}^{(q+1)}) | \mathbf{a}, \mathbf{w}^{(q)}$ for all $1 \leqslant i \leqslant n$ according to $z_i^{(q+1)} \sim \mathcal{M}(1, \tilde{z}_{i1}, \ldots, \tilde{z}_{iK})$ with, for $1 \leqslant k \leqslant K$,

$$\tilde{z}_{ik} = p(z_{ik} = 1 | \mathbf{a}, \mathbf{w}^{(q)}; \theta^{(q)}) = \frac{\alpha_k^{(q)} f_k(\mathbf{a}_i | \mathbf{w}^{(q)}; \theta^{(q)})}{\sum_{k'} \alpha_{k'}^{(q)} f_{k'}(\mathbf{a}_i | \mathbf{w}^{(q)}; \theta^{(q)})}$$

where $\mathbf{a}_i = (\mathbf{a}_{ij})_j$ and $f_k(\mathbf{a}_i | \mathbf{w}^{(q)}; \theta^{(q)}) = \Pi_{jl} p(\mathbf{a}_{ij}; \theta_{kl}^{(q)})^{w_{jl}^{(q)}}$;

(ii) generate the column partition $w_j^{(q+1)} = (w_{j1}^{(q+1)}, \ldots, w_{jL}^{(q+1)}) | \mathbf{a}, \mathbf{z}^{(q+1)}$ for all $1 \leqslant j \leqslant p$ according to $w_j^{(q+1)} \sim \mathcal{M}(1, \tilde{w}_{j1}, \ldots, \tilde{z}_{jL})$ with, for $1 \leqslant l \leqslant L$,

$$\tilde{w}_{jl} = p(w_{jl} = 1 | \mathbf{a}, \mathbf{z}^{(q+1)}; \theta^{(q)}) = \frac{\beta_l^{(q)} f_l(\mathbf{a}_j | \mathbf{z}^{(q+1)}; \theta^{(q)})}{\sum_{l'} \beta_{l'}^{(q)} f_{l'}(\mathbf{a}_j | \mathbf{z}^{(q+1)}; \theta^{(q)})}$$

where $f_l(\mathbf{a}_j | \mathbf{z}^{(q+1)}; \theta^{(q)}) = \Pi_{ik} p(\mathbf{a}_{ij}; \theta_{kl}^{(q)})^{z_{ik}^{(q+1)}}$.

(b) *Maximization step*: estimate $\theta^{(q+1)}$ conditionally on $\mathbf{z}^{(q+1)}$ and $\mathbf{w}^{(q+1)}$. This can be done with the same maximization step as that for the EM algorithm derived for FunHDDC inference (Bouveyron and Jacques, 2011). Mixture proportions are updated by $\alpha_k^{(q+1)} = (1/n)\Sigma_i z_{ik}^{(q+1)}$ and $\beta_l^{(q+1)} = (1/p)\Sigma_j w_{jl}^{(q+1)}$, whereas the block means are updated by

$$\mu_{kl}^{(q+1)} = \frac{1}{n_{kl}^{(q+1)}} \sum_i \sum_j \mathbf{a}_{ij}^{z_{ik}^{(q+1)} w_{jl}^{(q+1)}}$$

with $n_{kl}^{(q+1)} = \Sigma_i \Sigma_j z_{ik}^{(q+1)} w_{jl}^{(q+1)}$.

For the variance parameter updates, we introduce the sample covariance matrix of the block indexed by kl:

$$C_{kl}^{(q)} = \frac{1}{n_{kl}^{(q)}} \sum_{i=1}^{n} \sum_{j=1}^{p} z_{ik}^{(q+1)} \omega_{jl}^{(q+1)} (\mathbf{a}_{ij} - \mu_{kl}^{(q)})^{\mathrm{T}} (\mathbf{a}_{ij} - \mu_{kl}^{(q)}),$$

and $\mathbf{\Omega} = (\Omega_{jk})_{1 \leqslant j,k \leqslant m}$, the matrix of inner products between the basis functions, $\Omega_{jk} = \int_0^T \phi_j(t)\phi_k(t)\,\mathrm{d}t$. With this notation, the update formulae for the model parameters s_{klj}, b_{kl} and Q_{klj} are

(a) the d first columns of Q_k are updated by the eigenvectors that are associated with the largest eigenvalues of $\mathbf{\Omega}^{1/2}C_{kl}^{(q)}\mathbf{\Omega}^{1/2}$,
(b) the variance parameters s_{klj}, $j=1,\ldots,d$, are updated by the d largest eigenvalues of $\mathbf{\Omega}^{1/2}C_{kl}^{(q)}\mathbf{\Omega}^{1/2}$ and
(c) the variance parameters b_k are updated by $\mathrm{tr}(\mathbf{\Omega}^{1/2}C_{kl}^{(q)}\mathbf{\Omega}^{1/2}) - \Sigma_{j=1}^{d} s_{klj}^{(q)}$.

The SEM–Gibbs algorithm is run for a given number of iterations (from our experiments, three iterations are sufficient to ensure good behaviour of the algorithm). After a burn-in period, the final estimate of the parameters is the mean of the sample distribution. We denote the final estimate by $\hat{\theta}$. Then, a sample of $(\mathbf{z},\mathbf{w})$ is generated with the Gibbs sampling described above (the SE step) with θ set to $\hat{\theta}$. The final bipartition $(\hat{\mathbf{z}},\hat{\mathbf{w}})$ is estimated by the mode of their sample distributions.

2.4. Model selection

Regarding model selection, since a model-based approach is proposed here, two functional LBMs will be seen as different if they have different values of K and/or L. Therefore, the task of estimating K and L can be viewed as a model selection problem. Many model selection criteria have been proposed in the literature, such as the Akaike information criterion (Akaike, 1974) and the Bayesian information criterion (Schwarz, 1978). In this paper, because the optimization procedure considered involves binary matrices for $\mathbf{z}$ and $\mathbf{w}$, we rely on *the integrated information likelihood criterion*. This criterion was originally proposed by Biernacki *et al.* (2000) for Gaussian mixture models. Below we extend to the functional case the integrated information likelihood criterion that was developed by Lomet (2012) for co-clustering:

$$\mathrm{ICL}(K,L) = \log\{p(\mathbf{x},\hat{\mathbf{v}},\hat{\mathbf{w}};\hat{\theta})\} - \frac{K-1}{2}\log(n) - \frac{L-1}{2}\log(p) - \frac{KL\nu}{2}\log(np)$$

where $\nu = md + d + 1$ is the number of continuous parameters per block and

$$\log\{p(\mathbf{x},\hat{\mathbf{v}},\hat{\mathbf{w}};\hat{\theta})\} = \prod_{ik}\hat{z}_{ik}\log(\alpha_k) + \prod_{jl}\hat{w}_{jl}\log(\beta_l) + \sum_{ijkl}\hat{z}_{ik}\hat{w}_{jl}\log\{p(\mathbf{a}_{ij};\hat{\theta}_{kl})\}.$$

The couple (K,L) leading to the highest ICL-value is selected as the most appropriate model for the data at hand.

3. Numerical experiments

This section highlights the main features of the proposed approach on synthetic data. In particular, the validity of the inference algorithm and model selection criterion, which were both presented in the previous section, is demonstrated on simulated data.

3.1. Simulation set-up

To simplify the characterization and to facilitate the reproducibility of the experiments, we designed a common simulation scenario on which the main characteristics of the methodology proposed will be illustrated. The simulation set-up is as follows.

(a) We designed four different functions $f_1(t),\ldots,f_4(t)$, which will serve as block means, at equispaced time points $t = 0, 1/T, 2/T, \ldots, 1$. Fig. 1 shows those functions.
(b) Then, all curve points are sampled as follows:

$$x_{ij}(t)|\mathbf{z}_{ik}\mathbf{w}_{jl} = 1 \sim \mathcal{N}\{m_{kl}(t), s^2\},$$

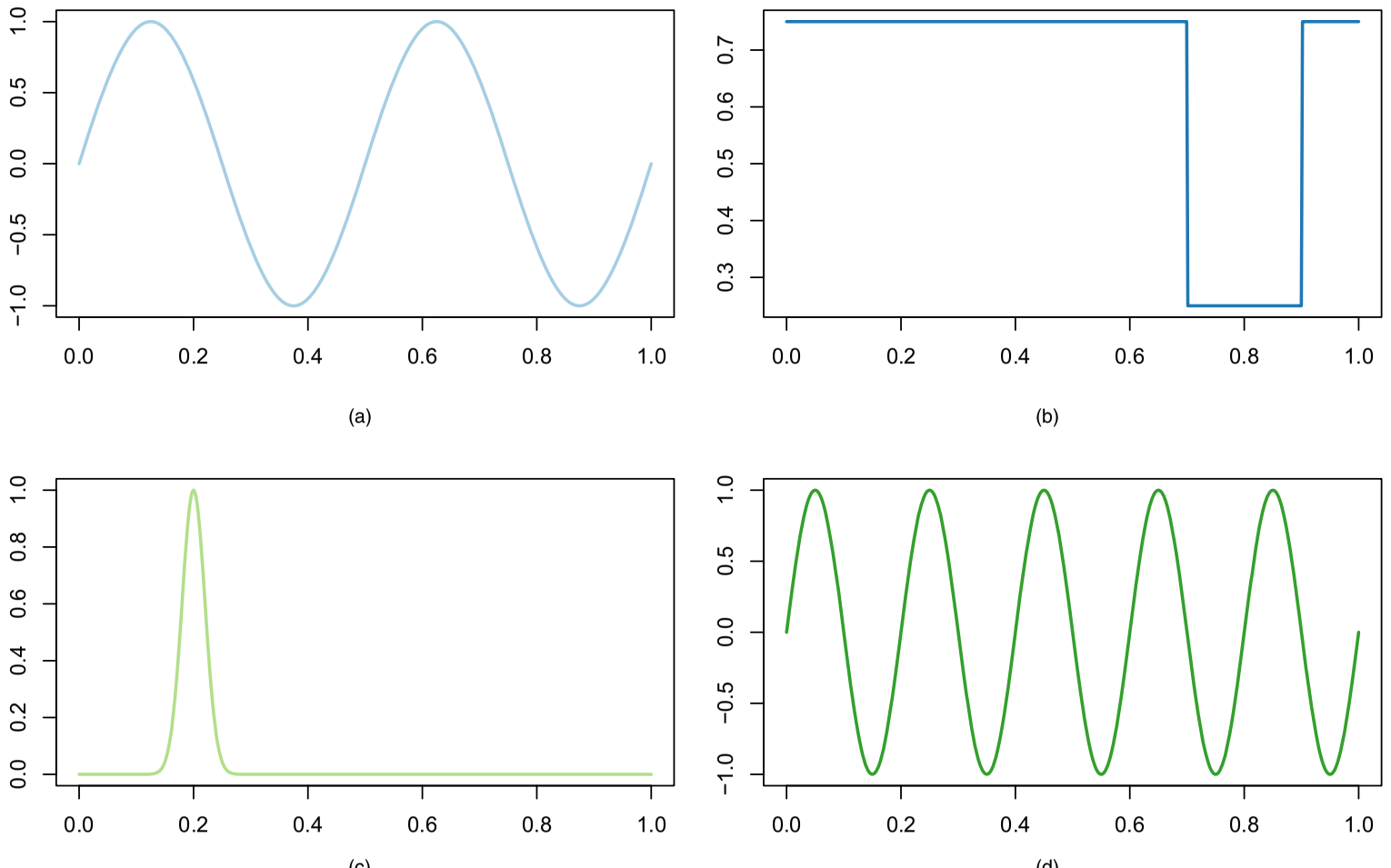
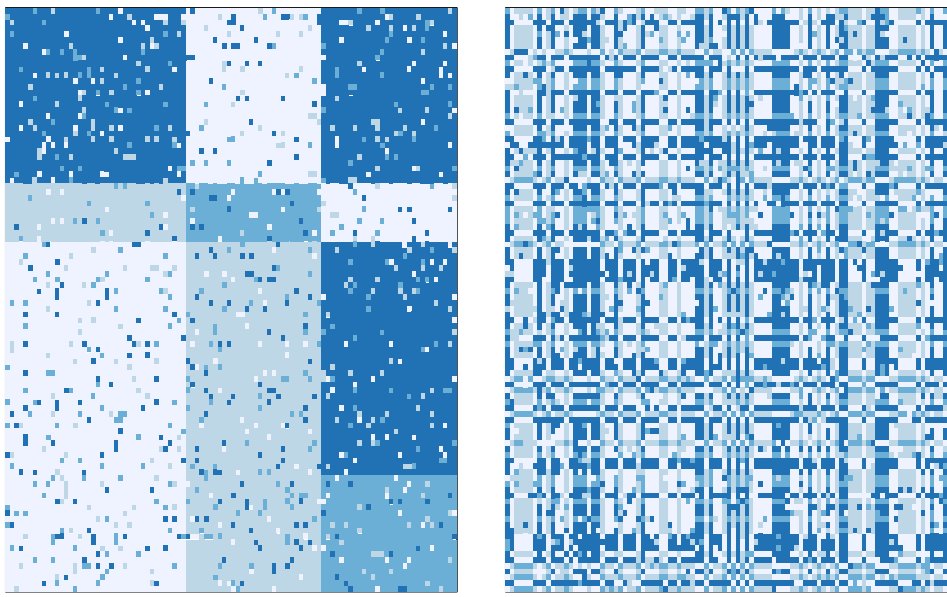


Fig. 1. The four mean functions for the blocks used in the simulations (a Gaussian noise term is then added to the mean functions to obtain the simulated curves): (a) function 1; (b) function 2; (c) function 3; (d) function 4

Table 1. Parameter values for the three simulation scenarios†

	Results for the following scenarios:		
	A	*B*	*C*
n (number of rows)		100	
p (number of columns)		100	
T (length of curves)		30	
K (number of row groups)	3	4	4
L (number of column groups)	3	3	3
α (row group proportion)	$(0.333,\ldots,0.333)$	$(0.2, 0.4, 0.1, 0.3)$	$(0.2, 0.4, 0.1, 0.3)$
β (column group proportion)	$(0.333,\ldots,0.333)$	$(0.4, 0.3, 0.3)$	$(0.4, 0.3, 0.3)$
τ (simulation noise)	0	0.1	0.3

†See the text for details.



(a) (b)

Fig. 2. A simulated data set with noise $\tau = 0.1$ (the colour intensities indicate the block mean functions used for simulation): (a) rows and columns sorted according to the actual groups; (b) same data matrix with random orders on rows and columns

where $s = 0.3$, $m_{11} = m_{21} = m_{33} = m_{42} = f_1$, $m_{12} = m_{22} = m_{31} = f_2$, $m_{13} = m_{32} = f_3$ and $m_{23} = m_{41} = m_{43} = f_4$.

(c) Finally, we add some noise within the blocks by randomly simulating a certain percentage τ of curves by using other block means.

Table 1 provides the parameter values for the three simulation scenarios. Fig. 2 shows a simulated data set according to scenario B ($K = 4$; $L = 3$; noise $\tau = 0.1$) where colours indicate the block mean functions used. It is worth noting that all simulation scenarios have been designed such that they do not follow the functional LBM and therefore they do not particularly favour

our model in comparisons. In all the following experiments, Fourier basis functions are used to reconstruct the functional form of the data.

3.2. An introductory example

As an introductory example, we consider a data set that was simulated according to scenario B: $K = 4$ groups of rows, $L = 3$ groups of columns, unbalanced row and column groups, and 10% of block noise. Fig. 2 shows such a data set. To illustrate the behaviour of the inference algorithm proposed, the SEM–Gibbs algorithm was run on the data with the actual numbers of row and column groups (the problem of model selection will be considered in the next section). First, Fig. 3(a) shows the behaviour of the complete-data likelihood over the iterations of the functional LBM algorithm. We can see that the functional LBM algorithm converges in a few iterations. Fig. 3 also presents the evolution of the SEM–Gibbs estimates for model parameters α and β along the iterations. Fig. 4 finally shows the clustering obtained, which here is perfect regarding both the simulated row and column partitions. Finally, it may be meaningful in practical cases to be able to visualize the functional means estimated by the functional LBM algorithm. Fig. 5 shows the estimated functional means for the $K \times L$ blocks and we can distinguish functions very close to the block mean functions that were used in the simulations (Fig. 1).

3.3. Initialization

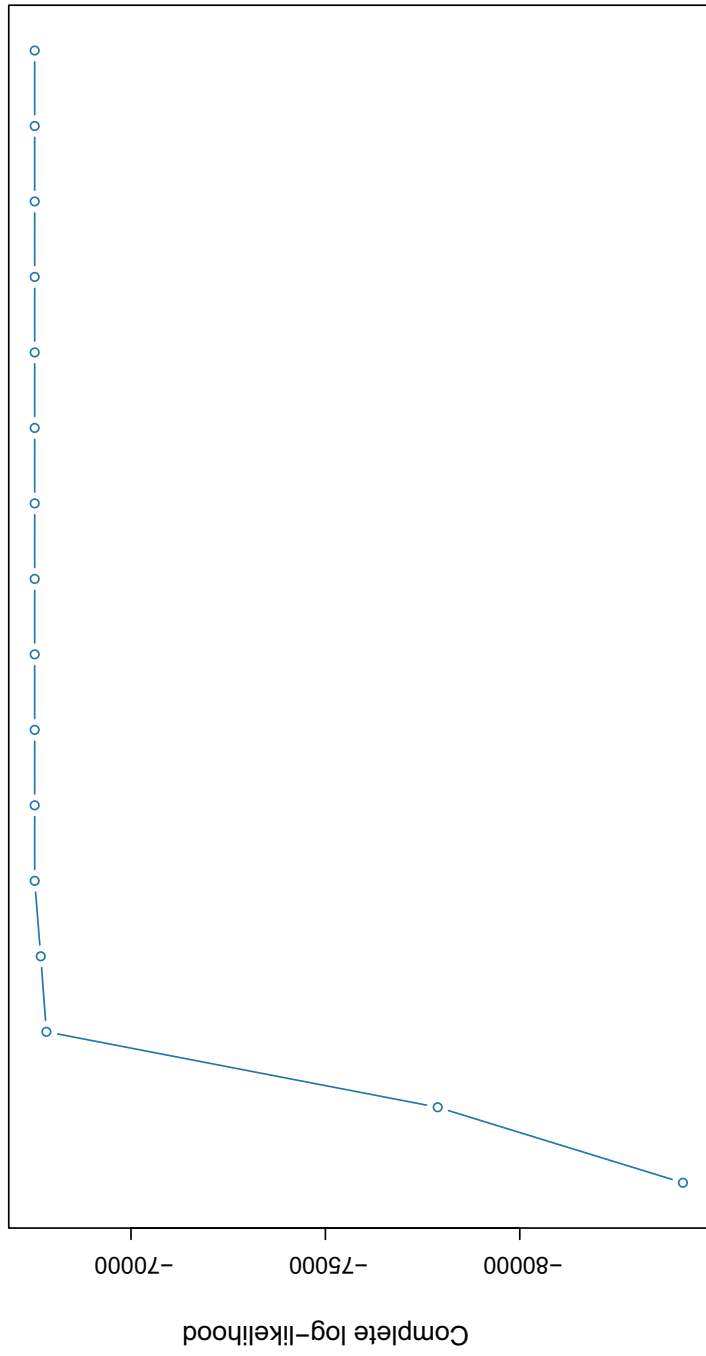
We now focus on the initialization of the SEM–Gibbs algorithm and consider three possible ways for that: random, k-means and functional. The first possible initialization procedure is to sample the $\mathbf{z}$- and $\mathbf{w}$-values randomly. The second consists in running the k-means algorithm on the binning of the time series values according to the rows and then the columns. Finally, we also propose to use a functional clustering algorithm, funFEM (Bouveyron *et al.*, 2015), instead of a k-means algorithm, on the binned functions.

We run our SEM–Gibbs algorithm with the three initialization strategies on 25 data sets simulated according to the three simulation scenarios. We evaluated the performance of the different results by using the adjusted Rand index (ARI) (Rand, 1971) on both row and column partitions. In the clustering community, the ARI serves as a widely accepted criterion for the difficult task of clustering evaluation. The ARI looks at all pairs of nodes and checks whether they are classified in the same group or not in both partitions. As a result, an ARI value close to 1 means that the partitions are similar and, in our case, that the functional LBM algorithm succeeds in recovering the simulated partitions. Fig. 6 presents the results of this study.

As we can see, the three techniques most of the time yield a satisfying result for both row and column partitions in scenario A, which is the easier situation. For scenario B, which has unbalanced groups and some noise, the functional-based initialization clearly outperforms the two other strategies. Conversely, the k-means initalization seems to be slightly superior in scenario C compared with the functional initialization. As a summary, the functional-based initialization may be viewed as the best overall solution and will be used in the following experiments.

3.4. Model selection

This third simulation study focuses on model selection and highlights the ability of our approach to catch the actual model. For this, 100 data sets were simulated for each scenario and the SEM–



(a)

Fig. 3 (*continued*)

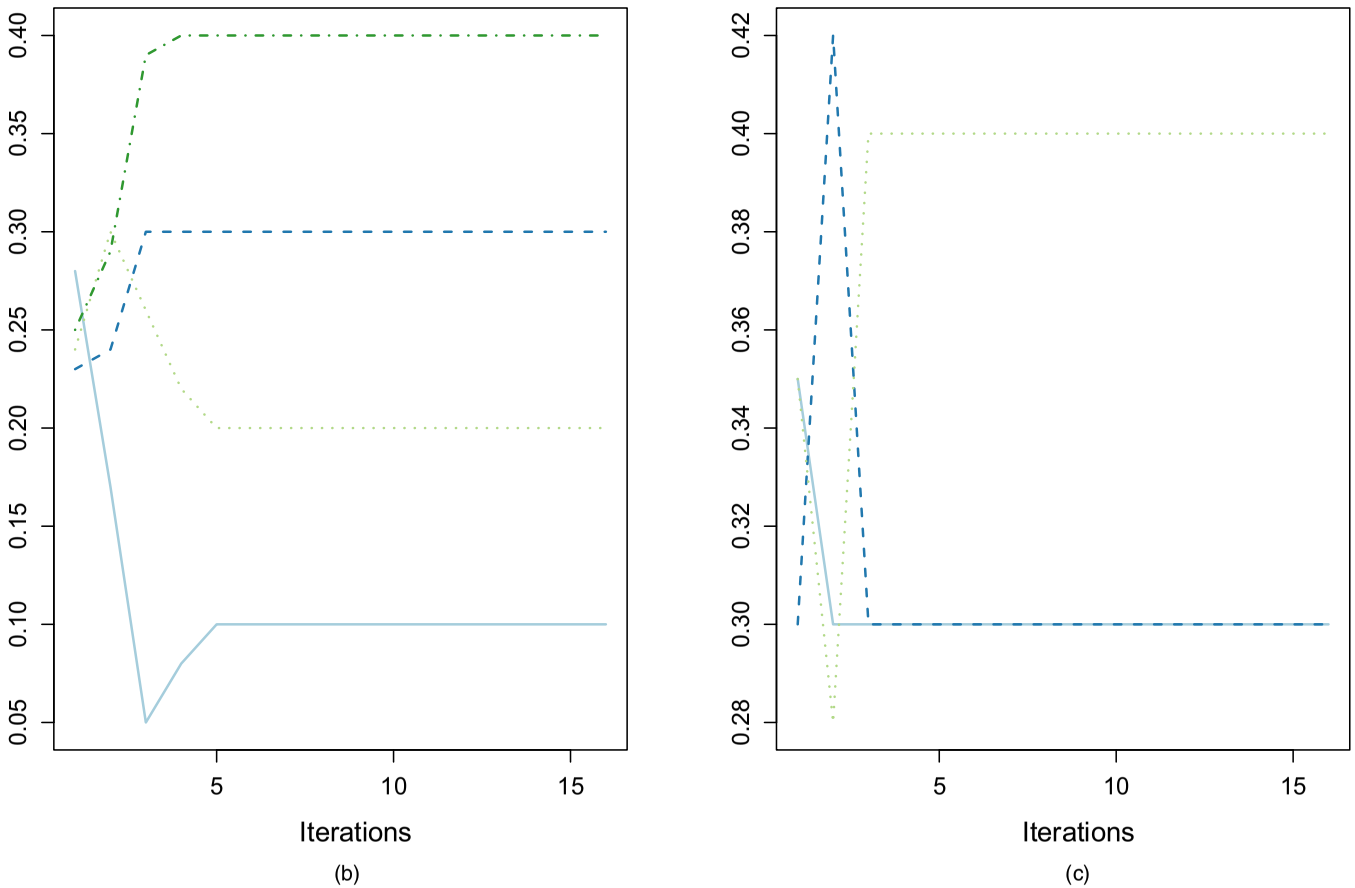
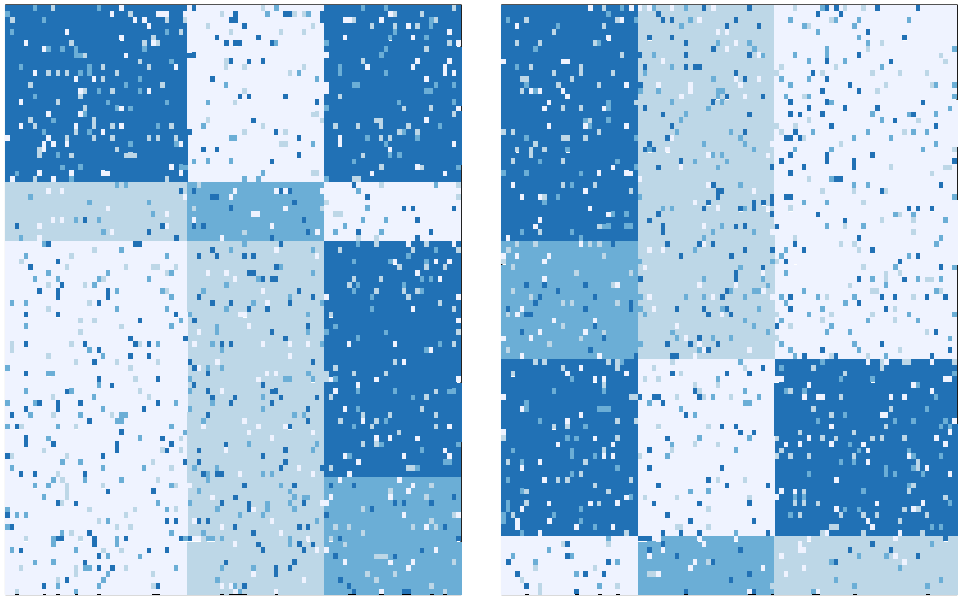


Fig. 3. (a) Complete-data likelihood over the iterations of the functional LBM algorithm and estimates (b) α and (c) β for mixture parameters along the iterations of the SEM–Gibbs algorithm on the introductory example



(a) (b)

Fig. 4. Clustering results on the introductory example: (a) data matrix with the rows and columns sorted according to the actual groups; (b) data matrix sorted according to the groups found by the functional LBM

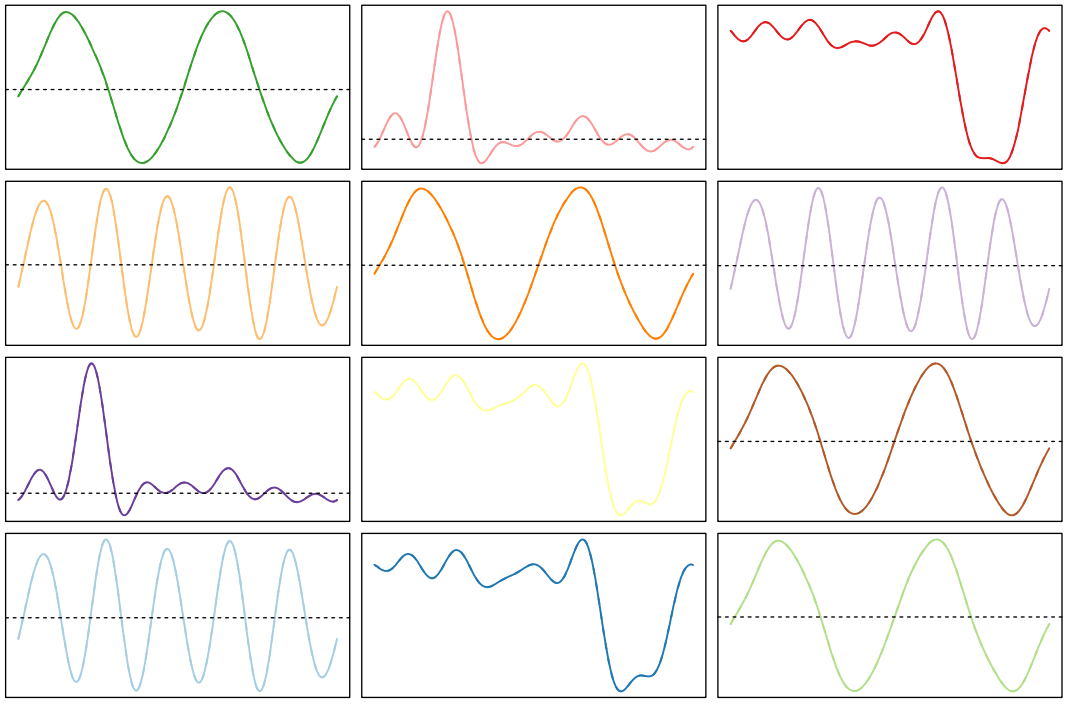


Fig. 5. Estimated functional means for the $K \times L$ blocks on the introductory example by the functional LBM

Gibbs algorithm (with the functional initialization) was applied in combination with our model selection criterion for values of K and L ranging from 1 to 6. Table 2 presents the percentage of selections by integrated information likelihood for each model (K, L) on the 100 simulated data sets of each of the three scenarios.

In the three situations, our ICL-criterion succeeds most of the time in identifying the actual combination of the number of row and column groups. For scenario A, the criterion enables our approach to identify perfectly the correct models. The task is of course slightly more difficult in the cases of the noisy scenarios B and C. ICL enables us nevertheless to recover the actual simulation model in more than seven cases out of 10. It is worth noting that, when ICL does not select the correct values for K and L, wrongly selected models are usually close to the simulated model. Recall also that, since the data are not strictly simulated according to a functional LBM, the ICL-criterion does not have the model which generated the data in the set of tested models. This experiment enables us to validate ICL as a model selection tool for the functional LBM.

4. Application to Électricité de France data

This section now presents the results of modelling with the functional LBM of electricity consumption curves measured through self-recording meters, known as Linky meters.

4.1. Context of the study

With the upcoming installation of Linky meters in all French households, the field of operational applications of co-clustering methods is very wide at EDF. For instance, the co-clustering results may be used to design new marketing offers, to propose demand response programmes or to detect outliers. Indeed, before launching a new offer or service, some experimental trials are always made to evaluate the effect of the offer by comparing a test group and a control group. The two groups must be similar except for the offer or the service tested. To ensure this hypothesis, we could select the samples among clusters built by the co-clustering technique. It is also possible to use the co-clustering results to design programmes which consist in giving incentives to customers to use less electricity at critical peak times. Two ways exist: price-based demand responses use changing prices to induce changes in customers' consumption of electricity and direct load control, where equipment can be shut down remotely by the programme operator. Once again, the groups of households and days that are found by the co-clustering may be used to parameterize these programmes. Finally, with co-clustering techniques, the electricity operator should be able to detect outliers and to warn customers if their consumption increases unusually, compared with the mean or the quantiles of the clients of the same cluster.

4.2. Data and protocol

The Linky data set that was provided by EDF comprises 1481 households in France (metropolitan) for which electricity consumption has been monitored every 30 min and over a period of almost 2 years (July 2010–March 2012). The data set therefore consists of a table with $n = 1481$ rows (households) and $p = 630$ columns (days), where each entry is a time series of daily electricity consumption with 48 measures. To transform the raw consumption data as meaningful functional data, the consumption data were first regressed against the observed temperatures at each household location to accommodate geographic variations. Then, the residuals of such a regression were projected on a basis of 15 Fourier functions. We finally end up with a $1481 \times 630 \times 15$ cube which contains for each row and column the 15 Fourier coefficients of the corresponding individual electricity consumption profiles. The SEM–Gibbs algorithm was run to infer func-

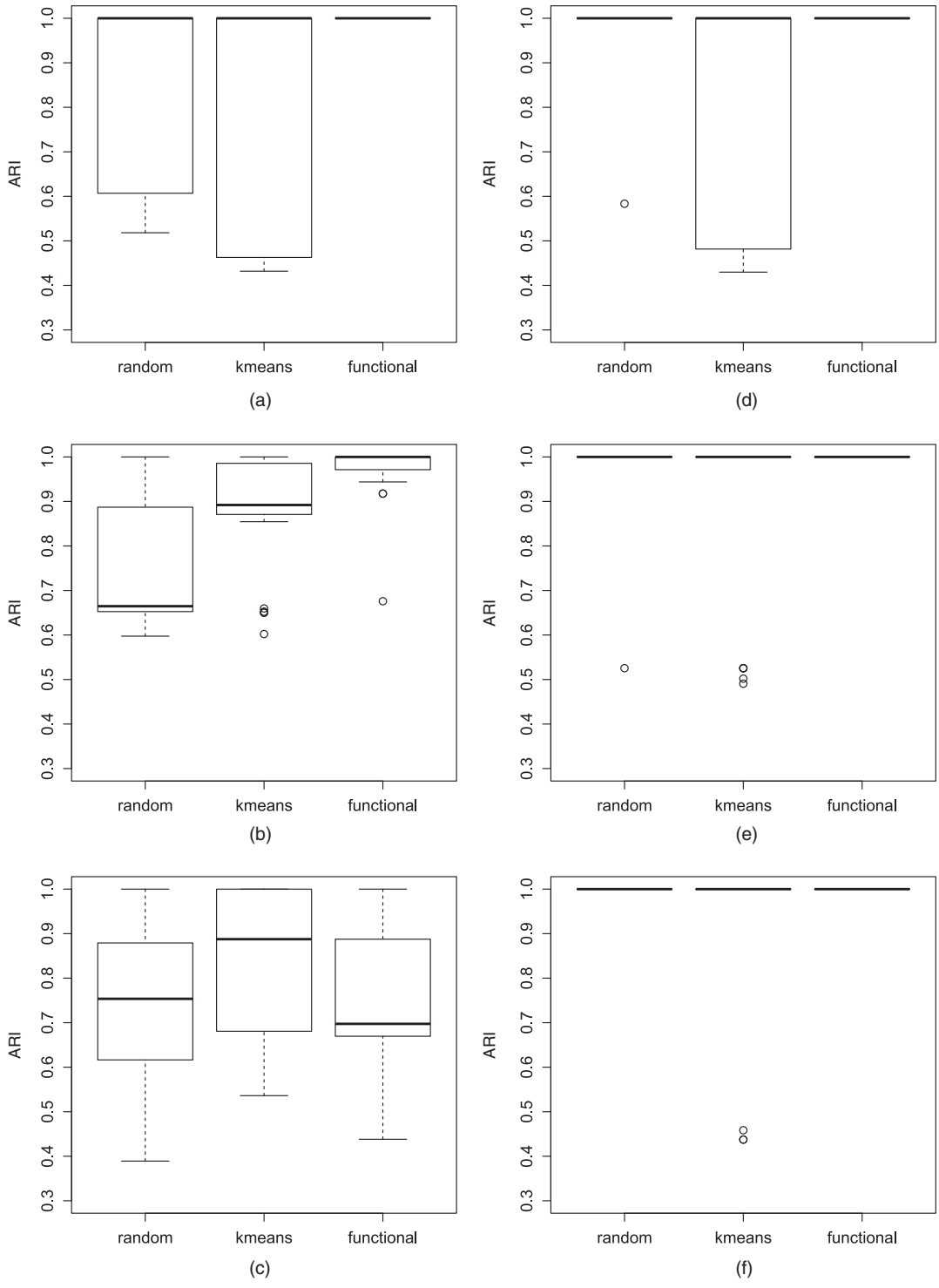


Fig. 6. ARI values (a)–(c) on rows and (d)–(f) on columns for the various initialization procedures for the functional LBM on the three simulation scenarios: (a), (d) scenario A; (b), (e) scenario B; (c), (f) scenario C

Table 2. Percentage of selections by integrated information likelihood for each model (K, L) on 100 simulated data sets of each of the scenarios†

K	*Results for scenario A* ($K=3$, $L=3$)						*Results for scenario B* ($K=4$, $L=3$)						*Results for scenario C* ($K=4$, $L=3$)					
	$L=1$	$L=2$	$L=3$	$L=4$	$L=5$	$L=6$	$L=1$	$L=2$	$L=3$	$L=4$	$L=5$	$L=6$	$Q=1$	$Q=2$	$Q=3$	$Q=4$	$Q=5$	$Q=6$
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0
3	0	0	*100*	0	0	0	0	0	0	0	0	0	0	0	*77*	0	0	0
4	0	0	0	0	0	0	0	0	*70*	0	1	0	0	0	5	0	0	0
5	0	0	0	0	0	0	0	0	26	1	0	0	0	0	1	0	0	0
6	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0

†Highlighted rows and columns correspond to the actual values for K and L.

Table 3. Selection of the most appropriate model for the Linky data by using the ICL-criterion†

Rank	K	L	*ICL* ($\times 10^6$)
1	*9*	*4*	*−116.09*
2	10	3	−116.11
3	9	3	−116.15
4	8	4	−116.18
5	8	3	−116.36

†The first row, highlighted in italics, is the best in terms of ICL.

tional LBMs on this final data set for a number K of row (household) groups and a number L of column (date) groups ranging from 2 to 10.

4.3. Numerical results

As shown by Table 3, ICL selects the functional LBM with nine groups of households and four groups of dates. Table 3 also shows the five models that were selected as the most appropriate for the data by integrated information likelihood. It is interesting that, among the best models, there is a relative consensus on the choice of K and L, since those models are 'centred' on the values of the best model. We therefore comment in what follows on the results corresponding to the functional LBM with $K=9$ and $L=4$.

Fig. 7 shows the estimated proportions $\hat{\alpha}$ and $\hat{\beta}$ for respectively the row and column groups. It is interesting that the functional LBM formed groups of rows and columns which are balanced, without extremely small or large groups. Fig. 8 then presents the average consumption profile of the nine groups of households for the four identified periods of time. Before going further in the analysis, we must explain that two major factors can have effects on the load curve profiles. First, the possession of an electrical heating system, which is particularly widespread in France (around 30%), results in load peaks during winter and, on average, as soon as the external temperature falls below 15 °C. Second, the possession of an electrical heating tank for sanitary

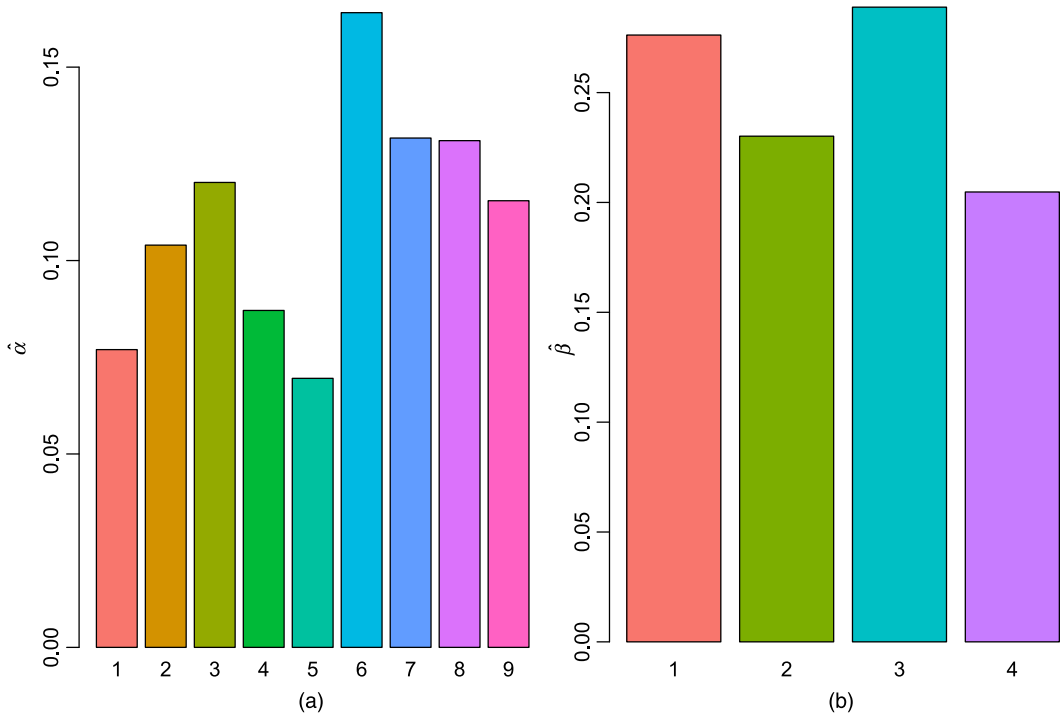


Fig. 7. Proportions $\hat{\alpha}$ and $\hat{\beta}$ for (a) the row and (b) the column groups estimated by the functional LBM

water also has a significant influence on load curve profiles. About 45% of dwellings in France are equipped with such systems. To flatten its overall load curve, EDF was the first operator to create a time-of-use tariff so that consumers benefit for 8 h a day at a lower price to encourage them to shift some use of appliances. Two main designs are proposed: one with 8 h of off-peak prices during the night and another with 2 h of off-peak price between noon and 5 p.m., and six remaining hours during the night.

At this point, it may be noted from Fig. 8 that some household groups have specific behaviours. On the one hand, the functional LBM enables us to identify six groups (groups 1, 3, 4, 5, 8 and 9) of households which have stable electricity consumption throughout the year, but which differ in their consumption profiles. These seem not to have electricity as their principal heating energy. For instance, the first and third groups have almost constant consumption of electricity during both the day and the year (they differ slightly in winter, probably because of the use of a secondary heating system which may be electric), whereas the fourth and fifth groups have profiles which vary strongly during a day but the day profile is stable during a year. Their daily profile is typical from customers who benefit from off-peak hours during the day. The eighth and ninth groups reveal the daily behaviour of customers with 8 h of off-peak price in a row during the night.

On the other hand, three groups (groups 2, 6 and 7) have consumption profiles which are dependent on the time periods. To understand those variations better, it is necessary to have a look at the groups of days that the functional LBM provided. Fig. 9 shows the group memberships of the 630 days of the data set in a calendar view. Briefly, the first period (class 1) corresponds to intermediate seasons (spring and autumn), the second period (class 2) gathers winter days, the third (class 3) corresponds to the beginning and end of winter and finally the

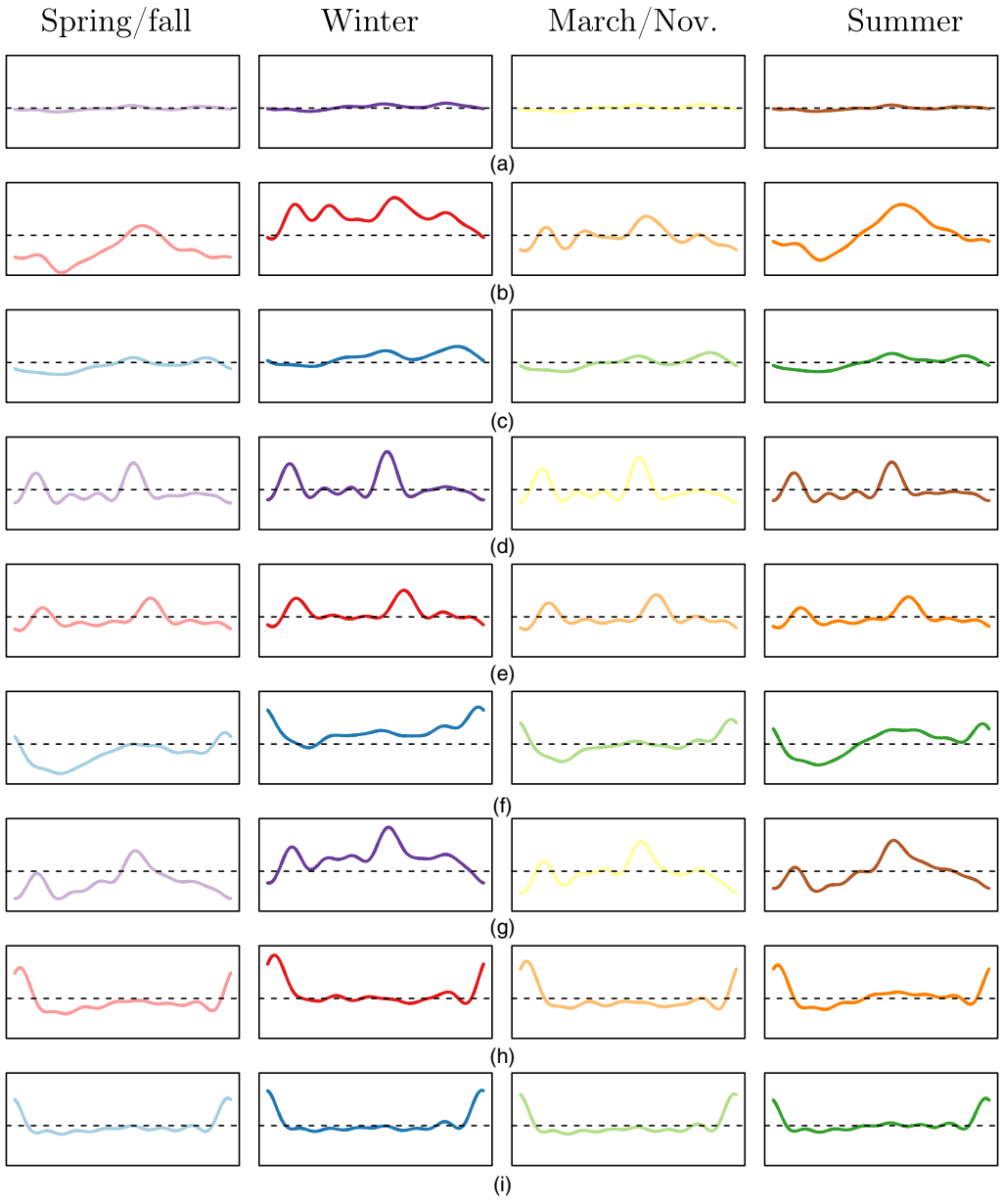


Fig. 8. Average curves of each block, as estimated by the SEM–Gibbs algorithm for the functional LBM: (a) group 1; (b) group 2; (c) group 3; (d) group 4; (e) group 5; (f) group 6; (g) group 7; (h) group 8; (i) group 9

fourth period (class 4) comprises summer and spring holidays. In view of this interpretation of the four periods, the sixth and seventh household groups have similar consumption profiles but with different intensity: consumption is higher in winter and summer than in the intermediate seasons. More interestingly, the second group (Fig. 8(b)) has significantly different profiles in winter and summer: in winter, they consume frequently during the day whereas, in summer, they consume only in the afternoon. This may be due to the installation of a programmable controller for the office heating system.

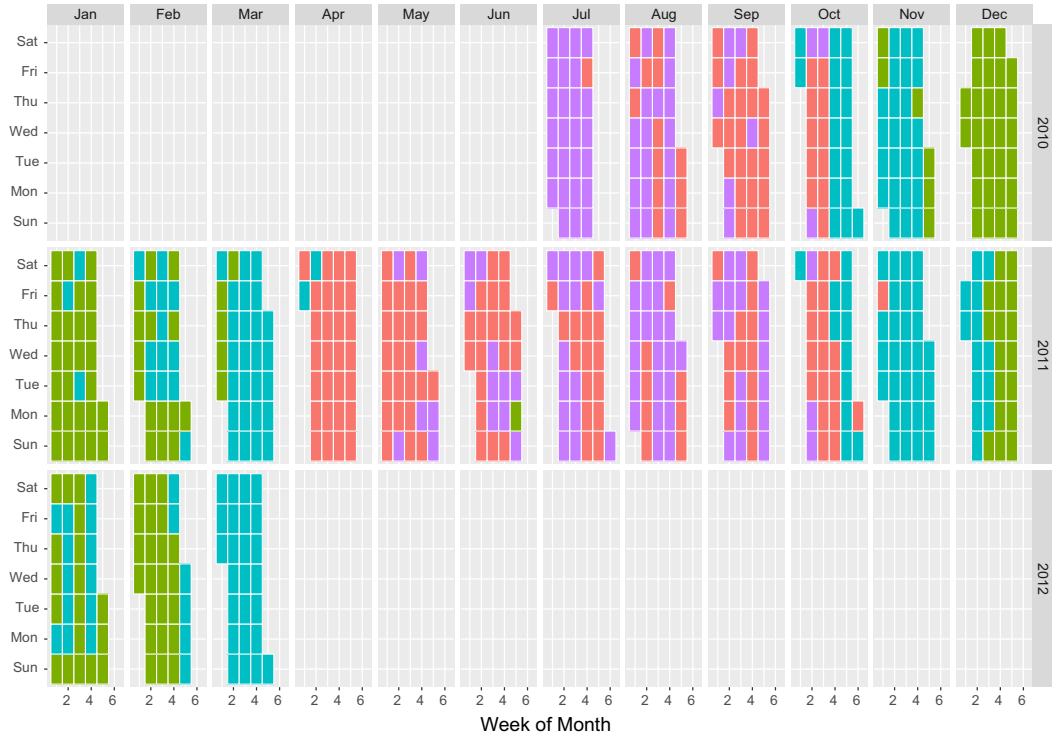


Fig. 9. Clustering of columns (dates), viewed as a calendar, as provided by the functional LBM: ■, class 1; ■, class 2; ■, class 3; ■, class 4

Finally, we highlight that one specific day has a surprising group membership: Wednesday, June 27th, 2011, is classified in the winter period. This may be explained by the fact that a heatwave was at its maximum on that day (37.5 °C at Clermont-Ferrand), forcing people to use maximum air conditioning to cool their homes or offices. This unusual use of air conditioning may be viewed here as similar in terms of electricity profiles to winter days where people use electric heating systems.

With this co-clustering method, EDF has obtained a very precise clustering of the load curves with clusters significantly different in terms of seasonality and daily forms. For EDF, co-clustering with the functional LBM is an innovative approach since it enables us to obtain these kinds of cluster in one batch whereas EDF usually had to perform several independent steps. Indeed, the current approach at EDF for this is to combine two different clusterings based on dimension reduction (PCA and Kohonen): one for the seasonality, and the other for the daily forms. We also highlight that the functional LBM framework enables us to identify the best combination of number K of household groups and number L of day groups automatically.

5. Conclusion

To address the upcoming problem of exploring extremely large sets of electricity consumption curves, we proposed in this work a new co-clustering methodology for functional data, based on the functional LBM. The resulting co-clustering algorithm enables us to build 'summaries' of these large consumption data which can be used efficiently by operators. The functional LBM

extends the LBM to the functional case by assuming that the curves of one block live in a low dimensional functional subspace. Model inference is done through an SEM–Gibbs algorithm and an ICL-criterion has been also derived to address the model selection problem (choosing the number of row and column groups). Numerical experiments on simulated and original Linky data have shown the usefulness of the methodology.

Among the possible extensions of this work, it would be interesting to take into account in the modelling the spatial dependences that exist between the consumption curves of neighbouring households. This would be possible by using Markov random fields for instance. Another possible extension would be to study the possibility of using model selection tools, such as the integrated information likelihood, to select the appropriate basis functions to smooth the raw curves.

Acknowledgements

This research has benefited from the support of the 'Fondation Mathématique Jacques Hadamard research initiative data science for industry' and from the support for this programme by EDF and by Thalès Optronique.

References

Abreu, J. M., Pereira, F. C. and Ferrão, P. (2012) Using pattern recognition to identify habitual behavior in residential electricity consumption. *En. Buildngs*, **49**, 479–487.

Aguilera, A., Escabiasa, M., Preda, C. and Saporta, G. (2011) Using basis expansions for estimating functional PLS regression: applications with chemometric data. *Chemometr. Intell. Lab. Syst.*, **104**, 289–305.

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Trans. Autom. Control*, **19**, 716–723.

Ben Slimen, Y., Allio, S. and Jacques, J. (2016) Model-based co-clustering for functional data. In *Proc. 48th Conf. French Statistical Society, Montpellier*.

Biernacki, C., Celeux, G. and Govaert, G. (2000) Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattn Anal. Mach. Intell.*, **7**, 719–725.

Bouveyron, C., Côme, E. and Jacques, J. (2015) The discriminative functional mixture model for the analysis of bike sharing systems. *Ann. Appl. Statist.*, **9**, 1726–1760.

Bouveyron, C. and Jacques, J. (2011) Model-based clustering of time series in group-specific functional subspaces. *Adv. Data Anal. Classificn*, **5**, 281–300.

Govaert, G. and Nadif, M. (2013) *Co-clustering*. Hoboken: Wiley.

Hartigan, J. (1972) Direct clustering of a data matrix. *J. Am. Statist. Ass.*, **67**, 123–129.

Jacques, J. and Biernacki, C. (2017) Model-based co-clustering for ordinal data. *Technical Report HAL 01448299*.

Jacques, J. and Preda, C. (2014) Functional data clustering: a survey. *Adv. Data Anal. Classificn*, **8**, 231–255.

Keribin, C., Govaert, G. and Celeux, G. (2010) Estimation d'un modèle à blocs latents par l'algorithme sem. In *Proc. 42nd Conf. French Statistical Society, Marseille*.

Keyno, H. S., Ghaderi, F., Azade, A. and Razmi, J. (2009) Forecasting electricity consumption by clustering data in order to decline the periodic variable's affects and simplification of the pattern. *En. Conversn Mangmnt*, **50**, 829–836.

Lomet, A. (2012) Sélection de modèle pour la classification croisée de données continues. *PhD Thesis*. Université de Technologie de Compiègne, Compiègne.

Melzi, F. N., Same, A., Zayani, M. H. and Oukhellou, L. (2017) A dedicated mixture model for clustering smart meter data: identification and analysis of electricity consumption behaviors. *Energies*, **10**, no. 10, article 1446.

Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*, 2nd edn. New York: Springer.

Rand, W. (1971) Objective criteria for the evaluation of clustering methods. *J. Am. Statist. Ass.*, **66**, 846–850.

Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.

Tsekouras, G. J., Hatziargyriou, N. D. and Dialynas, E. N. (2007) Two-stage pattern recognition of load curves for classification of electricity customers. *IEEE Trans. Powr Syst.*, **22**, 1120–1128.