Detection of energy waste in French households thanks to a co-clustering model for multivariate functional data

Amandine Schmutz, Julien Jacques, Charles Bouveyron, Laurent Bozzi, Laurence Cheze, Pauline Martin

Detection of energy waste in French households thanks to a co-clustering model for multivariate functional data

Amandine Schmutz*
Lim France, Chemin Fontaine de Fanny, Nontron, France
Julien Jacques,
Université de Lyon, Lyon 2, ERIC EA3083, Lyon, France
Charles Bouveyron,
Université Côte d'Azur, Inria, CNRS
LJAD, Maasai team, France
Laurent Bozzi,
EDF R&D, Paris-Saclay,France
Laurence Chèze,
Université de Lyon, Lyon 1, LBMC UMR T9406, Lyon, France
Pauline Martin,
Lim France, Chemin Fontaine de Fanny, Nontron, France,
CWD-VetLab, Ecole Nationale Vétérinaire d'Alfort,
Maisons-Alfort, F-94700, France

October 11, 2019

*The authors would like to thank the LabCom 'CWD-VetLab' for its financial support. The LabCom 'CWD-VetLab' is financially supported by the Agence Nationale de la Recherche (contract ANR 16-LCV2-0002-01). This research has also benefited from the support of the "FMJH Research Initiative Data Science for Industry".

Abstract

The exponential growth of smart devices in all aspects of everyday life leads to make common the collection of high frequency data. Those data can be seen as multivariate functional data: quantitative entities evolving along time, for which there is a growing needs of methods to summarize and understand them. The database that have motivated our project is supplied by the historical French electricity provider whose aim is to detect poorly insulated buildings, anomalies or long periods of absence. Their motivation is to answer COP24 requirements to reduce energy waste and to adapt electric load. To this end, a novel co-clustering model for multivariate functional data is defined. The model is based on a functional latent block model which assumes for each block a probabilistic distribution for multivariate functional principal component scores. A Stochastic EM algorithm, embedding a Gibbs sampler is proposed for model inference, as well as model selection criteria for choosing the number of co-clusters.

Keywords: latent block model, multivariate functional PCA, SEM-Gibbs algorithm, electricity consumption.

1 Introduction

Sensor networks are undergoing a great expansion in several domains such as industry 4.0, environment, transport, defense, smart cities, etc. The emergence of small size sensors at a reduced cost leads to an increasing use of connected devices for the mainstream. These sensors are put into smart houses to simultaneously follow, at a high frequency, the temperature of different rooms and the outdoor temperature for example. Those data can also be correlated to the electric consumption of the household. Studying electric consumption allows cities to have a better network management, users to reduce their energetic cost, and electric providers to adapt a strategy to meet demand. It is also an opportunity to gather customers consumption data and therefore improve customer knowledge. EDF, the historical and main French electricity provider, plans to access smart meters data every half an hour, which means 17472 measures per year for each of its 27 million clients. In the near future, many other information will be collected simultaneously by EDF thanks to IoT devices on all production facilities (nuclear power stations, hydrolic centrals, windmills, ...). Indeed, having the opportunity to cluster both wind production and electric consumption would allow a better steering of the energy use in order to use at the right time what is produced by the windmill for example. One of the major challenge of those devices is that they represent a mass of data to store and analyze. Thus, it may be necessary to build summaries allowing an easier storage and analysis. One way to achieve that is to cluster those data into homogeneous groups. In this work, we focus on the analysis of three features observed on a set of households: power consumption, indoor temperature and outdoor temperature. Of course, the proposed model can be extended to any set of multivariate temporal series.

There is an abundant literature on understanding electricity consumption patterns.

Simplest methods are based on averaging electric consumption per day and then applying a hierarchical clustering on those averaged values (Gouveia et al., 2015). Other works rely on studying temporal series monitored for a month with fuzzy methods such as fuzzy c-means algorithm (Zhou et al., 2017), but those methods neglect the time dependency inherent in such data. Thus, extensions of k-means methods are proposed to take into account the temporal component through preprocessing steps before applying the k-means algorithm (Tureczek et al., 2018) or with a dynamic k-means clustering algorithm where the similarity distances are calculated taking into account all Euclidean distances between each pair of objects at the same time stamp (Benitez et al., 2014). The same authors also proposed a dynamic clustering algorithm where two objects are compared thanks to a final distance which is the average of all comparisons of objects features (Benitez et al., 2016). Contrary to the previous methods, the dynamic of the data is taken into account by decomposing the temporal series into smaller linear surfaces which are compared by applying a Hausdorff-based similarity distance. Lastly, Bouveyron et al. (2017) study households electric consumption monitored for two years and proposed an algorithm that allows to cluster simultaneously households and days of measurements that have the same pattern. This method is based on both functional data analysis and co-clustering techniques that will be developed later in this paper. However, all of these works share one restrictive characteristic: they are limited to the study of only one functional variable.

In the present work, we want to follow three variables: electricity consumption, indoor and outdoor temperature. Such data can be seen as multivariate functional data: multiple quantitative entities evolving along time collected simultaneously for a same individual (Ramsay and Silverman, 2005; Bouveyron et al., 2019). In order to analyze and understand such data, it may be interesting to identify subgroups of individuals (households here) that

have the same profile for these three quantities. For example, to understand the profile of consumption depending on both indoor and outdoor temperatures in order to adapt the electric production or detect poorly insulated buildings. However, analyzing these profiles over a long period of time is in practice complex because it represents a large amount of data. The window of time that practitioners are used to analyze is the day (24 hours). Consequently the observed time period is cut into daily observations. The data set under study is consequently a large table, where rows correspond to households and columns to days, and in which each elements is a set of three curves: electricity consumption, indoor and outdoor temperatures.

To analyze such table, we propose to simultaneously cluster the rows into homogeneous groups of households and the columns into homogeneous groups of days. Such kind of analysis is called co-clustering (Govaert and Nadif, 2013). The co-clustering will result in exhibiting homogeneous blocks of households and days having similar behaviour according to the three functional variables under study.

One of the most famous model for co-clustering is the Latent Block Model (LBM, (Govaert and Nadif, 2013)). According to the LBM, the elements of a block are modeled by a parametric distribution. Each block is therefore interpretable thanks to the block-distribution's parameters. Moreover, model selection criterion, such as the ICL criterion (Biernacki et al., 2000), can be used for model selection purpose, including the choice of the number of co-clusters. This technique proved its efficiency for the co-clustering of numerous types of data: continuous (Nadif and Govaert, 2008), nominal (Bhatia et al., 2014), binary (Laclau et al., 2017), ordinal (Jacques and Biernacki, 2018; Corneli et al., 2019), functional data (Bouveyron et al., 2017; Chamroukhi and Biernacki, 2017; Slimen et al., 2018) or even mixed-type data (Selosse et al., 2019).

Slimen et al. (2018) proposed a co-clustering algorithm based on a vectorial LBM applied on the functional principal components scores of the curves. Bouveyron et al. (2017) extended this work by proposing a functional latent block model assuming that the functional principal components of the curves are block-specific and live into a low-dimensional subspace. Chamroukhi and Biernacki (2017) presented another co-clustering model based on a latent block model where the probability density function is estimated thanks to a regression model with a hidden logistic process. In the present work, a co-clustering algorithm for multivariate functional data is proposed as an extension of Bouveyron et al. (2017) to the multivariate case. Moreover, a more flexible probabilistic model is presented.

The paper is organized as follows. Section 2 presents the co-clustering model and Section 3 its inference. Results of the algorithm on simulated data is presented in Section 4. Section 5 is dedicated to the application on electricity consumption and indoor and outdoor temperatures. Our algorithm succeed in detecting poor insulated buildings and periods of low consumption. Then a discussion concludes the paper in Section 6.

2 A co-clustering model for multivariate functional data

Functional data, which are the observations of a random variable living into a infinite dimensional space, are in practice observed only at a finite set of time points. Consequently, a first step in functional data analysis is the reconstruction of the functional nature of data. This aspect is discussed in the second part of this section, just after having introduced the notations. Then, the proposed co-clustering model for multivariate functional data is presented.

2.1 Data and notations

Let $\boldsymbol{x} = (\boldsymbol{x}_{ij})_{1\leq i\leq n, 1\leq j\leq p}$ be the data table of dimension $n \times p$, where each element $\boldsymbol{x}_{ij}$ is a multivariate curve $\boldsymbol{x}_{ij} = (x_{ij}^1(t), \ldots, x_{ij}^S(t))$ with $t \in [0, T]$. Let us recall that i is the row index, j is the column index and s corresponds to the component of the multivariate curves. In our application i refers to the household, j to the day and s either to electricity consumption, indoor or outdoor temperature. Nevertheless, the model and its inference presented in this work can be used for any other matrix of multivariate functional data.

2.2 Functional data reconstruction

In practice, the functional expressions of the curves $x_{ij}^s(t)$ are not known and we only have access to discrete observations at a finite set of times: $x_{ij}^s(t_1), x_{ij}^s(t_2), \ldots$. A common way to reconstruct the functional form is to assume that the observations can be decomposed into a finite dimensional space spanned by a basis of functions. So each observed curve x_{ij}^s $(1 \leq i \leq n, 1 \leq j \leq p,\ 1 \leq s \leq S)$ can be expressed as a linear combination of basis functions $\{\phi_r^s\}_{r=1,\ldots,R_s}$:

$$x_{ij}^s(t) = \sum_{r=1}^{R_s} c_{ijr}^s \phi_r^s(t)$$

with R_s the number of basis functions. These basis functions can be for instance Fourier bases, spline bases, etc. The automatic choice of this basis (as well as the number of basis functions) is an open problem (Jacques and Preda, 2014a). In practice, this choice is done empirically such that the reconstruction is judged reasonable by the expert. Estimation of the basis expansion coefficients c_{ijr}^s is classically done by least squares smoothing. We refer the reader to Ramsay and Silverman (2005) for a complete survey on this aspect. Let $c = (c_{ij})_{ij}$ be the whole data set of coefficients, where $c_{ij} = (c_{ij1}^1, \ldots, c_{ijR_1}^1, \ldots, c_{ij1}^S, \ldots, c_{ijR_S}^S)$

contains the coefficients for individual i at day j which corresponds to the concatenation of coefficients c_{ijr}^s for all S functional variables. Let $c_i = (c_{ij})_j$ be the coefficients for the ith individual and similarly $c_j = (c_{ij})_i$ the coefficients for day j.

For clarity of the presentation, the same number of basis functions as well as the same basis function $\{\phi_r\}_{r=1,\ldots,R}$ is considered for each component of the multivariate functional variable. But extension is straightforward.

2.3 The proposed latent block model

The aim of a co-clustering model is to define row and column partitions in order to summarize the data matrix $\boldsymbol{x}$ into smaller subgroups, usually called blocks. Let $z = (z_{ik})_{1\leq i\leq n, 1\leq k\leq K}$ be the row partition of the n rows into K groups, and $w = (w_{jl})_{1\leq j\leq p, 1\leq l\leq L}$ the column partition of the p columns into L groups, such as $z_{ik} = 1$ if row i belongs to row-cluster k and 0 otherwise (and similarly for w_{jl}). Thus, one block is defined by a set of curves which belong to a row and column cluster such as $z_{ik}w_{jl} = 1$.

Let us first assume that z and w partitions are independent:

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

where Z the set of all possible rows partitions into K groups and W the set of all possible columns partitions into L groups. Let α_k and β_l be the row and column mixing proportions (belonging to $[0,1]$ and summing to 1), such that $p(z;\theta) = \prod_{ik}\alpha_k^{z_{ik}}$ and $p(w;\theta) = \prod_{jl}\beta_l^{w_{jl}}$. Let us also assume that, conditionally on (z,w), the basis expansion coefficients c_{ij} are independent and generated by a block-specific distribution: $p(c|z,w;\theta) = \prod_{ijkl} p(c_{ij};\theta_{kl})^{z_{ik}w_{jl}}$.

Thus,

$$p(c;\theta) = \sum_{z\in Z}\sum_{w\in W}\prod_{ik}\alpha_k^{z_{ik}}\prod_{jl}\beta_l^{w_{jl}}\prod_{ijkl}p(c_{ij};\theta_{kl})^{z_{ik}w_{jl}} \tag{2}$$

Depending on the grid of the study window, the studied time series can be very long leading to high dimensional coefficients c_{ij}. In order to suggest a parsimonious data modelling, we further suppose that the curves of each block kl ($k = 1, ..., K$, $l = 1, ..., L$) can be described into a low-dimensional functional latent subspace specific to each cluster, with intrinsic dimension $d_{kl} < S \times R$, through a principal component analysis for multivariate functional data (MFPCA, (Jacques and Preda, 2014b)) performed per block.

MFPCA is an extension of PCA for functional data (Ramsay and Silverman, 2005) to the multivariate functional case, representing the multivariate curves by a vector of principal scores into an eigen space formed by multivariate eigen functions. Thus each multivariate curve x_{ij}, conditionally to its belonging to block (k, l), can be represented by its scores $\delta_{ij} = (\delta_{ijr})_{1 \leq r \leq SR}$, with $S \times R$ the maximum number of non null principal components. Let us also define for later use Q_{kl}, a matrix of dimension $SR \times SR$ of eigen functions coefficients, which describes the linear mapping from the original space of c_{ij} to the low-dimensional subspace.

Conditionnaly to the block belonging, the scores are assumed to follow a Gaussian distribution with a parsimonious parametrization of the covariance matrix:

$$\delta_{ijr}^{kl} \sim \mathcal{N}(\mu_{kl}, \Delta_{kl}), \tag{3}$$

with $\mu_{kl} \in \mathbb{R}^{SR}$ and Δ_{kl} the diagonal covariance matrix defined as follows:

$$\Delta_{kl} = \begin{pmatrix} \boxed{\begin{matrix} a_{kl1} & & 0 \\ & \ddots & \\ 0 & & a_{kld_{kl}} \end{matrix}} & \mathbf{0} \\ \mathbf{0} & \boxed{\begin{matrix} b_{kl} & & 0 \\ & \ddots & \\ 0 & & b_{kl} \end{matrix}} \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} d_{kl} \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} SR - d_{kl} \end{matrix}$$

where $a_{kl1} > \ldots > a_{kld_{kl}} > b_{kl}$. With this assumption on Δ_{kl}, the first d_{kl} eigenvalues express the main part of the variability of the data, while the remaining ones reflect the noise and are modeled by a unique parameter b_{kl}. Thus, the space formed by the d_{kl} first eigen functions is a low-dimensional space which contains the main part of information about the data of a given block. The remaining information is considered as noise and modeled by a reduced number of parameters.

Thus, we can infer the distribution of one block curves coefficients according to:

$$c_{ij}|z_{ik}w_{jl} = 1 \sim \mathcal{N}(U_{kl}\mu_{kl}, U_{kl}\Sigma_{kl}U_{kl}^t + \Xi_{kl}), \tag{4}$$

where:

- $Q_{kl} = [U_{kl}, V_{kl}]$ with U_{kl} of size $SR \times d_{kl}$, V_{kl} of size $SR \times (SR - d_{kl})$ with $U_{kl}^t U_{kl} = I_{d_{kl}}$, $V_{kl}^t V_{kl} = I_{SR-d_{kl}}$ and $U_{kl}^t V_{kl} = 0$.
- Σ_{kl} is the matrix $diag(a_{kl1}, ..., a_{kld_{kl}})$.
- Ξ_{kl} is the noise variance matrix of size $SR \times SR$ such that $\Delta_{kl} = Q_{kl}^t(U_{kl}\Sigma_{kl}U_{kl}^t + \Xi_{kl})Q_{kl}$.

Thus, $\theta_{kl} = (\mu_{kl}, a_{kl}, b_{kl}, Q_{kl})$ and the whole set of model parameters is denoted by $\theta = (\alpha_k, \beta_l, \theta_{kl})_{1 \leq k \leq K, 1 \leq l \leq L}$.

In order to provide more parsimonious models, additional assumptions can be made on the different parameters $a_{kl\cdot}$, b_{kl} and d_{kl}, considering that they are common over cluster, over dimension, etc. This approach allows to generate a family of submodels of the general model introduced above. In this paper, we will detail the inference procedure to the submodel assuming $a_{klm} = a_{kl}, \forall m = 1, ..., d_{kl}$, since a good behaviour has been observed in practice. Nevertheless, the co-clustering method presented here can be derived for all models of the family extension, following the approach detailed in Schmutz et al. (2018).

3 Model inference

3.1 Model inference through a SEM-Gibbs algorithm

In co-clustering, the goal is to estimate the unknown row and column partitions z_{ik} and w_{jl}. Usually in mixture model, the maximum a posteriori rule is used, based on the estimation of model parameter θ maximizing the observed log-likelihood:

$$l(\theta; c) = \sum_{z,w} \log p(c; \theta). \tag{5}$$

Proof of this result is provided in Supplementary material A2. LBM implies a double missing structure (z and w) which makes inference harder than in a classical mixture model. Indeed, the use of the well known EM algorithm is not possible because the E step will need the computation of too much terms, which is not tractable in practice. We propose to use the stochastic version of the EM algorithm, embedding a Gibbs sampler for

generating the row and column partitions without having to compute their joint probability distribution (Keribin et al., 2010).

Starting from an initial column partition $w^{(0)}$ and initial parameter value $\theta^{(0)}$, the qth iteration of the partial SEM-Gibbs algorithm alternates between:

- **SE step** (Gibbs sampling): Alternates the two following steps until convergence:
  - simulate $z^{(q+1)}|c, w^{(q)}$ according to:

$$p(z_{ik} = 1|c, w^{(q)}; \theta^{(q)}) \quad = \quad \frac{\alpha_k^{(q)} f_k(c_i|w^{(q)}; \theta^{(q)})}{\sum_{k'} \alpha_{k'}^{(q)} f_{k'}(c_i|w^{(q)}; \theta^{(q)})}$$

    with $f_k(c_i|w^{(q)}; \theta^{(q)}) = \prod_{jl} p(c_{ij}; \theta_{kl}^{(q)})^{w_{jl}^{(q)}}$.
  - simulate $w^{(q+1)}|c, z^{(q+1)}$ according to:

$$p(w_{jl} = 1|c, z^{(q+1)}; \theta^{(q)}) = \frac{\beta_l^{(q)} f_l(c_j|z^{(q+1)}; \theta^{(q)})}{\sum_{l'} \beta_{l'}^{(q)} f_{l'}(c_j|z^{(q+1)}; \theta^{(q)}).}$$

    with $f_l(c_j|z^{(q+1)}; \theta^{(q)}) = \prod_{ik} p(c_{ij}; \theta_{kl}^{(q)})^{z_{ik}^{(q)}}$.

- **M step**: Update of $\theta^{(q+1)}$. The update of each parameter can be done in the same way than Schmutz et al. (2018). The mixing proportion and the block mean are updated by:
  - $\alpha_k^{(q+1)} = \frac{1}{n} \sum_i z_{ik}^{(q+1)}$ and $\beta_l^{(q+1)} = \frac{1}{p} \sum_j w_{jl}^{(q+1)}$,
  - $\mu_{kl}^{(q+1)} = \frac{1}{n_{kl}^{(q+1)}} \sum_{ij} c_{ij}^{z_{ik}^{(q+1)} w_{jl}^{(q+1)}}$ with $n_{kl}^{(q+1)} = \sum_{ij} z_{ik}^{(q+1)} w_{jl}^{(q+1)}$.

For the variance parameters a_{kl}, b_{kl} and Q_{kl}, let us define the sample covariance matrix C_{kl} of block kl:

$$C_{kl}^{(q)} = \frac{1}{n_{kl}^{(q)}} \sum_{i=1}^{n} \sum_{j=1}^{p} z_{ik}^{(q+1)} w_{jl}^{(q+1)} (c_{ij} - \mu_{kl}^{(q)})^t (c_{ij} - \mu_{kl}^{(q)}).$$

Let also introduce W the $SR \times SR$ matrix of inner products: $W = \int_0^T \phi^t(t)\phi(t)dt$, with $\phi(t)$ the matrix that gathers basis functions of all S functional variables:

$$\phi(t) = \begin{pmatrix} \phi_{11}(t) & ... & \phi_{1R}(t) & 0 & ... & 0 & ... & 0 & ... & 0 \\ 0 & ... & 0 & \phi_{21}(t) & ... & \phi_{2R}(t) & ... & 0 & ... & 0 \\ & & & & ... & & & & & \\ 0 & ... & 0 & 0 & ... & 0 & ... & \phi_{S1}(t) & ... & \phi_{SR}(t) \end{pmatrix}.$$

Then,

- the d_{kl} first columns of the matrix of eigen functions coefficients $Q_{kl}^{(q)}$ are updated by the eigen functions coefficients associated with the largest eigenvalues of $W^{1/2}C_{kl}^{(q)}W^{1/2}$,
- the variance parameters $a_{kl}^{(q+1)}$, are updated by the mean of the d_{kl} largest eigenvalues of $W^{1/2}C_{kl}^{(q)}W^{1/2}$,
- the variance parameters b_{kl} are updated by $\frac{1}{R-d_{kl}}(trace(W^{1/2}C_{kl}^{(q)}W^{1/2}) - d_{kl}a_{kl}^{(q)})$.

Proofs of those results are available in Supplementary material A3, A4 and A5.

In brief, the SEM-Gibbs algorithm is run for a given number of iterations. After a burn-in period, the final estimation $\hat{\theta}$ of the parameters is obtained by the mean of the sample distribution (without the burn-in iterations). Then, a new Gibbs sampler is used to sample $(\hat{z}, \hat{w})$ according to $\hat{\theta}$, and the final partition $(\hat{z}, \hat{w})$ is obtained by the marginal mode of this sample distribution.

Initialization of the algorithm As said previously, our algorithm relies on a SEM-Gibbs algorithm. This algorithm needs to be initialized with values for column partitions and parameters. In practice, column and row partitions are initialized, and corresponding

initial parameter values are deduced. To this end, three initialization strategies are available here : random, *k-means* and *funFEM*. In the random case, partitions are randomly sampled from a multinomial distribution with uniform probabilities. *k-means* strategy consists of initializing partitions with those obtained by *k-means* method directly applied on discretized data. Finally, *funFEM* aims to initialize partitions by applying the *funFEM* algorithm (Bouveyron et al., 2015). We will see later, in the numerical experimentation section, that *funFEM* is the one that gives the best results.

3.2 Choice of the number of clusters

We now discuss the choice of the hyper-parameters K and L, *i.e.* the number of row clusters and column clusters respectively. The choice of these hyper-parameters is viewed here as a model selection problem. Well established model selection tools are Akaike information criterion (AIC, (Akaike, 1974)), Bayesian information criterion (BIC, (Schwarz, 1978)) and Integrated Classification Likelihood (ICL, (Biernacki et al., 2000)). However, in the co-clustering case, the likelihood is not tractable for the same reason than the EM algorithm is not usable. Consequently, AIC and BIC are not tractable. Conversely, the ICL criterion can be considered since it relies on the completed log-likelihood, which is tractable. Adapted to our model, the ICL criterion is:

$$ICL(K,L) = \log p(c, \hat{z}, \hat{w}; \hat{\theta}) - \frac{K-1}{2} log(n) - \frac{L-1}{2} log(p) - \frac{\nu}{2} log(np)$$

where $\nu = KLSR + 2KL + \sum_{kl} d_{kl}(SR - \frac{d_{kl}+1}{2})$ is the number of continuous parameters per block and

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

The couple (K, L) leading to the highest ICL value is selected as the most appropriate number of row and column clusters.

4 Numerical experimentation on simulated data

This section presents numerical experiments on simulated data in order to illustrate the behavior of the proposed methodology in presence of different noise ratio in data and to study the selection of the number of row and column clusters. The R code for our multivariate functional co-clustering algorithm is available under request and will be soon available on CRAN as an R package.

For all examples, we set to 50 iterations the burn-in period of the algorithm and the SEM-Gibbs maximal number of iterations is set to 100.

4.1 Introductory example

4.1.1 Simulation setup

A sample of $n = 100$ bivariate curves are simulated with $K = 4$, $L = 3$ and $p = 100$. The proportions of row clusters α used is (0.2,0.4,0.1,0.3) and column clusters β is (0.4,0.3,0.3). The first functional variable is designed from four different functions that are used as blocks mean at 31 equi-spaced time points, $t = 0, 1/30, 2/30, ..., 1$:

$$x_{ij}(t)|z_{ik}w_{jl} = 1 \sim \mathcal{N}(m_{kl}(t), s^2),$$

where $s = 0.3$ and the mean function is taken from $m_{11} = m_{21} = m_{33} = m_{42} = f_1$, $m_{12} = m_{22} = m_{31} = f_2$, $m_{13} = m_{32} = f_3$ et $m_{23} = m_{41} = m_{43} = f_4$, with $f_1(t) = sin(4\pi t)$, $f_2(t) = 0.75 - 0.5\mathbb{1}_{t\in]0.7,0.9[}$, $f_3(t) = h(t)/max(h(t))$ where $h(t) = \mathcal{N}(0.2, \sqrt{0.02})$ and $f_4(t) =$

$sin(10\pi t)$. Then the second variable is designed according to the same process than the first one but with four different functions: $f_1(t) = cos(4\pi t)$, $f_2(t) = 0.75 - 0.5\mathbb{1}_{t\in]0.2,0.4[}$, $f_3(t) = h(t)/max(h(t))$ where $h(t) = \mathcal{N}(0.2, \sqrt{0.05})$ and $f_4(t) = cos(10\pi t)$. The block means functions are shown in Figure 1.

Starting from this simulation setting, five scenarios are build by adding some noise fraction within the blocks by randomly simulating a percentage τ of curves using other block means: 0% (scenario 1), 10% (scenario 2), 30% (scenario 3), 50% (scenario 4) and 80% (scenario 5).

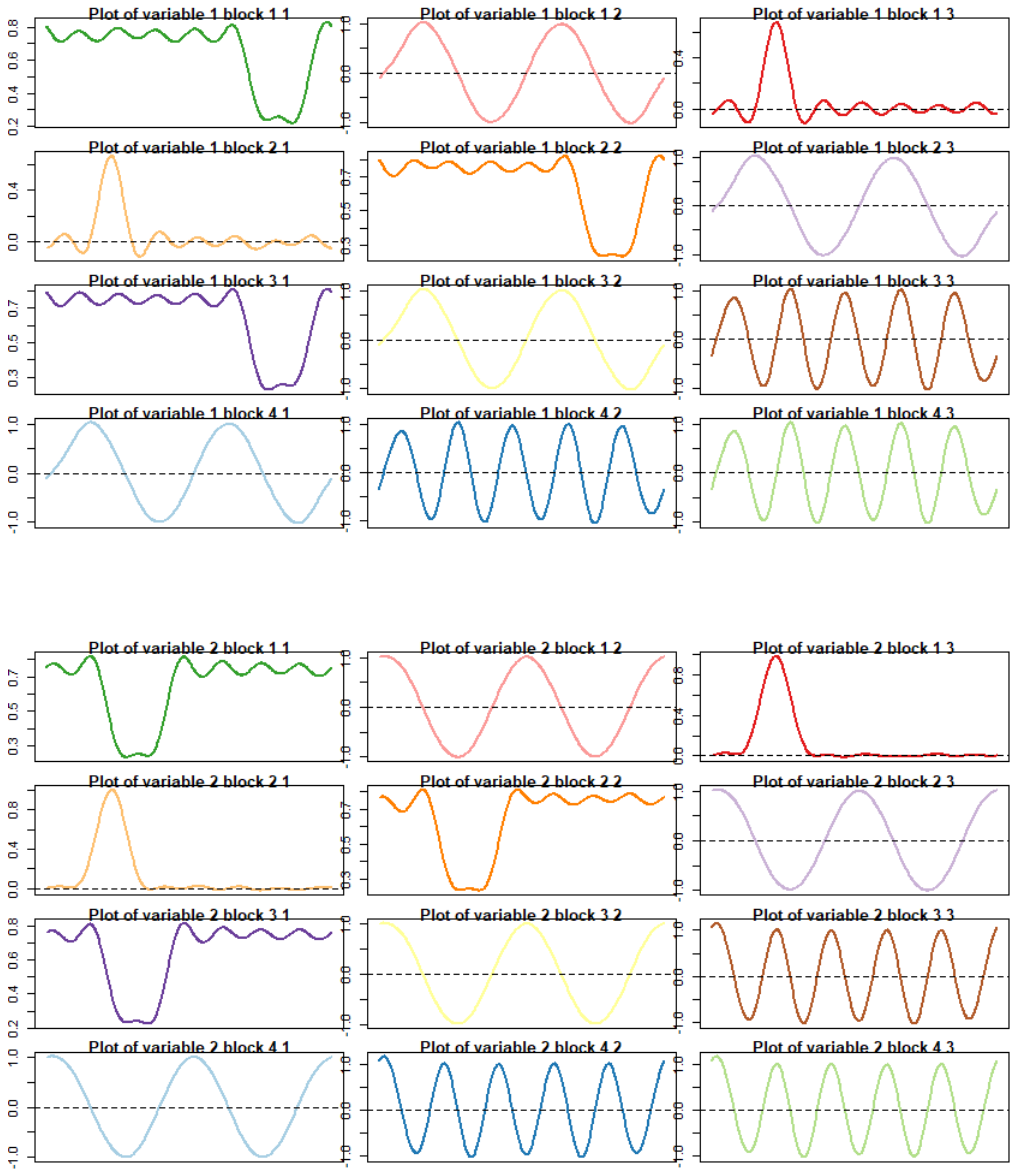

Figure 1: Block means functions for the first variable (top) and second variable (bottom)

4.1.2 Results

In order to illustrate the good behaviour of our algorithm in the case of noisy data and to compare the influence of algorithm initialization on co-clustering results, 20 simulations have been performed for each scenario with both *k-means*, *funFEM* and *random* initializations. The algorithm is applied for $K = 4$ and $L = 3$ and with Fourier smoothing with 15 basis functions. The quality of estimated partitions is assessed with the Adjusted Rand Index (ARI, (Rand, 1971)).

Results are shown in Figure 2. We can see that co-clustering results are almost perfect for the 4 first scenarios with *funFEM* initialization, and the 2 first scenarios for *k-means* initialization. As expected, the algorithm performance decreases while noise increases, but median ARI value is always above 0.8 in the case of four first scenarios with *k-means* and *funFEM* initialization. Moreover *k-means* initialization performs better than random when the noise ratio is upper than 50%. And the *funFEM* initialization performs better than the *k-means* one.

To conclude, in view of the good behaviour of *funFEM* initialization in both previous examples, we recommend to use firstly *funFEM* initialization rather than the two others available in the algorithm.

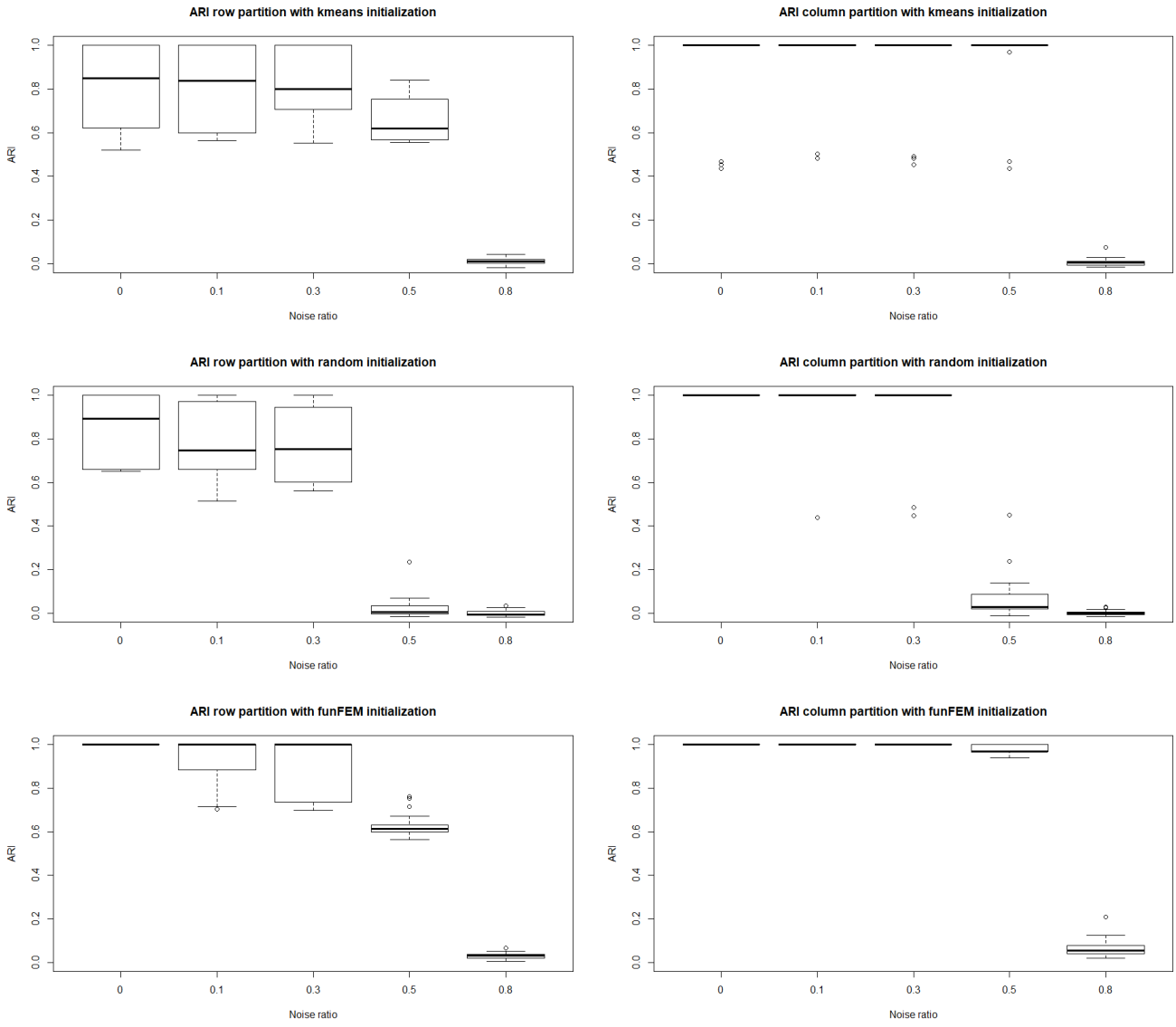

Figure 2: Results of ARI for each scenario for *k-means* (top), random (middle) and *funFEM* (bottom) initialization

Table 1: Number of times each partition is selected by ICL for 20 simulations (in percent). Highlighted rows and columns correspond to the actual values for K and L.

Scenario $\tau = 0$						Scenario $\tau = 0.1$					
K/L	2	3	4	5	6	K/L	2	3	4	5	6
2	0	0	0	0	0	2	0	0	0	0	0
3	0	0	0	0	0	3	0	0	0	0	0
4	0	**100**	0	0	0	4	0	**100**	0	0	0
5	0	0	0	0	0	5	0	0	0	0	0
6	0	0	0	0	0	6	0	0	0	0	0

Scenario $\tau = 0.3$						Scenario $\tau = 0.5$					
K/L	2	3	4	5	6	K/L	2	3	4	5	6
2	0	0	0	0	0	2	0	0	0	0	0
3	0	0	0	0	0	3	0	0	0	0	0
4	0	**100**	0	0	0	4	0	**90**	0	0	0
5	0	0	0	0	0	5	0	10	0	0	0
6	0	0	0	0	0	6	0	0	0	0	0

4.2 Model selection

In this section, the selection of the number of clusters is investigated. Data are generated as previously. The simulation setting is repeated 20 times with $n = 500$, $p = 500$ and $t = 30$. The algorithm is run for 2 to 6 clusters with *funFEM* initialization and the best model selected by ICL is noted down. Results are shown in Table 1.

Model selection with *funFEM* is perfect with a noise ratio from 0 to 30% of the data

volume. Then, as expected, the performance of the criterion decreases, for a noise ratio of 50% the ICL criterion gets back to the true partition in 90% of cases.

5 Co-clustering for energy waste detection in French households

This section focus on the analysis of French households' electric consumption according to their indoor and outdoor temperatures. With more and more involvement of the mainstream on environmental issues and with the COP24 aim of evaluating government efforts to tackle against global warming, it is important to find ways to save energy. This type of data can help detecting potentially poor insulated homes and target those households with rehabilitation offers which would lead to less energy waste.

5.1 Data

This data set deals with electric consumption, indoor temperature and outdoor temperature of 356 French households representatives of metropolitan France. It has been provided by EDF company, a French electricity provider. Data are collected every 30 minutes for 173 days (48 time points per day, cf. Figure 3), from July 2009 to December 2009. Their objectives are to model insulation efficiency, detect weeks of absence in order to adapt electric load and evaluate the impact of indoor temperature on the electric consumption. The data set contains some missing observations. Since the first step of our analysis consists in estimating the basis expansion approximations of the curves, it is not a problem if there is some missing observations (except if they are at the beginning or at the end of the

period). Indeed, smoothing the data into a basis expansion can be performed even if some of the 48 time points are not observed. Nevertheless, we remove from the analysis daily curves with too many missing time points (more than 41) or for which missing values occur at the beginning or at the end of the period.

The functional form of the whole database is reconstructed using a Fourier smoothing with 15 basis functions. Our algorithm has been applied with *funFEM* initialization with a varying number of row and column clusters, from 2 to 20 on normalized data. The ICL criteria is used to choose an appropriate number of row and column clusters.

Table 2: ICL values for the 10 first partitions

Number of row clusters K	Number of column clusters L	ICL
5	14	-3944668
5	13	-3946480
5	11	-3951143
4	14	-3958896
3	13	-3961862
5	10	-3963644
4	12	-3966577
3	14	-3967567
6	8	-3968532
6	9	-3970104

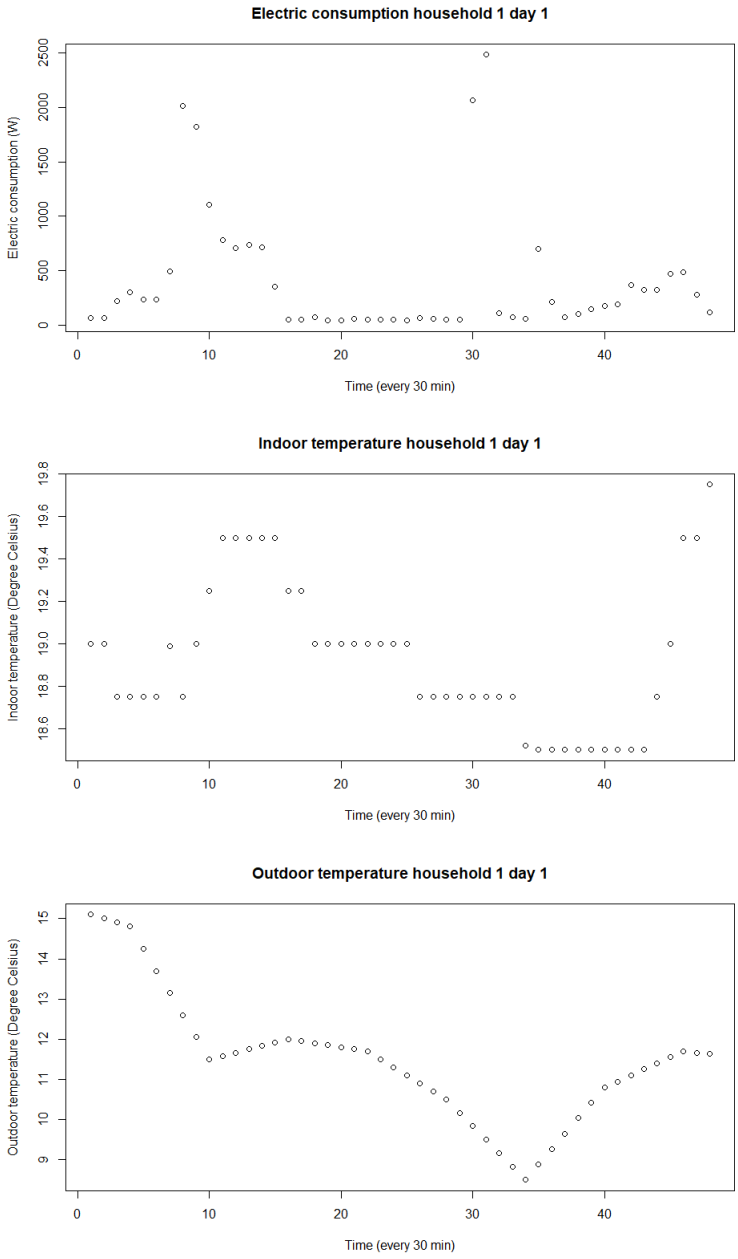

Figure 3: Raw data for one individual and one day of measurement

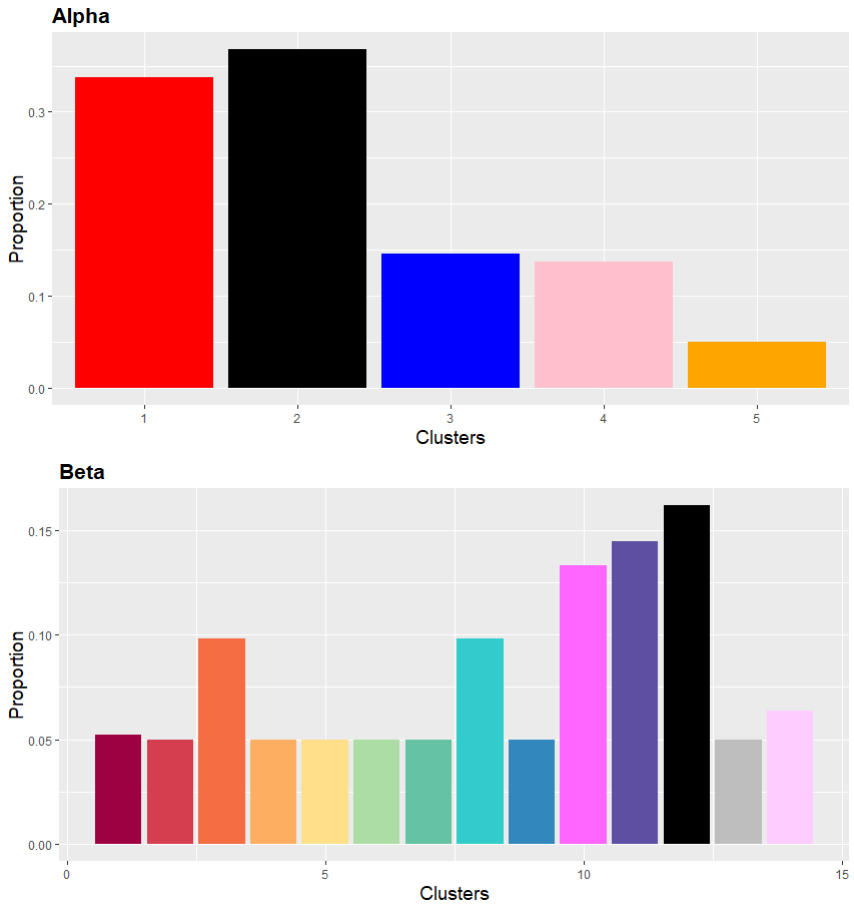

Figure 4: Proportion of households in row clusters (top) and days in column clusters (bottom)

5.2 Results

According to ICL, the best partition is with 5 row clusters and 14 column clusters (cf. Table 2). Clusters proportions can be seen on Figure 4.

The obtained block clusters can be described with their mean curves. Since the number of column clusters is large, we focus our interpretation on four typical column clusters in order to ease the analysis: Figure 5 plots the mean curves for row-cluster 1 to 5 and column cluster 1, 6, 10 and 14. The whole set of mean curves is plotted in supplementary material A1 (Figures 10, 11 and 12).

On average French households indoor temperature is between 17°C and 25°C over the 6 months measurement period (Figure 5) and whatever the geographical location (Figure 8). Presumably it is more dictated by the way of life than by the area climate type. For instance, the third cluster of households (row-cluster) is the one with the lowest electric consumption (third row on Figure 7) although it is distributed all over the country (blue points on Figure 8). This cluster may correspond to well insulated houses or households whose heating system is not electrical. The first and second cluster of households (red and black points on Figure 8) have a medium consumption (first and second rows on Figure 7) and are distinguished by the outdoor temperature (first and second rows on Figure 6) which is a bit more flat for the first cluster (red): the delta of temperature for one day is less important for red households on the measured period than for black ones. The fifth cluster of households (orange points on Figure 8) have a high and uncontrolled electricity consumption (last row on Figure 7) and an indoor temperature well below the one of other clusters on the whole time period (last row on Figure 5). We may suppose that those households belong to retired persons who have periods of extended vacations because they are the only one where the indoor temperature stays at 15°C during 4 time periods. Those

households are also the only ones with a very variable aspect of the indoor temperature mean curves and consumption curves, it may be explained by switching on and off their heating system many times along a day, whereas other households may be equipped with a timer that holds the wanted temperature. Lastly, the fourth cluster of households (pink points on Figure 8) has a very high electricity consumption (about twice that of other clusters) whereas the outdoor temperature is not especially lower (fourth row on Figure 6) and the indoor temperature not especially greater (between 18°C to 22°C, fourth row on Figure 5). Those households may be poor insulated and it may be important to alert them on their consumption in order to assist them in reducing their invoices by undertaking renovations to their property and also to reduce their ecological footprint.

Finally, Figure 9 presents the distribution of the column clusters (clusters of days) along the period under study. Note that there is no link between the color of the row and column clusters. We can see that the pink cluster (14th column cluster) corresponds to the coldest days of the period (outdoor temperature below 5°C). For those days the electric consumption of all households is higher than their average one on the other periods (last column on Figure 7).

To the contrary, the hottest days where mainly in July (first and second column cluster on Figure 12). We can see that the maximum electric consumption reached in a day does not differ a lot between those days and days of medium outdoor temperatures. We suppose that for these periods the main consumption areas are other electric devices than the heating system.

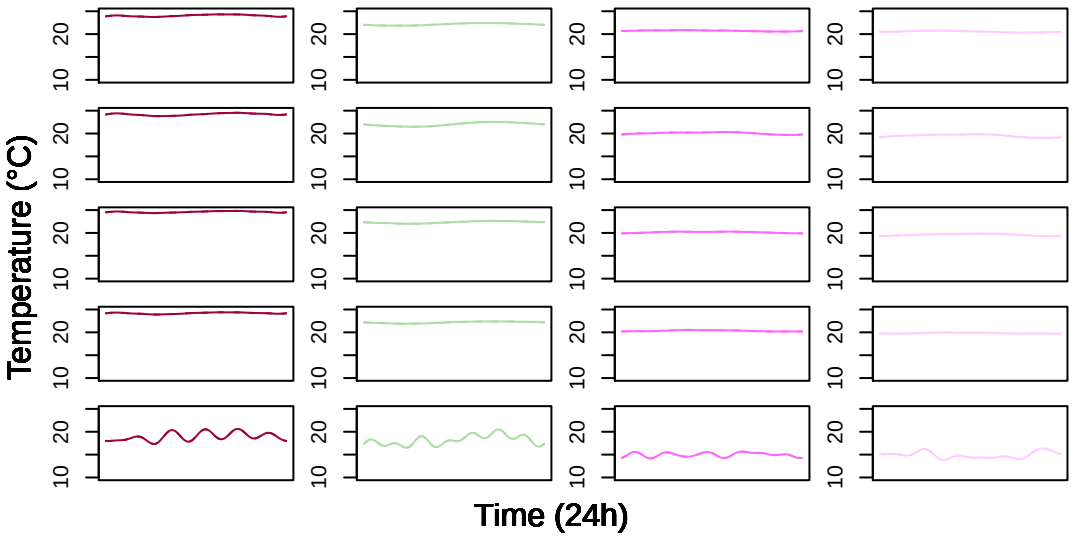

Figure 5: Co-clustering results for indoor temperature for column cluster 1, 6, 10 and 14

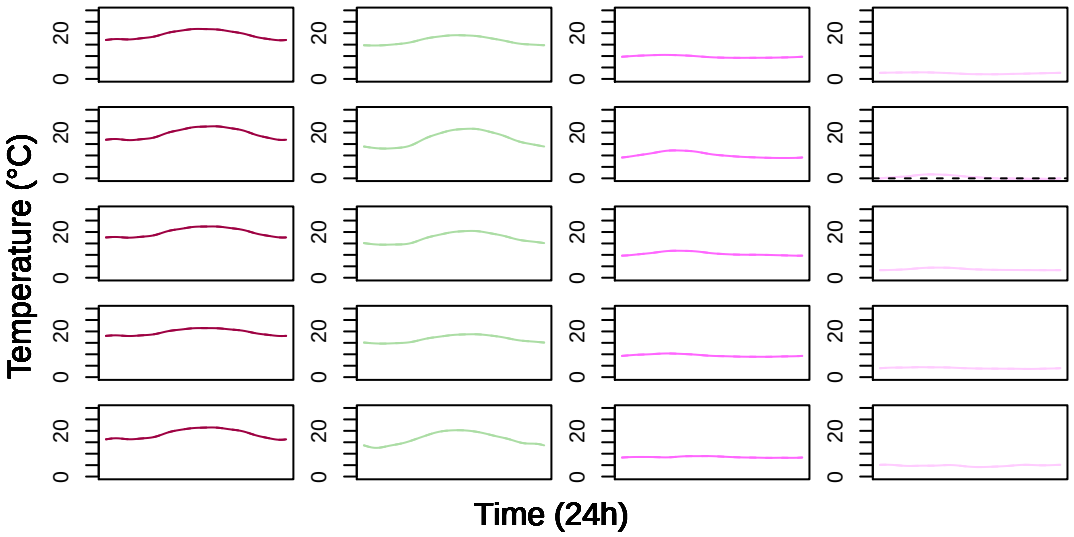

Figure 6: Co-clustering results for outdoor temperature for column cluster 1, 6, 10 and 14

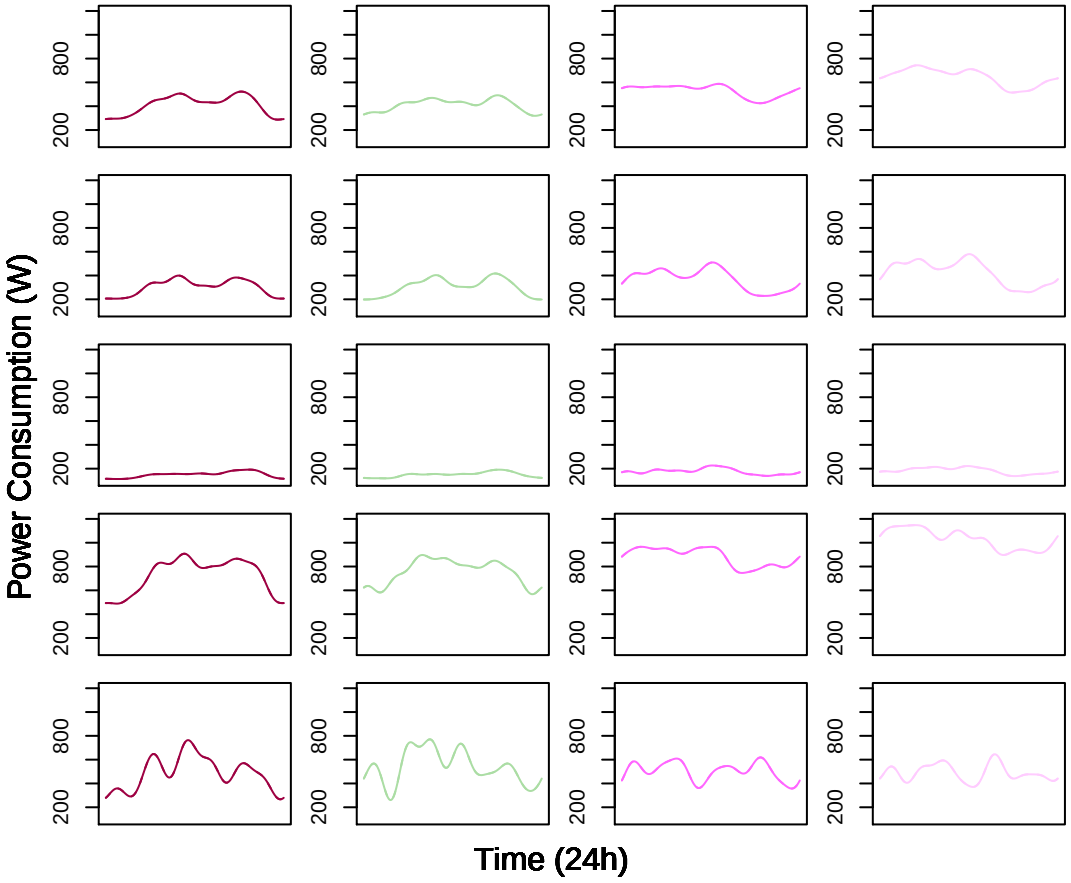

Figure 7: Co-clustering results for electric consumption for column cluster 1, 6, 10 and 14

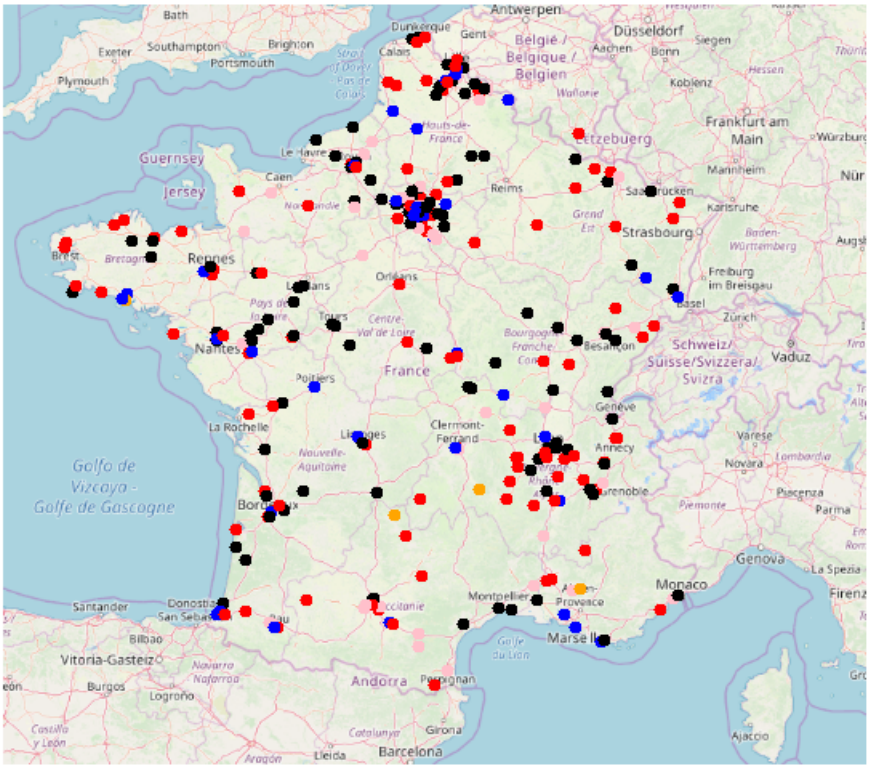

Figure 8: Geographical position of households by row clusters (colored by cluster)

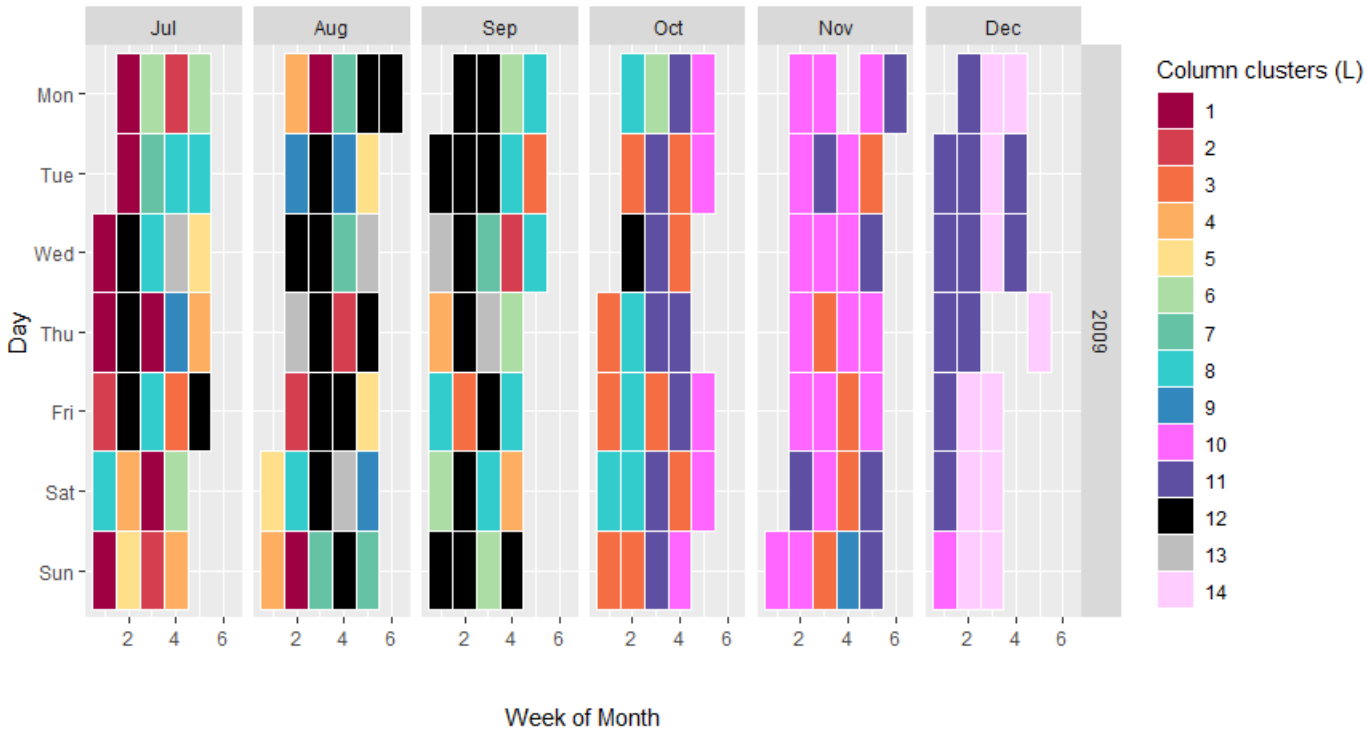

Figure 9: Calendar position of column clusters (colored by cluster)

To conclude, the analysis of this large multivariate functional data set with the co-clustering model has allowed the distinction of different habits profiles. Those results have especially highlighted that French households' electric consumption and indoor temperature do not depend on the geographical location and thus on the outdoor temperature, but more on the way of life of people and the insulation level. Actions can be undertaken to further raise awareness on the impact of reducing their environmental footprint by reducing their indoor temperature to a maximum value of, at least, 21°C. Moreover some households have been identified by their abnormally high electric consumption, those houses should be targeted by the electric provider with rehabilitation offers to reduce their energy waste.

6 Discussion and conclusion

This work was motivated by the will to detect poor insulated buildings and long periods of absence in order to reduce energy waste. The proposed co-clustering model answers both objectives, which will allow EDF company to alert users on their over-consumption and advise them on renovation work. Furthermore, the field of operational applications of co-clustering methods for multivariate functional curves is wide at EDF company. For instance, it can help design new marketing offers or services such as co-clustering load curves and photo-voltaic production data in order to identify clusters with low yields or detect anomalies. In the future, EDF will have more and more data coming from connected devices, like connected thermostats or connected weather stations, this data flood deserves to be classified to ease it interpretation. The proposed algorithm answers this need. The co-clustering method for multivariate functional data, allows to cluster both individuals and variables simultaneously. The proposed approach relies on a functional latent

block model, which assumes for each block a probabilistic distribution for the scores of the multivariate curves obtained from a multivariate functional principal component analysis. Model inference relies on a SEM-Gibbs algorithm which alternates a SE-step where row and column partitions are simulated according to Gibbs algorithm, and a M-step where model parameters are updated thanks to the previous simulated partitions. Lastly, the best number of row and column clusters is selected thanks to the ICL criterion. As far as the authors know, it is the first algorithm available for functional multivariate co-clustering.

Acknowledgements The authors would like to thank the LabCom 'CWD-VetLab' for its financial support. The LabCom 'CWD-VetLab' is financially supported by the Agence Nationale de la Recherche (contract ANR 16-LCV2-0002-01).

This research has also benefited from the support of the "FMJH Research Initiative Data Science for Industry".

SUPPLEMENTARY MATERIAL

A1. Co-clustering results for $K = 5$ and $L = 14$

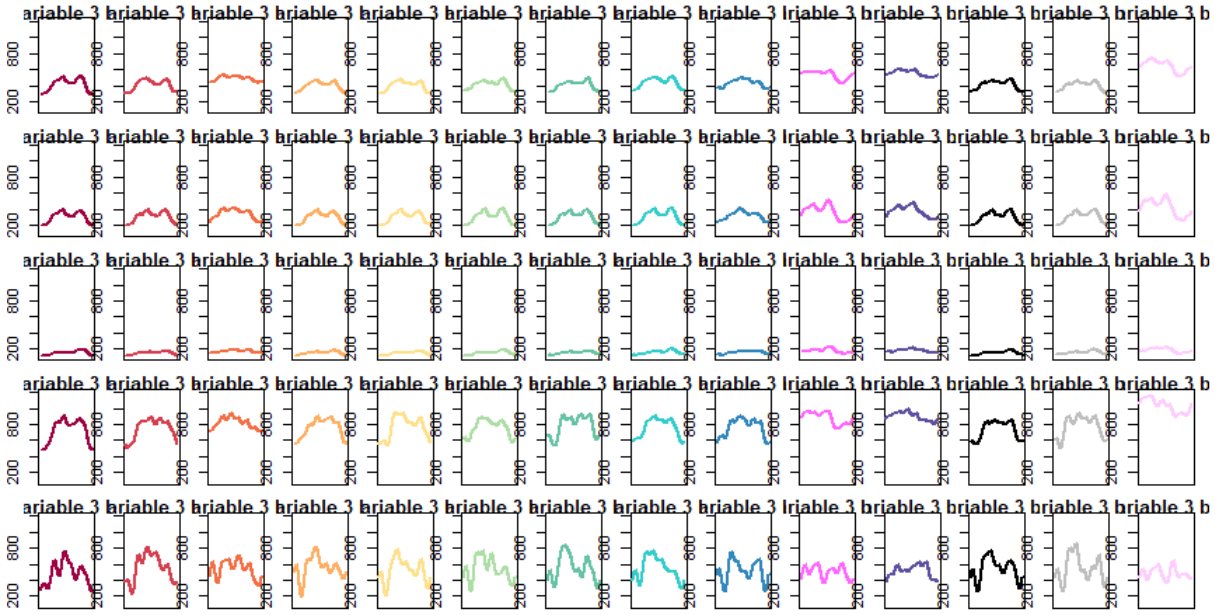

Figure 10: Co-clustering results for electric consumption

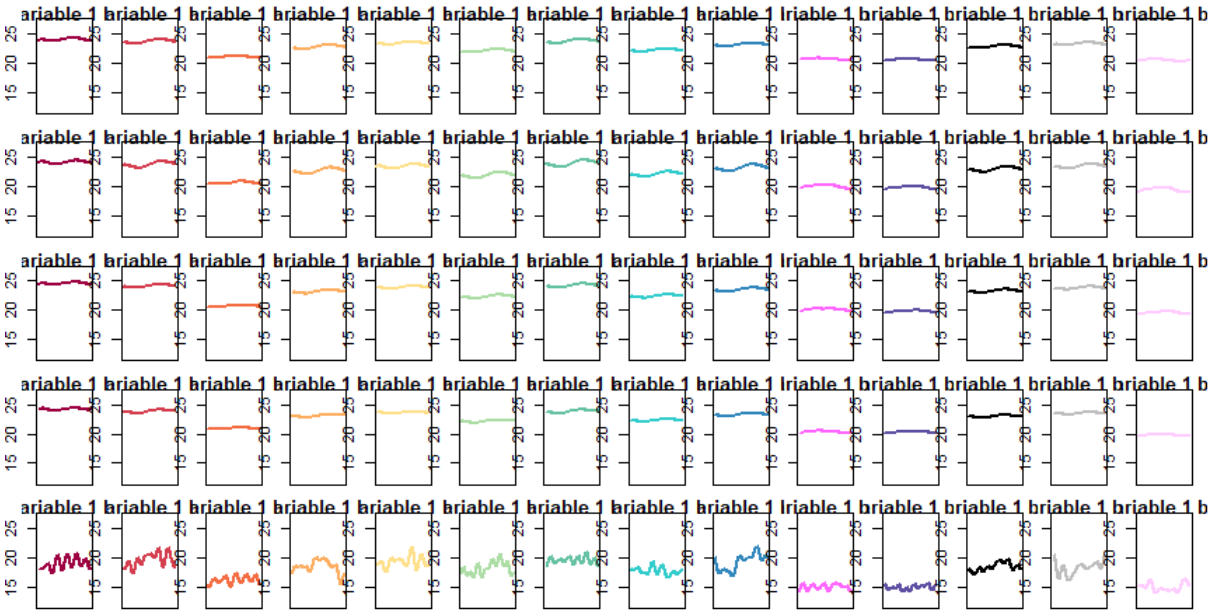

Figure 11: Co-clustering results for indoor temperature

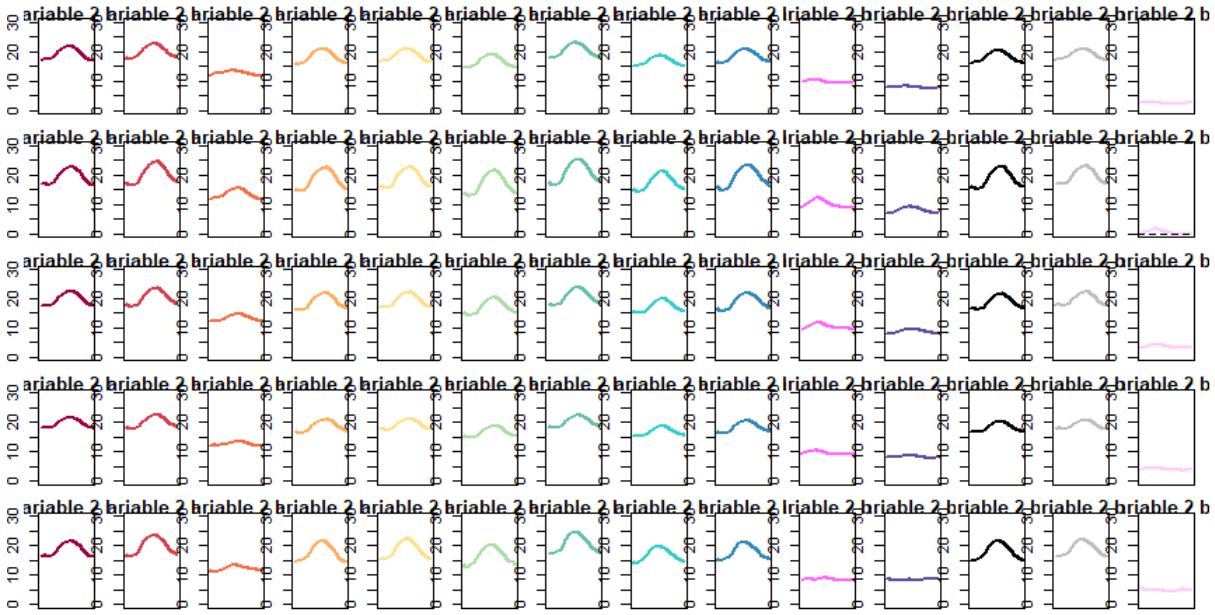

Figure 12: Co-clustering results for outdoor temperature

A2. Log-likelihood formula, proof of Equation 5:

$$
\begin{aligned}
l(\theta) &= \sum_{i=1}^{n}\sum_{j=1}^{SR}\sum_{k=1}^{K}\sum_{l=1}^{L} z_{ik}w_{jl}log(\pi_{kl}f(c_{ij},\theta_{kl})) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{SR}\sum_{k=1}^{K}\sum_{l=1}^{L} z_{ik}w_{jl}log(\frac{\pi_{kl}}{(2\pi)^{SR/2}|\Sigma_{kl}|^{1/2}}exp(-\frac{1}{2}(c_{ij}-\mu_{kl})^t\Sigma^{-1}(c_{ij}-\mu_{kl}))) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{SR}\sum_{k=1}^{K}\sum_{l=1}^{L} z_{ik}w_{jl}[log(\pi_{kl})-\frac{1}{2}log|\Sigma_{kl}|-\frac{SR}{2}log(2\pi)-\frac{1}{2}(c_{ij}-\mu_{kl})^t\Sigma^{-1}(c_{ij}-\mu_{kl})] \\
&= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{SR}\sum_{k=1}^{K}\sum_{l=1}^{L} z_{ik}w_{jl}[-2log(\pi_{kl})+d_{kl}log(a_{kl})+(SR-d_{kl})log(b_{kl}) \\
&+ (c_{ij}-\mu_{kl})^t\Sigma^{-1}(c_{ij}-\mu_{kl})] - \frac{nSR}{2}log(2\pi)
\end{aligned}
$$

Let $n_{kl}=\sum_{i=1}^{n}\sum_{j=1}^{SR} z_{ik}w_{jl}$ be the number of block kl curves,

$$
\begin{aligned}
l(\theta) &= -\frac{1}{2}\sum_{k=1}^{K}\sum_{l=1}^{L} n_{kl}[-2log(\pi_{kl})+d_{kl}log(a_{kl})+(SR-d_{kl})log(b_{kl}) \\
&+ \frac{1}{n_{kl}}\sum_{i=1}^{n}\sum_{j=1}^{SR} z_{ik}w_{jl}(c_{ij}-\mu_{kl})^tQ_{kl}\Delta_{kl}^{-1}Q_{kl}^t(c_{ij}-\mu_{kl})] - \frac{nSR}{2}log(2\pi)
\end{aligned}
$$

The quantity $(c_{ij}-\mu_{kl})^tQ_{kl}\Delta_{kl}^{-1}Q_{kl}^t(c_{ij}-\mu_{kl})$ is a scalar, so it is equal to it trace:

$$
\frac{1}{n_{kl}}\sum_{i=1}^{n}\sum_{j=1}^{SR} z_{ik}w_{jl}(c_{ij}-\mu_{kl})^tQ_{kl}\Delta_{kl}^{-1}Q_{kl}^t(c_{ij}-\mu_{kl}) = \frac{1}{n_{kl}}\sum_{i=1}^{n}\sum_{j=1}^{SR} z_{ik}w_{jl}tr((c_{ij}-\mu_{kl})^tQ_{kl}\Delta_{kl}^{-1}Q_{kl}^t(c_{ij}-\mu_{kl})).
$$

Well $tr([(c_{ij}-\mu_{kl})^tQ_{kl}]\times[\Delta_{kl}^{-1}Q_{kl}^t(c_{ij}-\mu_{kl})]) = tr([\Delta_{kl}^{-1}Q_{kl}^t(c_{ij}-\mu_{kl})]\times[(c_{ij}-\mu_{kl})^tQ_{kl}])$,

consequently:

$$\begin{aligned}
& \frac{1}{n_{kl}} \sum_{i=1}^{n} \sum_{j=1}^{SR} z_{ik} w_{jl} (c_{ij} - \mu_{kl})^t Q_{kl} \Delta_{kl}^{-1} Q_{kl}^t (c_{ij} - \mu_{kl}) \\
= & \frac{1}{n_{kl}} \sum_{i=1}^{n} \sum_{j=1}^{SR} z_{ik} w_{jl} tr(\Delta_{kl}^{-1} Q_{kl}^t (c_{ij} - \mu_{kl})(c_{ij} - \mu_{kl})^t Q_{kl}) \\
= & tr(\Delta_{kl}^{-1} Q_{kl}^t [\frac{1}{n_{kl}} \sum_{i=1}^{n} \sum_{j=1}^{SR} z_{ik} w_{jl} (c_{ij} - \mu_{kl})^t (c_{ij} - \mu_{kl})] Q_{kl}) \\
= & tr(\Delta_{kl}^{-1} Q_{kl}^t C_{kl} Q_{kl}),
\end{aligned}$$

where $C_{kl} = \frac{1}{n_{kl}} \sum_{i=1}^{n} \sum_{j=1}^{SR} z_{ik} w_{jl} (c_{ij} - \mu_{kl})^t (c_{ij} - \mu_{kl})$ is the empirical covariance matrix of the k-th element of the mixture model. The Δ_{kl} matrix is diagonal, so we can write:

$$\begin{aligned}
\frac{1}{n_{kl}} \sum_{i=1}^{n} \sum_{j=1}^{SR} z_{ik} w_{jl} (c_{ij} - \mu_{kl})^t Q_{kl} \Delta_{kl}^{-1} Q_{kl}^t (c_{ij} - \mu_{kl}) &= \sum_{j=1}^{d_{kl}} \frac{q_{klj}^t W^{1/2} C_{kl} W^{1/2} q_{klj}}{a_{klj}} \\
&+ \sum_{j=d_{kl}+1}^{SR} \frac{q_{klj}^t W^{1/2} C_{kl} W^{1/2} q_{klj}}{b_{kl}},
\end{aligned}$$

where q_{klj} is j-th column of Q_{kl}.

Finally,

$$\begin{aligned}
l(\theta) = & -\frac{1}{2} \sum_{k=1}^{K} \sum_{l=1}^{L} n_{kl} [-2log(\pi_{kl}) + d_{kl} log(a_{kl}) + (R - d_{kl}) log(b_{kl}) \\
& + \sum_{j=1}^{d_{kl}} \frac{q_{klj}^t W^{1/2} C_{kl} W^{1/2} q_{klj}}{a_{klj}} + \sum_{j=d_{kl}+1}^{SR} \frac{q_{klj}^t W^{1/2} C_{kl} W^{1/2} q_{klj}}{b_{kl}}] + \frac{nSR}{2} log(2\pi)
\end{aligned}$$

A3. Parameter Q_{kl} update:

We have to maximize the log-likelihood under the constraint $q_{klj}^t q_{klj} = 1$, with q_{klj} the jth column of Q_{kl}. This is equivalent to look for a saddle point of the Lagrange function:

$$\mathcal{L} = -2l(\theta) - \sum_{j=1}^{SR} \omega_{klj}(q_{klj}^t q_{klj} - 1)$$

where ω_{klj} are Lagrange multipliers. Thus, we can write:

$$\begin{aligned}
\mathcal{L} &= \sum_{k=1}^{K}\sum_{l=1}^{L} n_{kl}[d_{kl}log(a_{kl}) + \sum_{j=1}^{d_{kl}} \frac{q_{klj}^t W^{1/2} C_{kl} W^{1/2} q_{klj}}{a_{klj}} \\
&+ (SR - d_{kl})log(b_{kl}) + \sum_{j=d_{kl+1}}^{SR} \frac{q_{klj}^t W^{1/2} C_{kl} W^{1/2} q_{klj}}{b_{klj}} - 2log(\pi_{kl})kl)] \\
&+ \frac{nSR}{2} log(2\pi) - \sum_{j=1}^{SR} \omega_{klj}(q_{kj}^t q_{klj} - 1).
\end{aligned}$$

Therefore, the gradient of $\mathcal{L}$ in relation to q_{kjl} is:

$$\begin{aligned}
\nabla_{q_{klj}}\mathcal{L} &= \nabla_{q_{klj}}(\sum_{k=1}^{K}\sum_{l=1}^{L} n_{kl}[\sum_{j=1}^{d_{kl}} \frac{q_{klj}^t W^{1/2} C_{kl} W^{1/2} q_{klj}}{a_{klj}} + \sum_{j=d_{kl}+1}^{SR} \frac{q_{klj}^t W^{1/2} C_{kl} W^{1/2} q_{klj}}{b_{kl}}] \\
&- \sum_{j=1}^{SR} \omega_{klj}(q_{klj}^t q_{klj} - 1)).
\end{aligned}$$

As a reminder, when W is symmetric, then $\frac{\partial}{\partial x}(x-s)^T W(x-s) = 2W(x-s)$ and $\frac{\partial}{\partial x}(x^T x) = 2x$ (cf. Petersen and Pedersen (2012)), so:

$$\nabla_{q_{klj}}\mathcal{L} = n_{kl}[2\frac{W^{1/2} C_{kl} W^{1/2}}{\sigma_{klj}} q_{klj}] - 2\omega_{klj} q_{klj}$$

where σ_{klj} is the j-th diagonal term of matrix Δ_k.

Thus,

$$\begin{aligned} q_{klj}^t \nabla_{q_{klj}} \mathcal{L} = 0 \quad &\Leftrightarrow \quad \omega_{klj} q_{klj} = \frac{n_{kl}}{\sigma_{klj}} q_{klj}^t W^{1/2} C_{kl} W^{1/2} q_{klj} \\ &\Leftrightarrow \quad W^{1/2} C_{kl} W^{1/2} q_{klj} = \frac{\omega_{klj}\sigma_{klj}}{n_{kl}} q_{klj}. \end{aligned}$$

q_{klj} is the eigenfunction of $W^{1/2}C_{kl}W^{1/2}$ which match the eigenvalue $\lambda_{klj} = \frac{\omega_{klj}\sigma_{klj}}{n_{kl}} = W^{1/2}C_{kl}W^{1/2}$. We can write $q_{klj}^t q_{klm} = 0$ if $j \neq m$. So the log-likelihood can be written:

$$\begin{aligned} -2l(\theta) &= \sum_{k=1}^{K}\sum_{l=1}^{L} n_{kl}[d_{kl}log(a_{kl}) + \sum_{j=1}^{d_{kl}} \frac{\lambda_{klj}}{a_{kl}} + (SR - d_{kl})log(b_{kl}) + \sum_{j=d_{kl}+1}^{R} \frac{\lambda_{klj}}{b_{kl}} - 2log(\pi_{kl})] \\ &+ \frac{nSR}{2} log(2\pi), \end{aligned}$$

we substitute the equation $\sum_{j=d_{kl}+1}^{R} \lambda_{klj} = tr(W^{1/2}C_{kl}W^{1/2}) - \sum_{j=1}^{d_{kl}} \lambda_{klj}$:

$$\begin{aligned} -2l(\theta) &= \sum_{l=1}^{L}\sum_{k=1}^{K} n_{kl}[-2log(\pi_{kl}) + d_{kl}log(a_{kl}) + \sum_{j=1}^{d_{kl}} \frac{\lambda_{klj}}{a_{kl}} + (SR - d_{kl})log(b_{kl}) \\ &+ \frac{1}{b_{kl}}(tr(W^{1/2}C_{kl}W^{1/2}) - \sum_{j=1}^{d_{kl}} \lambda_{klj})] + \frac{nSR}{2} log(2\pi) \\ &= \sum_{k=1}^{K}\sum_{l=1}^{L} n_{kl}[d_{kl}log(a_{kl}) + \sum_{j=1}^{d_{kl}} \lambda_{klj}(\frac{1}{a_{kl}} - \frac{1}{b_{kl}}) + (SR - d_{kl})log(b_{kl}) \\ &+ -2log(\pi_{kl}) + \frac{tr(W^{1/2}C_{kl}W^{1/2})}{b_{kl}}] + \frac{nSR}{2} log(2\pi). \end{aligned}$$

In order to minimize $-2l(\theta)$ compared to q_{klj}, we minimize the quantity $\sum_{k=1}^{K}\sum_{l=1}^{L} n_{kl} \sum_{j=1}^{d_{kl}} \lambda_{klj}(\frac{1}{a_{kl}} - \frac{1}{b_{kl}})$ compared to λ_{klj}. Knowing that $(\frac{1}{a_{kl}} - \frac{1}{b_{kl}}) \leq 0$, λ_{kl} has to be as high as feasible. So, the j-th column q_{klj} of matrix Q_{kl} is estimated by the eigen function associated to the j-th highest eigenvalue of $W^{1/2}C_{kl}W^{1/2}$.

A4. Parameter a_{kl} update:

Partial derivative of $l(\theta)$ according to a_{kl} correspond to:

$$\begin{aligned}\frac{\partial l(\theta)}{\partial a_{kl}} &= -\frac{1}{2}n_{kl}(\frac{d_{kl}}{a_{kl}} - \sum_{j=1}^{d_{kl}} \frac{q_{kjl}^t W^{1/2} C_{kl} W^{1/2} q_{kjl}}{a_{kl}^2}) \\ &= -\frac{1}{2}n_{kl}(\frac{d_{kl}}{a_{kl}} - \sum_{j=1}^{d_{kl}} \frac{\lambda_{klj}}{a_{kl}^2})\end{aligned}$$

The prerequisite $\frac{\partial l(\theta)}{\partial a_{kl}} = 0$ implies:

$$\begin{aligned}&\frac{\partial l(\theta)}{\partial a_{kl}} = 0 \\ \Leftrightarrow\ & -\frac{1}{2}n_{kl}(\frac{d_{kl}}{a_{kl}} - \sum_{j=1}^{d_{kl}} \frac{\lambda_{klj}}{a_{kl}^2}) = 0 \\ \Leftrightarrow\ & \frac{n_{kl}d_{kl}}{a_{kl}} = \frac{n_{kl}}{a_{kl}^2}\sum_{j=1}^{d_{kl}} \lambda_{klj} \\ \Leftrightarrow\ & a_{kl} = \frac{1}{d_{kl}}\sum_{j=1}^{d_{kl}} \lambda_{klj}\end{aligned}$$

with λ_{kl} the eigen values of block kl.

A5. Parameter b_{kl} update:

Partial derivative of $l(\theta)$ according to b_{kl} correspond to:

$$\frac{\partial l(\theta)}{\partial b_{kl}} = -\frac{1}{2}[\frac{SR - d_{kl}}{b_{kl}} - \sum_{j=d_{kl}+1}^{SR} \frac{q_{klj}^t W^{1/2} C_{kl} W^{1/2} q_{klj}}{b_{kl}^2}]$$

The prerequisite $\frac{\partial l(\theta)}{\partial b_{kl}} = 0$ implies:

$$-\frac{1}{2}[\frac{SR - d_{kl}}{b_{kl}} - \sum_{j=d_{kl}+1}^{SR} \frac{q_{klj}^t W^{1/2} C_{kl} W^{1/2} q_{klj}}{b_{kl}^2}] = 0$$

$$\Leftrightarrow \frac{SR - d_{kl}}{b_{kl}} = \sum_{j=d_{k+1}}^{SR} \frac{q_{klj}^t W^{1/2} C_{kl} W^{1/2} q_{klj}}{b_k^2}$$

$$b_{kl} = \frac{1}{SR - d_{kl}}[tr(W^{1/2} C_{kl} W^{1/2}) - \sum_{j=1}^{d_{kl}} \lambda_{klj}]$$

R-package for co-clustering algorithm: R-package containing code to perform all analyses described in the article. (GNU zipped tar file)

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control 9*, 716–723.

Benitez, I., J.-L. Dez, A. Quijano, and I. Delgado (2016). Dynamic clustering of residential electricity consumption time series data based on hausdorff distance. *Electric Power Systems Research 140*, 517 – 526.

Benitez, I., A. Quijano, J.-L. Dez, and I. Delgado (2014). Dynamic clustering segmentation applied to load profiles of energy consumption from spanish customers. *International Journal of Electrical Power Energy Systems 55*, 437 – 448.

Bhatia, P., S. Iovleff, and G. Govaert (2014, December). blockcluster: An R Package for Model Based Co-Clustering. working paper or preprint.

Biernacki, C., G. Celeux, and G. Govaert (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. PAMI 22*, 719–725.

Bouveyron, C., L. Bozzi, J. Jacques, and F.-X. Jollois (2017, December). The Functional Latent Block Model for the Co-Clustering of Electricity Consumption Curves. *Journal of the Royal Statistical Society: Series C Applied Statistics*.

Bouveyron, C., G. Celeux, T. B. Murphy, and A. E. Raftery (2019). *Model-Based Clustering and Classification for Data Science: With Applications in R.* Cambridge University Press.

Bouveyron, C., E. Côme, and J. Jacques (2015). The discriminative functional mixture model for a comparative analysis of bike sharing systems. *Annals of Applied Statistics*, in press.

Chamroukhi, F. and C. Biernacki (2017, July). Model-Based Co-Clustering of Multivariate Functional Data. In *ISI 2017 - 61st World Statistics Congress*, Marrakech, Morocco.

Corneli, M., C. Bouveyron, and P. Latouche (2019, January). Co-Clustering of ordinal data via latent continuous random variables and a classification EM algorithm. working paper or preprint.

Gouveia, J., J. Seixas, S. Luo, N. Bilo, and A. Valentim (2015, 06). Understanding electricity consumption patterns in households through data fusion of smart meters and door to door surveys. pp. ECEEE SUMMER STUDY PROCEEDINGS.

Govaert, G. and M. Nadif (2013). *Co-Clustering* (1st ed.). Wiley-IEEE Press.

Jacques, J. and C. Biernacki (2018, July). Model-Based Co-clustering for Ordinal Data. *Computational Statistics and Data Analysis 123*, 101–115.

Jacques, J. and C. Preda (2014a). Functional data clustering: a survey. *Advances in Data Analysis and Classification 8*(3), 231–255.

Jacques, J. and C. Preda (2014b). Model based clustering for multivariate functional data. *Computational Statistics and Data Analysis 71*, 92–106.

Keribin, C., G. Govaert, and G. Celeux (2010). Estimation d'un modèle à blocs latents par l'algorithme SEM. In *42èmes Journées de Statistique*, Marseille, France, France.

Laclau, C., I. Redko, B. Matei, Y. Bennani, and V. Brault (2017, August). Co-clustering through Optimal Transport. In *34th International Conference on Machine Learning*, Volume 70 of *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, pp. 1955–1964. Proceedings of Machine Learning Research.

Nadif, M. and G. Govaert (2008). Algorithms for model-based block gaussian clustering. In *Proceedings of The 2008 International Conference on Data Mining, DMIN 2008, July 14-17, 2008, Las Vegas, USA, 2 Volumes*, pp. 536–542.

Petersen, K. B. and M. S. Pedersen (2012, nov). The matrix cookbook. Version 20121115.

Ramsay, J. O. and B. W. Silverman (2005). *Functional data analysis* (Second ed.). Springer Series in Statistics. New York: Springer.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association 66*(336), 846–850.

Schmutz, A., J. Jacques, C. Bouveyron, L. Cheze, and P. Martin (2018, July). Clustering multivariate functional data in group-specific functional subspaces. working paper or preprint.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461–464.

Selosse, M., J. Jacques, and C. Biernacki (2019, October). Model-based co-clustering for mixed type data. *Computational Statistics and data analysis*.

Slimen, Y. B., S. Allio, and J. Jacques (2018). Model-Based Co-clustering for Functional Data. *Neurocomputing 291*, 97–108.

Tureczek, A., P. Nielsen, and H. Madsen (2018, 04). Electricity consumption clustering using smart meter data. *Energies 11*, 859.

Zhou, K., C. Yang, and J. Shen (2017). Discovering residential electricity consumption patterns through smart-meter data mining: A case study from china. *Utilities Policy 44*, 73 – 84.