Dynamic clustering and modeling of temporal data subject to common regressive effects

Louise Bonfils, Allou Same, Latifa Oukhellou

Dynamic clustering and modeling of temporal data subject to common regressive effects

Louise Bonfils*[a], Allou Samé[a], Latifa Oukhellou[a]

[a] *University Gustave Eiffel, COSYS-GRETTIA, 77420, Champs-sur-Marne, France*

Abstract

Clustering is used in many applicative fields to organize data into a few groups. Motivated by behavioral extraction issues from urban data, this paper proposes a new clustering method to model clusters with dynamic profiles while considering common regressive effects. As maximum likelihood estimation is not suitable in this case, the parameters of the proposed model were estimated using variational approximation. The ability of the model to estimate parameters was evaluated using various simulated data and compared with two other models. The article also proposes an application of this model to the extraction of occupant behavior in buildings using a real open source indoor temperature database. The objective is to classify individual houses according to indoor temperature while estimating the effect of meteorological variables and class profiles that can be interpreted as occupancy behaviors.

Keywords:
Clustering, Dynamic latent variable model, Variational inference approximation, Urban data, Occupant behavior in buildings

1. Introduction

In many application domains, clustering data into a small set of clusters is meaningful to highlight common aspects within the clusters. Considering urban data collected in the energy or mobility domains, clustering gives insight into typical user behavior patterns [9, 13]. Customers' habits and preferences can also be classified to build recommendation systems [23].

Usually, the clustering of user behaviors or customer preferences does not consider potential changes or evolutions. Especially in the field of energy, authors have worked on clustering energy consumption patterns [24] by considering time series. However, incentive policies, price changes or innovations can lead

*Corresponding author
Email addresses: louise.bonfils@univ-eiffel.fr (Louise Bonfils*), allou-badara.same@univ-eiffel.fr (Allou Samé), latifa.oukhellou@univ-eiffel.fr (Latifa Oukhellou)

to changes in these behaviors and habits. Thus, it may be interesting to consider the dynamic and evolving aspect of behavior in the classification task. This evolution of behaviors in clustering problems is often taken into account by using segmentation methods to identify periods where behaviors are static and constant, then performing clustering on these specific periods. The segmentation phase can be performed manually based on solid assumptions or using stochastic methods such as Hidden Markov Models [20].

This paper presents a model that attempts to group similar observations into a small set of clusters while estimating class profiles by a dynamic approach using autoregressive processes. We position ourselves in a framework where a set of temporal data from independent entities are determined, partially, by latent processes characterizing clusters. It is assumed that part of these data can be explained by known and common exogenous factors. The proposed model seeks to estimate the effect of these common factors and, at the same time, identify in an unsupervised way the latent clusters and their dynamics from a set of observations.

The proposed model is compared and evaluated with two other models on simulated data sets. The simulation is fully controlled, so the coefficients of the exogenous effects, the class profiles, and the partition are known, making it possible to calculate performance indicators based on estimation errors.

After the evaluation with simulated data, an application to real data from the REFIT data base [16] is proposed in this article. Indoor temperature data from a set of English houses are used. In our case, the objective is to classify these houses and extract class profiles that provide important information about the occupancy patterns of the inhabitants. The dataset includes outdoor weather conditions that will be used as exogenous factors. The idea is to consider that the ambient temperature of a house is influenced by the outdoor weather conditions as well as by the occupancy and the activity of the inhabitants. In fact, in [19], this database and meteorological variables were used to compute and estimate heating behaviors. The classification of behavior and estimation of occupancy is a significant topic in the energy field and the lack of knowledge and the complexity of these behaviors make it difficult to predict the energy consumption of a building [10]. In the literature, occupancy and activity in houses are frequently estimated using models based on Markov Chains [1, 11]. Occupancy is generally a component of Bottom-up models that are an essential family in estimating energy consumption [28]. Lastly, according to [2], another interest in estimating occupancy in dwellings is that it could allow for better adaptation of the amount of hot water stored in hot water tanks for dwellings that have this equipment. These studies illustrate the importance of occupancy in predicting energy consumption. The model proposed here could, in the future, be used to infer occupancy.

In [19], the authors used outside temperature and ambient temperature of the REFIT dataset in order to compute and estimate the heating behavior in a typical English household. In our case, the proposed model classifies these houses and extracts class profiles that provide important information about the

occupancy patterns of the inhabitants.

Finally, this paper's contributions can be summarized in three main points:

- First, this article presents a latent process model to classify temporal data, estimate the effect of common exogenous factors, and model cluster profiles as stochastic processes.
- Then, the proposed model is compared to a two-step regression model based on the k-means algorithm and to a constant class centers model estimated with the EM algorithm. These models are evaluated according to three criteria based on the estimation error of the different components of the model. The parameter estimation on simulated and controlled data allows us to compute the estimation error made on class centers, on the effect of exogenous factors, and on the classification of observations.
- Finally, the REFIT open-access database [16] is used to evaluate and compare the models on a real dataset. This comparison is based on approximate log-likelihoods. By classifying the temperature data, classes were identified whose profiles can be interpreted in terms of occupancy within dwellings.

The rest of the paper is organized as follows. The second section of this paper presents a brief review of the literature to explain the methodological positioning of the model. The third section presents the construction of the model as well as the inference method and the algorithm used for parameter estimation. Then, the fourth is devoted to the performance results obtained using simulated datasets and the comparison with two other methods. Finally, the last section presents the application of the model on a real ambient temperature dataset.

2. Literature Review

Unsupervised classification is a major topic in statistical modeling and data analysis. It is used to summarize the information of a set of observations, by a small number of groups. The most widely used methods are static methods such as K-means or hierarchical methods (CAH). Mixture models, reviewed in [22], are also widespread because they are powerful and offer greater flexibility than k-means models since they contain fewer constraints. The Gaussian mixture model, for example, estimates a probabilistic classification [7] instead of allocating data strictly to identified clusters. Several variants of mixture models exist, among which regression mixture models, which classify data but also estimate cluster-specific regression coefficients relative to known factors [12, 26].

To deal with the clustering of multivariate temporal data, [3] and [27] proposed regression mixture models applied to three-dimensional data, with a set of n observations, of d variables, over p time instants or locations. These models are interesting because they estimate the effect of group-specific factors with time-dependent variables. However, the observations at each time step are considered as independent of each other, so the dynamics or links that may exist

between observations of the same entity are not taken into account. In [3], the authors proposed an application in the field of genotype classification, where observations are grouped into similar clusters for each time step or location. In our case, the temporal data are clustered based on the entire temporal sequence observed. However, clustering of multidimensional temporal data is also an important topic in other domains, such as the energy domain, where clustering of consumption behaviors allows for better understanding and thus better prediction of certain behaviors [24]. In the field of temporal data classification, some authors use time series decomposition methods to classify the extracted elements [13, 8]. Along the same lines, it is also possible to use segmentation methods to classify temporal data according to the hidden states identified [20].

The above mentioned methods attempt to classify temporal data without necessarily modeling the evolution of estimated class profiles over time. However, as previously mentioned, it could be interesting to estimate clusters from temporal data while modeling the class centers and their evolution over time. The Gaussian mixture model [15] or Kalman filter mixture models [6] classify, at each time step, the temporal data and estimate the centers of the classes by fixing an a priori on the evolution of the latter. The modeling of class profiles from autoregressive processes enables the evolution and dynamics of clusters in time to be taken into account.

The two methods mentioned above classify observations for each time sequence because the objective is to model the evolution of clusters over time. In our case, clusters are constructed in an unsupervised way and each observation belongs to a cluster only during the observed period. Moreover, the regressive part corresponding to the observed exogenous effect is considered as common to all the observations, contrary to regression mixture models which identify different effects for the different components of the mixture. The approach adopted here is that of latent variable models using mixture models. These models allow some flexibility to build classification models with dynamic class profiles or time regressive components ([15], [6], [26]).

In order to estimate the parameters of mixture models, the Expectation-Maximization algorithm (EM) is usually used because the maximum of the likelihood is intractable [21]. In [18], however, the authors point out that the EM algorithm is not suitable for some complex generative models involving multiple and temporal latent variables. In [5] the authors estimated time dependent effects via *Variable Neighborhood Search* algorithms. However, this algorithm does not deal with the autoregressive latent variables. This topic was explored in [15] using variational inference methods. Indeed, when exact inference is impossible, as is the case with the model proposed here, the variational inference method is able to approach the optimum and to estimate the parameters.

These elements and this context led us to build a mixture model with dynamic latent variables, which is presented in the next section.

3. Model definition and parameter estimation

This section first presents the model definitions and notations, then, the variational inference method used to address the optimization problem is introduced. Finally, a subsection is dedicated to the iterative algorithm built for parameter estimation.

3.1. Model definition

To formalize the model, we consider the following notations:

- $(\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_n)$ a set of n observations, where $\mathbf{x}_i = (x_{it})_t$ is a sequence of T observed data, with $\forall t$, $x_{it} \in \mathbb{R}$,
- $\mathbf{u}_t$ $(t \in [\![1, T]\!])$ a $(p+1)$-dimensional vector representing p exogenous and observable factors. We include the constant value 1 in the vector to take into account a level parameter (bias).

The model proposed in this article assumes that the series $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ can be grouped into K clusters. It is characterized by a regressive common component reflecting the effect of known and observed factors, and by cluster-specific profiles reflecting the effect of latent dynamic factors. According to this assumption, we consider that x_{it} can be explained by the following model:

$$\forall i \in [\![1, n]\!], \quad \forall t \in [\![1, T]\!]; \qquad x_{it} = \mathbf{u}_t' \mathbf{a} + \sum_{k=1}^{K} z_{ik} b_{kt} + e_{it}, \tag{1}$$

where z_{ik} is a binary variable equal to 1 if the observation i belongs to the class k and 0 otherwise. We assume that z_i, satisfying $z_i = k$ if $z_{ik} = 1$, follows a Multinomial distribution with parameters $\boldsymbol{\pi} = (\pi_k)_{k=1,\ldots,K}$. Also, the profile $(b_{kt})_{t=1,\ldots,T}$ corresponds to the unobservable group-specific profiles, e_{it} is a centered and normally distributed noise with variance v_k and $\mathbf{a} = (a_0, \ldots, a_p) \in \mathbb{R}^{(p+1)}$ refers to the regression coefficients associated to exogenous factors and a_0 denotes the level coefficient.

The latent profiles $(b_{kt})_{t=1,..,T}$ are modeled as first-order autoregressive processes as follows:

$$\forall t \in [\![1, T]\!], \forall k \in [\![1, K]\!], \quad b_{kt} = \Phi_k b_{kt-1} + \nu_{kt}, \tag{2}$$

where, ν_{kt} is a centered Gaussian noise with variance w_k, and b_{k0} is normally distributed with μ_{k0} and σ_{k0} as mean and variance parameters. The coefficient Φ_k satisfies the stationarity constraint $|\Phi_k| < 1$. Then, using the previous elements, the vector of parameters of the model is as follows: $\boldsymbol{\Theta} = \{(v_k, w_k, \pi_k, \Phi_k, \mu_{k0}, \sigma_{k0})_{k=1,\ldots K}, \mathbf{a}\}$.

The model defined by Equation (1) is not identifiable. In fact, the coefficient a_0 can be confused with class profiles $(b_{kt})_{(k,t)}$. In this case, it is necessary to

add a constraint to the model. In the present case, by setting $\tilde{\mathbf{a}} = (a_1, \ldots, a_p)$, and noting $\tilde{\mathbf{u}}_t$ the corresponding p-dimensional exogenous variables, we have:

$$\mathbf{u}_t'\mathbf{a} + \sum_k z_{ik} b_{kt} = a_0 + \tilde{\mathbf{u}}_t'\tilde{\mathbf{a}} + \sum_k z_{ik} b_{kt} = (a_0 - \alpha) + \tilde{\mathbf{u}}_t'\tilde{\mathbf{a}} + \sum_k z_{ik}(b_{kt} + \alpha).$$

Thus, depending on the value of α, there is an infinite number of choices for a_0 and b_{kt}. To ensure the identifiability, we add the following constraint to the model: $\sum_{k=1}^{K} \sum_{t=1}^{T} b_{kt} = 0$.

The modeling of class profiles characterizes this model as a first-order autoregressive process. This choice of modeling latent processes to characterize classes was already proposed in [6] and [15]. They also chose this a priori on the class centers in the framework of dynamic classification models. They show that this modeling has better estimation performances on simulated datasets than other models such as regression mixtures or simpler Gaussian mixture models. Moreover, it should be noted that modeling based on autoregressive processes of order one is parsimonious, limiting the number of parameters to be estimated. Other latent process models can however be considered. For example, in [25], the authors characterize the classes from time-dependent polynomials in a context of non-linear regression with other regimes.

After presenting the model parameters and assumptions, the next subsection is dedicated to the theory related to the variational inference method and algorithm used for parameter estimation.

3.2. Variational Inference Methodology

After the model construction presented in Subsection 3.1, an estimation method needs to be found. In the case of a log-likelihood maximization problem, this function can be written as follows:

$$\begin{aligned}\mathcal{L}(\mathbf{x}; \boldsymbol{\Theta}) &= \log(P(\mathbf{x}; \boldsymbol{\Theta})) \\ &= \log\left(\sum_z p(\mathbf{z}; \boldsymbol{\Theta}) \int_{\mathbf{b}} p(\mathbf{b} \mid \mathbf{z}; \boldsymbol{\Theta}) p(\mathbf{x} \mid \mathbf{b}, \mathbf{z}; \boldsymbol{\Theta}) d\mathbf{b}\right). \end{aligned} \tag{3}$$

But in our case, the complex structure of the model and the presence of latent variables make the log-likelihood intractable and the parameter estimation via a direct maximization method intractable. It is therefore necessary to get around this problem by using variational inference methods.

To do so, a function, called "Evidence Lower Bound", $F(q(\mathbf{z}, \mathbf{b}), \boldsymbol{\Theta})$ is introduced and defined such that:

$$F(q, \boldsymbol{\Theta}) = \mathbf{E}_q(\mathcal{L}_c(\boldsymbol{\Theta})) + H(q), \tag{4}$$

where $H(q)$ is the entropy of the variational distribution $q(.)$, and $\mathcal{L}_c$ refers to the complete log-likelihood of the model. The previous function $F(.)$ is called "The Evidence Lower Bound" because it respects the following equation:

$$\mathcal{L}(\mathbf{x}; \boldsymbol{\Theta}) \geq F(q(\mathbf{z}, \mathbf{b}), \boldsymbol{\Theta}).$$

The main goal is to estimate the variational distribution and to estimate the model parameters by maximizing the Evidence Lower Bound.

In a general case, the distribution q can take any form. There is therefore an infinite number of possibilities, and the maximization problem is difficult to solve. To simplify the problem and ensure a solution, it is possible to reduce the possible form of the variational density to a restricted function family. In [4], the authors justify the choice of the *mean-field family* because it greatly simplifies the optimization problem while offering good performances. The mean-field family corresponds to a factorization assumption such that the function $q(.)$ has the following form:

$$q(\mathbf{z}, \mathbf{b}) = \prod_{i=1}^{n} q_z(z_i) \prod_{t=0}^{T} \prod_{k=1}^{K} q_b(\mathbf{b}_{kt}), \tag{5}$$

where q_z is the distribution of the latent variable z_i and q_b, the distribution of the processes (b_{kt}). In this model, variables b_{kt} are Gaussian with mean parameters m_{kt} and standard error λ_k. The variables z_i are distributed according to a Multinomial distribution with parameters $(\tau_{ik})_{i=1,...n;k=1,...,K}$. With the previous element, the function $q(.)$ can be rewritten as follows:

$$q(\mathbf{z}, \mathbf{b}) = q(\mathbf{m}, \boldsymbol{\tau}, \boldsymbol{\lambda}) = \prod_{i=1}^{n} \prod_{k=1}^{K} \tau_{ik}^{z_{ik}} \prod_{t=0}^{T} \prod_{k=1}^{K} \mathcal{N}(b_{kt}, m_{kt}, \lambda_k). \tag{6}$$

This variational density function leads us to introduce Variational parameters that will be estimated by maximizing the Evidence Lower Bound. The variational parameters of the models are as follows:

- $\boldsymbol{\tau} = \{(\tau_{ik})_{k=1,..,K;i=1,..,n}\}$,
- $\mathbf{m} = \{(m_{kt})_{k=1,..,K;t=0,..,T}\}$,
- $\boldsymbol{\lambda} = \{(\lambda_k)_{k=1,..,K}\}$.

Using previous elements presented in Section 3.1, the Evidence Lower Bound can be explicitly written.

$$\begin{aligned} F(\mathbf{m}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \boldsymbol{\Theta}) = & \sum_{i,t,k} \tau_{ik} \left(\log(\pi_k \varphi(x_{it}; m_{kt} + \mathbf{u}_t' \mathbf{a}, v_k)) - \frac{1}{2} \lambda_k (v_k^{-1}) \right) \\ & + \sum_{k,t} \log\left(\varphi(m_{kt}; \Phi_k m_{kt-1}, w_k)\right) - \frac{1}{2} \lambda_k ((w_k^{-1}) + (w_k^{-1} \Phi_k^2)) \\ & + \sum_{k} \log\left(\varphi(m_{k0}, \mu_{k0}, \sigma_{k0})\right) - \frac{1}{2} \lambda_k (\sigma_{k0}^{-1}) \\ & - \sum_{i,k} \tau_{ik} \log(\tau_{ik}) + \frac{d(T+1)}{2} \sum_{k} \log(2\pi e) + \log(\lambda_k). \end{aligned} \tag{7}$$

The Evidence lower bound is used to construct the iterative algorithm for parameter estimation that is presented in the next subsection.

3.3. Iterative Algorithm for parameter estimation

Algorithm 1 is an algorithm with an initialization step and an iterative part consisting in updating the variational parameters and the parameters of the model, which ends when a stopping criterion is reached.

input : Observed data $(\mathbf{x}_1, ..., \mathbf{x}_n)$, with $\mathbf{x}_i \in \mathbb{R}^T$, the number of clusters K and exogenous factor vectors $(\mathbf{u}'_t)_{t=1,...T}$
output: $(\mathbf{m}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \boldsymbol{\Theta})$

initialization:
$\boldsymbol{\Theta}^{(0)}$, $\boldsymbol{\lambda}^{(\mathbf{0})}$, $\mathbf{m}^{(0)}$ using K-means algorithm and initial setting values;
repeat
  for $k = 1$ **to** K **do**
    for $t = 1$ **to** T **do**
      Compute class profiles $\left(m_{kt}^{(q)}\right)$;
      for $i = 1$ **to** n **do**
        Compute the probabilities of membership $\left(\tau_{ik}^{(q)}\right)$
      end
    end
    Compute variational variances (λ_k);
    Compute parameters $\left(\mu_{k0}^{(q)}, \sigma_{k0}^{(q)}\right)$;
    Compute Class proportions $\left(\pi_k^{(q)}\right)$;
    Compute Variances $\left(v_k^{(q)}\right)$ and $\left(w_k^{(q)}\right)$;
  end
  Compute regression coefficients $\mathbf{a}$;
until *The stop criterion is reached*;

Algorithm 1: Variational inference algorithm for parameter estimation

The initialization consists in setting a starting point for the parameters. Initial values are chosen for variance parameters $(v_k^{(0)}, w_k^{(0)}, \sigma_0^{(0)})$, proportion parameters $(\pi_k^{(0)})$, and variational variances $(\lambda_k^{(0)})$. Then, initial values are computed for $(m_{kt}^{(0)})$, $(\tau_{ik}^{(0)})$ and coefficients $\mathbf{a}$ using the K-means algorithm.

The iterative algorithm consists in updating each variational parameter and model parameter, one by one, by considering the others as fixed to the previous updated value. The formulas used for the updating are obtained by the maximization of the Evidence Lower bound 4 according to each parameter while considering the others as fixed. Updating the variational parameters of class centers $(m_{kt}^{(q+1)})_{(k,t)}$ requires an adapted version of the Kalman filter [15].

It is assumed that the algorithm has converged to a solution when the updated class centers are sufficiently close to those obtained in the previous iteration. In other words, the stopping criterion for this algorithm is, with $\varepsilon \to 0$, $\frac{1}{KT}\sum_{t,k}(m_{kt}^{(q+1)} - m_{kt}^{(q)})^2 < \varepsilon$. Once this condition is reached, the algorithm stops.

The algorithm was implemented and tested using simulated data. The following section is devoted to the presentation of the simulated data and the comparison of the performance of the proposed model with two other reference models.

4. Evaluation of the model on simulated datasets

In order to evaluate the model performances, it is important to simulate various datasets. We define three criteria to evaluate the accuracy of the proposed model. These results were then compared with the performances of two other models used as references.

4.1. Simulation of various datasets

The simulation of a dataset can be decomposed into four steps. An example of generated data is presented in figure 4.1.

1. For a given number of clusters K and a length of sequence T, class profiles are generated as first-order autoregressive processes: these profiles are drawn more or less distinct depending on the chosen level of difficulty, and centered (see Figure 4.1(B)).
2. For a given number of observations n, generate cluster labels using the mixture proportions. Depending on the level of difficulty, the mixture can be more or less heterogeneous.
3. Define coefficients associated to exogenous factors which can be real or simulated data (see Figure 4.1(A)).
4. Using formula 1 and the previous simulated elements, each observation is generated (see Figure 4.1(C).

The model is evaluated by generating various data sets with two and four classes, and different numbers of observations. For each configuration, the models were tested on two hundred different datasets. First, we consider the fixed time window $T = 100$ and vary the number of observations ($n = 20$ and $n = 150$). Then, we fix the number of observations to $n = 100$ and set the time window to $T = 80$ and $T = 300$. The following results, presented in Figure 2 and 3 and Table 1, were obtained with four clusters. In order to explore as many cases as possible, the datasets were generated considering different levels of difficulty using different mixture proportions and distances between simulated class profiles.

Using the simulated datasets, the performance evaluation is based on three criteria presented in the next subsection.

4.2. Criteria for model performance evaluation

As a reminder, the model is assumed to be able to identify the global exogenous effect, classify observations, and estimate class centers as dynamic processes. The objective is to evaluate the model on these three aspects using three

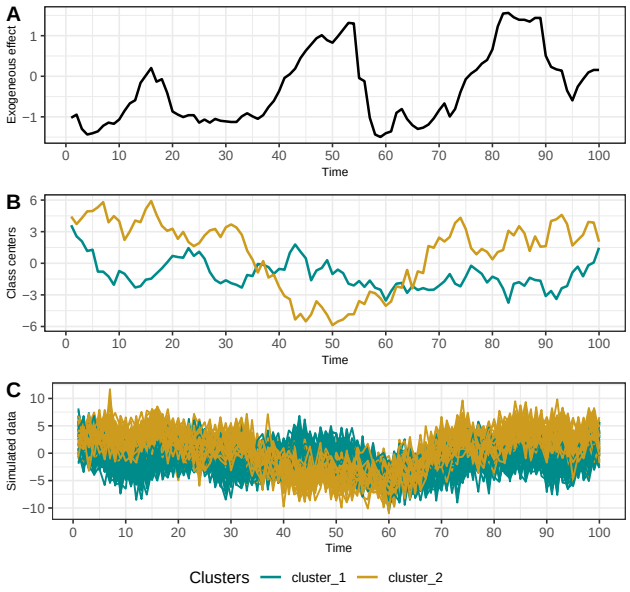

Figure 1: Simulation of one dataset of 150 observations and 100 time sequences. (A) corresponds to the exogenous factors effect. (B) represents the class profiles simulated as autoregressive processes. (C) represents the set of n observations simulated using the previous elements.

criteria. Note that cluster labels have been reorganized to maximize the classification rate. First, the mean square error is used to evaluate the ability of the model to estimate class profiles:

$$\text{CRIT}_1 = \frac{1}{KT}\sum_{t=1}^{T}\sum_{k=1}^{K}(\hat{m}_{kt} - b_{kt})^2. \tag{8}$$

Then, the ability of the model to identify and estimate the exogenous effect is evaluated using the mean square error computed on exogenous factors such that:

$$\text{CRIT}_2 = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{u}_t'\hat{\mathbf{a}} - \mathbf{u}_t'\mathbf{a})^2. \tag{9}$$

Finally, the correct classification rate is used to evaluate the model:

$$\text{CRIT}_3 = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{z_i=\hat{z}_i\}}. \tag{10}$$

4.3. Two reference models for performance comparison

The three criteria for performance evaluation were computed for the proposed model and compared with two other models presented in this subsection.

First, let's consider the same notations as those presented in Section (3.1). The first model, used as a reference and called *"Constant class centers model"* models observations as the results of a regression part and a mixture of Gaussian densities such that:

$$\forall i \in [\![1,n]\!], \quad \forall t \in [\![1,T]\!]; \qquad p(x_{it};\boldsymbol{\theta}_1) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(x_{it};\mu_k + \mathbf{u}_t'\mathbf{a},\sigma_k\right), \tag{11}$$

with $\boldsymbol{\theta}_1 = ((\pi_k,\mu_k,\sigma_k)_{k=1,..K},\mathbf{a})$ the model parameters vector. First, the coefficient vector $\mathbf{a}$ is estimated using the following formula:

$$\hat{\mathbf{a}} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{u}_t\mathbf{u}_t'\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{u}_t x_{it}\right). \tag{12}$$

Then, the EM algorithm is used on $\widetilde{x_{it}} = x_{it} - \mathbf{u}_t'\hat{\mathbf{a}}$ in order to estimate mixture parameters $(\pi_k)_{k=1,..,K}$ and Gaussian parameters $(\mu_k,\sigma_k)_{k=1,...K}$ to characterize the clusters. In this model, the observations $(x_{it})_{i,t}$ are considered as independent observations and the algorithm classifies observations for each time sequence t. The final classification is obtained by taking the most frequent clusters for each observation $\mathbf{x}_i$.

The second model, called *"Two-step regression model"*, is based on a first part of the regression using formula 12. Then, the residuals, denoted by $\widetilde{\mathbf{x}}_i = (\widetilde{x}_{i1},..,\widetilde{x}_{iT}), \forall i = 1,..,n$ such that $\widetilde{x}_{it} = x_{it} - \mathbf{u}_t'\hat{\mathbf{a}}$, are classified using a K-means algorithm. The observations are grouped into clusters and the centers of these clusters are T dimensional vectors. The K-means algorithm is chosen because this method is widely used for clustering time series [24].

The next section presents the results of the performance criteria computed on various simulated datasets for the proposed model and the two reference models.

4.4. Results of criteria computed on various simulated datasets

Table 1 presents the mean value of the criteria for the three models in different cases. The "Constant class center model" has higher values for the three criteria in each of the presented cases.

Figure 2 shows box-plots of the three criteria computed for the proposed model and the two-step regression model, on one hundred and fifty datasets with different levels of difficulty with four clusters when the number of observations is equal to $(n = 20, T = 100)$ and $(n = 150, T = 100)$. The Constant class center model, which does not provide good performances, is not presented for readability reasons. Levels of difficulty depend on the similarity between class profiles and the degree of cluster mixing. For example, similar class profiles or highly mixed clusters correspond to the highest levels of difficulty.

	CRIT_1		CRIT_2		CRIT_3	
T=100	n=20	n=150	n=20	n=150	n=20	n=150
Proposed Model	**1.6**	**0.97**	**0.078**	**0.054**	**0.92**	**0.99**
Constant Model	9.05	8.53	0.17	0.1631	0.42	0.46
Two-step Regression	1.88	1.04	0.17	0.1631	0.91	0.99
n=100	T=80	T=300	T=80	T=300	T=80	T=300
Proposed Model	**1.74**	**1.58**	**0.062**	**0.051**	**0.99**	**1**
Constant Model	5.9	5.68	0.078	0.067	0.62	0.81
Two-step Regression	1.77	1.65	0.078	0.067	0.99	0.98

Table 1: Performance results obtained for the three models. CRIT_1 corresponds to the mean square error computed on the class centers, CRIT_2 corresponds to the mean square error computed on the exogenous effects and CRIT_3 corresponds to the classification rate. For the three criteria, the proposed model provides the best performances. Also, the more observations there are, the more accurate the model is.

These results show different performances for the two models. In fact, the proposed model, for the three criteria, seems to be more accurate on the simulated datasets. It can be seen that for the classification rate (Figure 2 (C)), the two-step regression model has more extreme values than the proposed model for n=150 as the latter estimates probabilities of membership and a mixture of clusters whereas the former estimates a strict classification.

In addition, we can note that the more observations the dataset contains, the more accurate the model is on the basis of the three criteria.

Figure 3 shows box plots of the three criteria computed for two of the models, on two hundred datasets with different levels of difficulty with four clusters when the number of observations is equal to $(n = 100, T = 80)$ and $(n = 100, T = 300)$. As before, the third model is not represented, but the results are displayed in table 1.

First, according to the three criteria, the complete model performs better than the two-step regression model. The means of the first criterion (Figure 3(A)) are close, as well as the means of the second criterion (Figure 3(B)) but the proposed model performs better and shows less extreme values. This may be due to the difficult cases of highly mixed clusters or similar class centers, since the proposed model estimates the class profiles and the effect of exogenous factors better. In this case, the longer the sequence T is, the more accurate the model is.

The results displayed in Figure 2 and Figure 3, show that, as expected, the values of the first two criteria decrease with the size of the temporal window (T) and the number of observations (n). This means that the more data there are, the more accurate the model is. The previous figures and tables also show that,

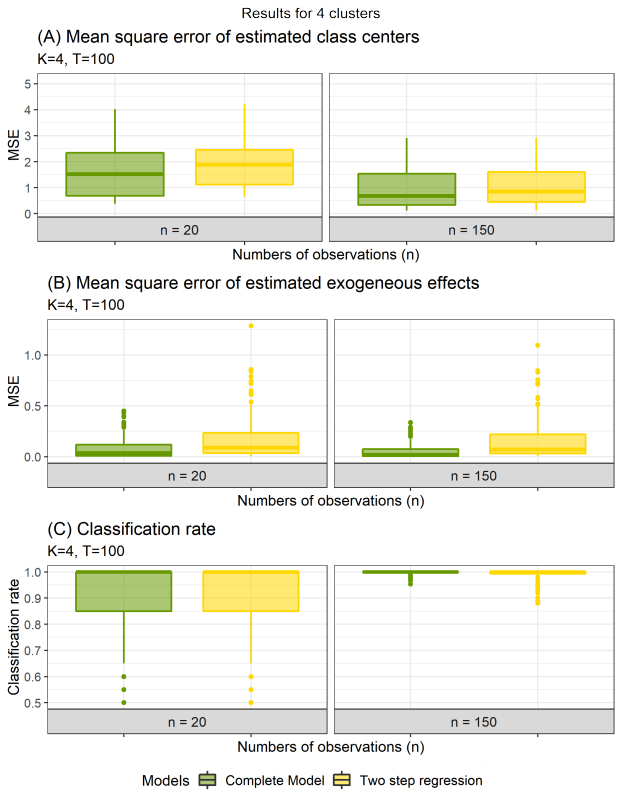

Figure 2: Box-plots of three criteria obtained for the proposed complete model (green) and the Two-step regression model (yellow) with dataset of 100 time sequences and different numbers of observations (n=20 and n=150) and two clusters. The box plots were obtained using 200 datasets for each sample size considered with different levels of difficulty. These difficulty levels are managed using the distance between the simulated class profiles and the degree of cluster mixing.

on all three criteria, the proposed model outperforms the other approaches. The performances of the proposed model compared to the model with constant class centers highlight the interest of estimating the class profiles dynamically.

The third model is based on the assumption of constant class centers through time and classifies the time series using the clustering made for each time stamp.

These results show that this model is not well-adapted to the simulated data. It is important to note however that the regression step, used for the two reference models, provides a good estimate of the exogenous effect.

After showing that the proposed model performs well on simulated data, the next section presents an application on real indoor temperature data of twenty houses in the UK.

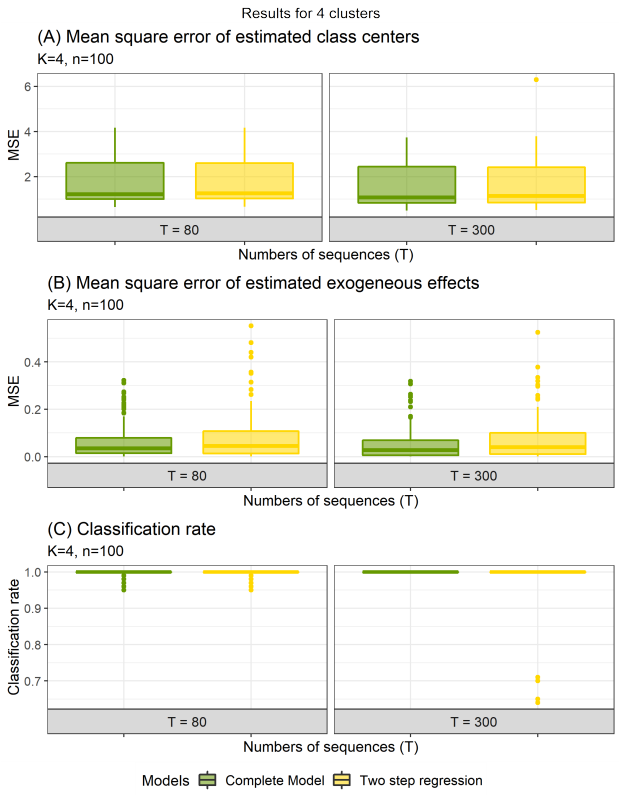

Figure 3: Three-criteria box plots obtained for the proposed full model (green) and the two-step regression model (yellow) with a dataset of 100 observations and different numbers of sequences (T=80 and T=300) and four clusters. The box plots were obtained using 200 datasets for each sample size considered with different levels of difficulty. These difficulty levels are managed using the distance between the simulated class profiles and the degree of cluster mixing.

5. Application of the model to real ambient temperature data from a set of individual houses

The model was then applied on a real dataset in order to classify apartments according to the occupants' behavior, based on their energy consumption. We consider that variables such as indoor ambient temperature could be influenced by a set of exogenous and known factors (outside temperature for example) and a non-observable part relating to the occupation and activity of the inhabitants.

The dataset [16] used in this application consists of 20 individual houses in Loughborough, UK. This dataset contains, among other things, the indoor temperature of the houses' living rooms, as well as a set of characteristics concerning the inhabitants of the dwellings, and finally meteorological data such as the outside temperature or the solar irradiance. Motion detection data are also available for nineteen of the houses during relatively short periods.

5.1. Indoor temperature data from 20 homes

Indoor temperature data are available for 20 homes every 30 minutes over a period from September 2013 to August 2015. However, during this long period, some data are missing for unknown reasons. To avoid data interpolation and loss of information, we focus on a smaller period from the 10^{th} of February to the 18^{th} of June, 2014. No data are missing during this period for the 20 houses. This period was used to compare the models on real datasets in order to use as much information as possible. For the second part, however, the application was made on a dataset between the 24^{th} and the 30^{th} of November, 2014 because motion detection data are also available for this specific period, enabling a class profile interpretation.

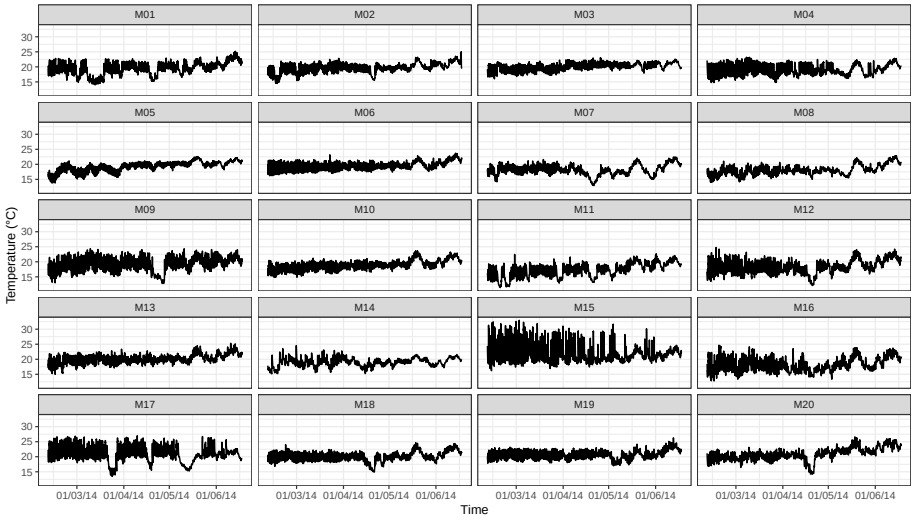

Figure 4: Temperature inside living rooms for 20 houses in the UK in (°C) from February 10 to June 18, 2014. Indoor temperature was recorded in the living-room of each house at 15-minute intervals. The Figure shows aggregated data with 30-minute intervals for each house.

Figure 4 gives the ambient temperature for the houses during almost 5 months. It shows that houses do not have the same heating behavior and show different temperature patterns.

5.2. The choice of exogenous factors

The proposed model has a regressive component common to all the observations, composed of common and known exogenous factors. We consider that the indoor temperature data are impacted by meteorological factors, especially outside temperature and solar irradiance, as these variables are used in the case of thermal simulation using physical modeling [14] for example and also in [17] for modeling and predicting the indoor temperature in a building. Figure 5 shows the two meteorological variables considered in this application.

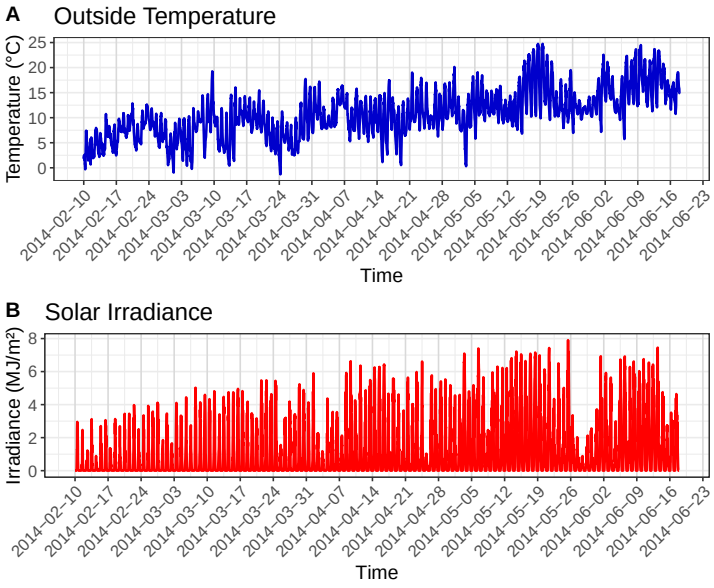

Figure 5: Normalized meteorological variables considered as exogenous factors for the models comparison,from February 10 to June 18, 2014. (A) is the outside Temperature (°C) and (B) is the total solar irradiance (MJ/m^2). Meteorological data recorded by Loughborough University.

In addition, the hour of the day is taken into account as an exogenous factor and introduced as a periodic variable. Because data are available with a 30-minute interval, the two hour variables, called h_1 and h_2 were built such that:

$$\forall t \in \{1, 3, 5, ..T-1\} \begin{cases} h_{1_t} = \sin(\frac{2\pi t}{24}) \\ h_{2_t} = \cos(\frac{2\pi t}{24}) \\ h_{1_{t+1}} = \sin(\frac{2\pi t}{24}) \\ h_{2_{t+1}} = \cos(\frac{2\pi t}{24}). \end{cases} \tag{13}$$

5.3. Criterion used for model comparison

In the case of simulated data, both cluster membership and parameters are known and the comparison of models is possible using criteria based on the estimation error of these models. However, the three criteria defined in section 4 for model performance evaluation are not computable for real data since the true classification, the class profiles and the coefficients of the exogenous factors are unknown. For this reason, a new criterion is defined in this section to compare the models.

The variational method is used in the case where the likelihood cannot be directly maximized and provides an approximation of this quantity. For this reason, the comparison of criteria based on the likelihood can be interesting. If, for a fixed number of parameters, a higher likelihood is obtained with the

proposed model than with the other models, this means that the data set is better explained by the proposed model.

The proposed evaluation criterion, which can be seen as an approximation of the log-likelihood, whatever model is adopted, is defined as:

$$\text{CRIT} = \frac{1}{nT} \sum_{k} \sum_{\hat{z}_i=k} \sum_{t} -\log(\hat{\sigma}_k^2) - \frac{1}{\hat{\sigma}_k^2}(x_{it} - \hat{b}_{kt} - \mathbf{u}_t \hat{\mathbf{a}}_m)^2, \tag{14}$$

where $\hat{\mathbf{b}}_k$ and $\hat{\mathbf{a}}$ are respectively the estimated class centers and regression coefficients and $\hat{\sigma}_k$ the estimated variance parameters obtained by each model. The variable $\hat{z}_i$ is equal to k if the observation $\mathbf{x}_i$ belongs to the cluster k. For the two-step regression model, $\forall k = 1, ...K,\ \hat{\sigma}_k = \hat{\sigma}$. In addition, for the Constant Class center model, $\forall t = 1, ...T,\ \hat{b}_{kt} = \hat{b}_k$.

This criterion was computed for the three models and for different real data sets. The results obtained are presented and compared in the next subsection.

5.4. Results

In order to fully compare the models, one-, two-, three-, and four-week rolling one-week window data sets were extracted and used. These periods correspond to data sets with, respectively, $T = 336$, $T = 672$, $T = 1008$ and $T = 1344$ observations. On these different datasets, the three models were estimated with different numbers of clusters. Table 2 displays the criteria obtained for the three models on the collection of data sets. The proposed model gives a higher approximate log-likelihood for all the considered periods and hyper-parameters K. This means that for real datasets, with various dimensions, the proposed model gives a better likelihood than the two baselines.

The objective of this model applied to indoor temperature is to cluster houses and estimate profiles corresponding to latent variables related to endogenous factors such as activity or occupancy. The next subsection presents the results obtained with the complete model on a one-week dataset.

5.5. Estimated class profiles and classification on one-week of real data

The following subsection is dedicated to the model application on a one-week dataset of ambient temperature within 18 houses. The period chosen was between November 24^{th} and November 30^{th} 2014 because motion detection data are also available for this period, but, in return, data for House 2 and House 17 are not available. These motion detection variables provide additional information that is useful to confirm the class profile interpretation. Figure 6 represents the model input. Indoor temperature data are normalized and used as the set of 18 observations with T=336, and the outside temperature, solar irradiance, and two periodic hour variables, as the exogenous factors for the model estimation.

		CRIT (approximate log-likelihood)		
Sequence T	Number of clusters	Proposed Model	Two-step regression model	Constant class centers model
T = 336	*K=2*	**0.222**	0.191	-0.417
	K=3	**0.541**	0.456	-0.590
	K=4	**0.807**	0.674	-0.776
	K=6	**1.329**	1.009	-2.481
T = 672	*K=2*	**0.126**	0.103	-0.615
	K=3	**0.417**	0.328	-0.950
	K=4	**0.690**	0.507	-1.248
	K=6	**1.189**	0.830	-3.270
T = 1008	*K=2*	**0.048**	0.046	-0.790
	K=3	**0.368**	0.247	-1.015
	K=4	**0.657**	0.416	-1.430
	K=6	**1.102**	0.719	-1.891
T = 1344	*K=2*	**0.022**	0.024	-0.809
	K=3	**0.310**	0.210	-1.226
	K=4	**0.597**	0.369	-1.282
	K=6	**1.142**	0.665	-1,444

Table 2: Approximate Log-likelihoods computed for various real datasets using the proposed model and the two reference models. Criteria are computed for datasets of different sizes. The table shows that the proposed model allows higher log-likelihood approximations than the other two on real datasets.

The objective is to estimate the effect of the four exogenous factors on the indoor temperature, identify clusters of similar houses and estimate class profiles using the model presented in section 3.

The number of clusters is a hyper-parameter of the model. The BIC criterion is widely used for model selection. Figure 7 represents the BIC criterion computed for $K = 2, ..., 17$. The BIC criterion is defined for this model using the Evidence Lower bound 7 such that:

$$BIC(k) = -2F(\hat{\mathbf{m}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\tau}}, \boldsymbol{\Theta}) + N_k \log(nT), \tag{15}$$

where n is the number of observations, T the length of the observed sequence, and N_k the number of free parameters of the model with k clusters.

The BIC criterion decreases until $K = 16$. However, the goal is to build a small number of clusters to summarize the behaviors and interpret the clusters. As, for K ¿ 5, the additional clusters contain only one house, K = 5 was therefore chosen in order to estimate behaviors for a small number of interpretable clusters with more than one observation in each cluster.

The following results were obtained from the proposed model estimation on the data (see Figure 6) with $K = 5$. The model estimates class centers ($\hat{\mathbf{b}}$), exogenous effect, ($\hat{\mathbf{a}}$) and also, for each observation and cluster, the probability of membership $(\hat{\tau}_{ik})_{i,k}$). Then, using these elements,the estimated data can be

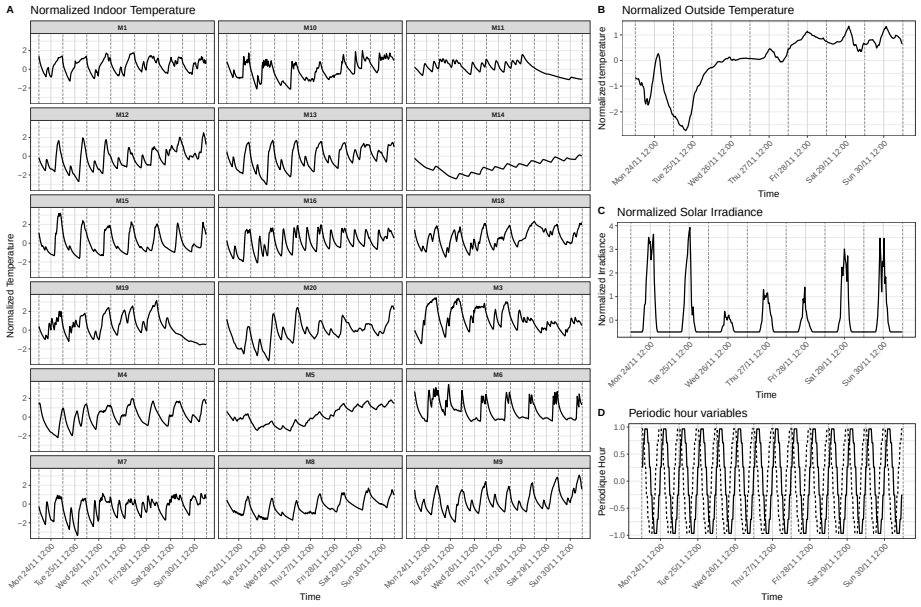

Figure 6: Normalized Indoor Temperature (A) used as inputs, Outside Temperature (B) and Solar Irradiance (C) and periodic hour variables (D) used as exogenous factors for the estimation.

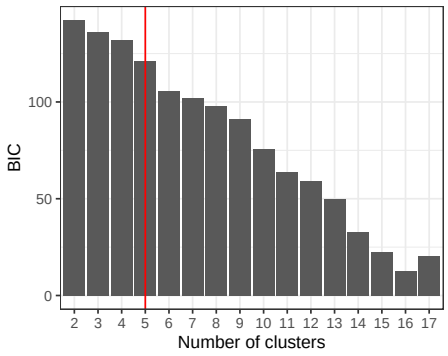

Figure 7: BIC criterion computed when the number of clusters K varies between $K = 2, ..., 17$.

computed as follows:

$$\forall i = 1, .., n \text{ and } t = 1, ...T \quad \hat{x}_{it} = \mathbf{u}_t'\hat{\mathbf{a}} + \sum_{k=1}^{K} \hat{\tau}_{ik}\hat{b}_{kt}. \tag{16}$$

Figure 8 shows input observations and estimated indoor temperature depending on the class to which each apartment belongs. In fact, the model computes probabilities of membership, but afterwards, in our case, each house is assigned

to a cluster. The variable $\hat{z}_i$ denotes the class to which the observation i belongs and is defined using the Maximum a posteriori estimation (MAP).

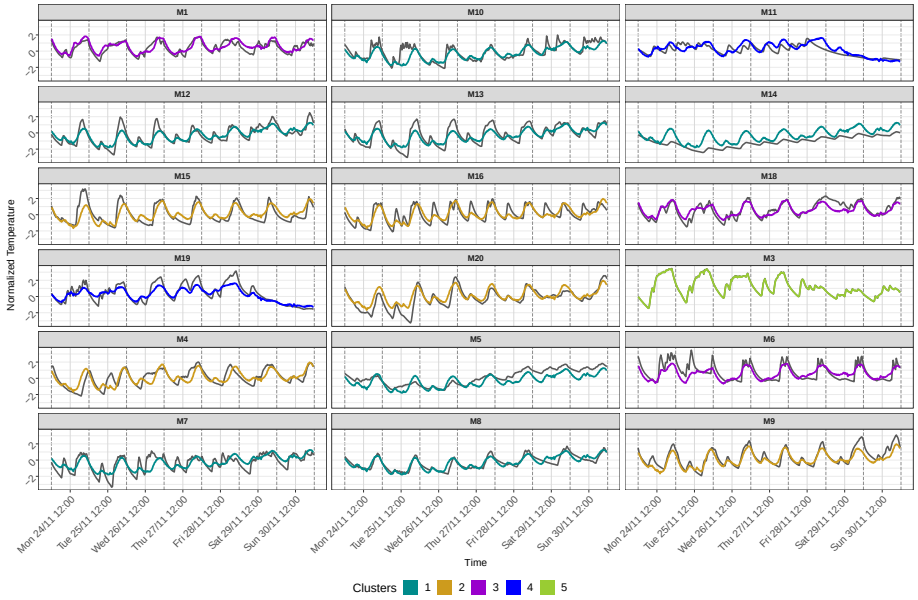

Figure 8: Real normalized indoor temperature for eighteen houses. As an overlay, the curves colored according to the class to which these houses belong represent the data estimated via the proposed model with K=5. The estimated data are obtained using the estimated effect of the exogenous factors, the class profiles and the membership class determined using the MAP.

Based on the results shown is Figure 9 and Figure 8, we first observe that cluster 5 includes only one house. Also, considering the four other clusters' dynamics, some dissimilarities can be observed. Neither the high nor the low peaks occur at the same time. There are also dissimilarities in the number and shape of the peaks. For cluster 2,for instance, we can observe one high peak at the end of each day and other smaller high peaks at the end of mornings during working days (Monday to Friday). For cluster 3 the high peaks seem to be longer than for clusters 1 and 2. In cluster 4, no significant peak variations can be observed. In addition, the dynamics change during the week-end for all of the clusters.

This period was chosen because motion detection data are also available. In fact, the cluster dynamics can be related to occupancy behaviors. Motion detection data were therefore used to confirm this intuition and interpret the profiles.

Figure 10 depicts the estimated cluster dynamics and average number of detections within the houses for each cluster. As can be seen, there is, a priori, a link between the number of motion detections and the identified class profile. The duration of the peak observed on the estimated class profiles can be related

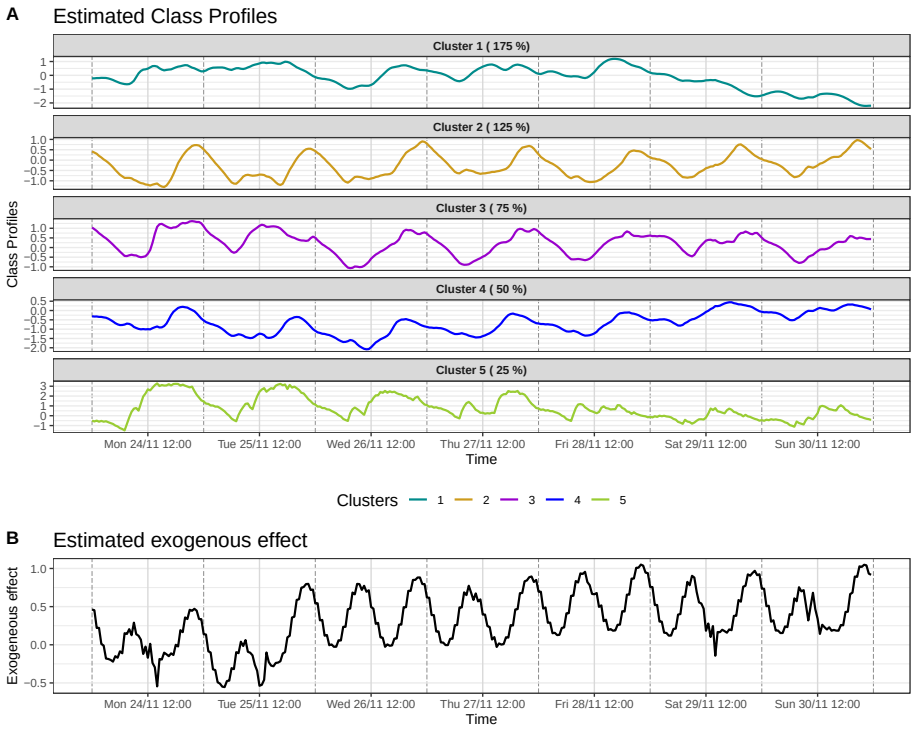

Figure 9: Estimated class profiles (A) and Estimated exogenous effect (B). Plot (A) represents class profiles estimated using the proposed model. Cluster labels were reorganized according to the class proportions. Plot (B) represents the estimated effect of exogenous factors. These results were obtained using the normalized meteorological data and hour variables multiplied by the estimated regression parameters.

to the period during which the number of observed movements is high. In the same way, the time of the peaks seems to be related to the times of the peaks of presence with perhaps a slight shift due to the inertia of the temperature for example. This link is even more obvious for profile 4 during the weekend, which decreases. There is also a very low average number of movements during the weekend for cluster 4.

Table 3 shows the calculated linear correlation coefficients between the class profiles of each group and the average number of motion detections in these groups. In addition, since there are temporal data, the correlation coefficient is also calculated between the class profiles and the average number of motion detections in the previous temporal sequence. Table 3 shows a higher correlation between the class dynamics and the average number of previous motion detections. This result may be due to temperature inertia. For example, if the inhabitants turn on the heat when they get home, there may be a timelag between that moment and the moment when the temperature increases.

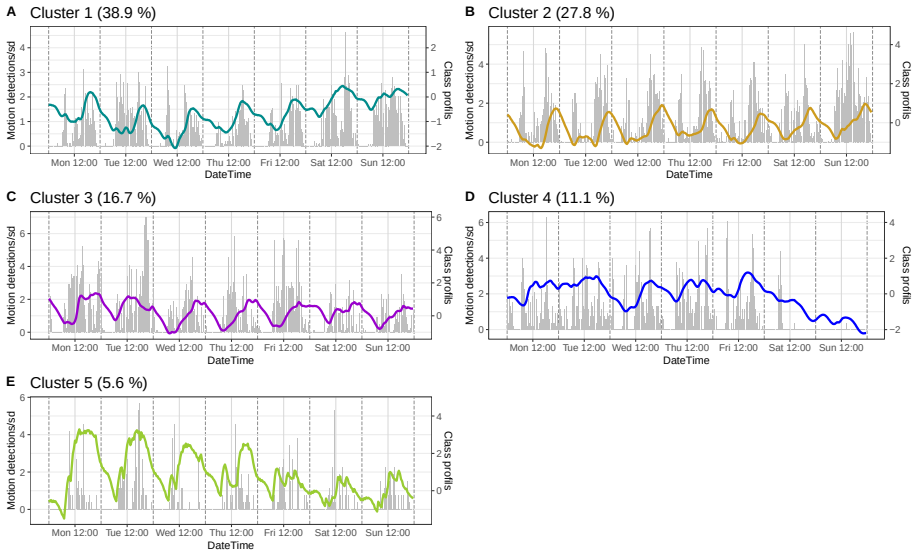

Figure 10: Class profiles and average reduced number of motion detections for each cluster during the period from the 24^{th} to the 30^{th} of November, 2014. Motion detections are recorded each time a movement is detected with a one-minute precision. For each house, motion detections have been aggregated into 30-minute intervals. The plots display, in gray, the average number of motion detections for each cluster. In order to adjust class profiles and motion detection data, the latter have been divided by the standard error.

	$(b_{kt}, \text{MOTION}_{kt})$				
Cluster	**1**	**2**	**3**	**4**	**5**
r	0.296	0.075	0.044	0.385	0.213
p.value	3.17e-08	0.167	0.416	2.54e-13	8.09e-05
	$(b_{kt}, \text{MOTION}_{kt-1})$				
Cluster	**1**	**2**	**3**	**4**	**5**
r	0.334	0.190	0.109	0.401	0.226
p.value	3.46e-10	4.79e-04	0.0455	2.34e-14	2.93e-05

Table 3: Linear correlation coefficients and p-values resulting from the independence test. The coefficient r is the linear correlation: $r = cov(x, y)/s_x s_y$, with s_x and s_y, respectively, the standard error computed on samples x and y. The p-value is the result of a statistical test with the null hypothesis of the nullity of the correlation.

To confirm this observation, table 3 also provides the p-values resulting from the independence tests performed. This test consists of a statistical test with the null hypothesis of the nullity of the correlation. For a small p-value, ($<$ 0.05), this hypothesis is rejected and we can conclude that the correlation is significantly different from zero. In this case, correlations between class profiles and lags of average motion detections are below the current threshold of 5%. This means that we cannot reject the hypothesis of a linear dependency between

class dynamics and the lag of the average number of motion detections for all of the 5 clusters.

These results allow us to interpret class profiles as occupancy patterns.

Cluster 1 shows peaks in the late afternoon/early evening and a higher level and less variation over the weekend. This indicates a more constant and high presence during the weekend. The periods of presence during the week are similar to office hours with inhabitants who probably do not work on weekends, which would explain the lower variations and higher level.

For cluster 2, the peaks of presence are observed during the evening (8pm/9pm) on weekdays and they are slightly earlier during the weekend. Moreover, during the weekend, the variations are lower, which also indicates a more constant presence on weekends. In addition, the slight peaks during the morning indicate a pattern of presence in the morning, but lower and shorter than in the evening. There appears to be a slight lag between the attendance peaks and the high peaks in the cluster profiles. This may be due to the time it takes for the room to heat up.

In cluster 3, the high peaks are quite long and start earlier in the week than for the previous profiles. This indicates that people are present from the end of the afternoon and during the evening. In addition, there is no peak during the morning. Both the level and the variations are lower during the weekend.

Profile 4 has fewer periodic peaks, indicating an average presence throughout the week. In addition, the level during the weekend is lower, which could indicate the absence of the inhabitants of these houses during the weekend.

Finally, cluster 5 is composed of only one house which can be characterized as atypical. The indoor temperature of this house presents marked variations during the beginning of the week and much less during the weekend. In addition, periods of presence are long and occur mainly during the afternoon of working days. However, the comparison between class profile and the number of motion detections for this cluster confirms the observation made previously.

Thus, to conclude this application of the classification model and estimation of the effect of exogenous factors as well as class dynamics, 5 classes were estimated and identified among 18 individual houses. The class centers were compared with motion detector data. This step allowed the interpretation of the clusters and class profiles in terms of house occupancy patterns, showing that each cluster highlights occupancy dynamics with different occupancy times.

6. Conclusion

This paper has presented a dynamic latent variable model to solve a classification problem by considering the evolution of class centers over time. The main objective of the model was to estimate class profiles as autoregressive processes. Moreover, the model is able to estimate the effect of known exogenous factors on the observations.

The comparison of the proposed model with a two-step regression model based on the k-means algorithm, and a constant class centers model based on the EM algorithm confirms the interest of considering dynamic class centers and of including estimation of the regression coefficients within the iterative algorithm as the proposed model shows better performances for the three selected criteria computed on various simulated and controlled datasets.

The open source database REFIT [16] was used to compare the three models on real data and to estimate class profiles and house clusters. The estimated class dynamics can be interpreted in terms of occupancy while the correlation between number of motion detections and class profiles is significantly different from zero. The identified profiles highlight different occupancy habits in terms of working days schedule and week-end presence.

7. Discussion

In this article, the number of clusters K is assumed to be known or is determined using the BIC criterion. Further investigations can be conducted in order to consider this hyper-parameter as a model parameter thanks to Bayesian methods.

The presented model represents a first step towards a more general model where exogenous effects are not global but specific to each cluster, which highlights structural effects within clusters. This model will be closer to mixture regression models with time-dependent factors.

One of the advantages of modeling class profiles is to use estimation to predict future profiles. The predictive aspect of this model has not been addressed in this paper but could be an interesting perspective. In addition, application on real indoor temperature data makes it possible to interpret class profiles in terms of occupancy habits. In future work, the estimated class dynamics could be used to estimate occupancy within dwellings.

Finally, the choice of modeling the class profiles from a first order autoregressive process is an issue for discussion. We can imagine an additional step in order to select a higher order that would be better adapted to the data. This requires either selecting this hyper-parameter in a previous step or developing a method to integrate the order of the processes as parameters of the model.

References

[1] Albert, A., Rajagopal, R., 2013. Smart meter driven segmentation: What your consumption says about you. IEEE Transactions on Power Systems 28, 4019–4030. 10.1109/TPWRS.2013.2266122.

[2] Amayri, M., Ploix, S., Kazmi, H., Ngo, Q.D., El Safadi, A., 2019. Estimating occupancy from measurements and knowledge using the bayesian network for energy management. Journal of Sensors 2019, 1–12. https://doi.org/10.1155/2019/7129872.

[3] Basford, K.E., McLachlan, G.J., 1985. The mixture method of clustering applied to three-way data. Journal of Classification 2, 109–125. https://doi.org/10.1007/BF01908066.

[4] Blei, D.M., Kucukelbir, A., McAuliffe, J.D., 2017. Variational inference: A review for statisticians. Journal of the American Statistical Association 112, 859–877. http://dx.doi.org/10.1080/01621459.2017.1285773.

[5] Bonhomme, S., Manresa, E., 2015. Grouped patterns of heterogeneity in panel data. Econometrica 83, 1147–1184. https://doi.org/10.3982/ECTA11319.

[6] Calabrese, A., Paninski, L., 2011. Kalman filter mixture model for spike sorting of non-stationary data. Journal of neuroscience methods 196, 159—169. https://doi.org/10.1016/j.jneumeth.2010.12.002.

[7] Celeux, G., Govaert, G., 1992. A classification em algorithm for clustering and two stochastic versions. Computational Statistics Data Analysis 14, 315–332. https://doi.org/10.1016/0167-9473(92)90042-E.

[8] Cheifetz, N., Noumir, Z., Same, A., Sandraz, A.C., Feliers, C., Heim, V., 2017. Modeling and clustering water demand patterns from real-world smart meter data. Drinking Water Engineering and Science Discussions , 1–12https://doi.org/10.5194/dwes-2017-19.

[9] Cote, J., Diana, M., 2017. Exploring the benefits of a traveller clustering approach based on multimodality attitudes and behaviours. Transportation Research Procedia 25, 2556–2569. https://doi.org/10.1016/j.trpro.2017.05.295.

[10] De Wilde, P., 2014. The gap between predicted and measured energy performance of buildings: A framework for investigation. Automation in Construction 41, 40–49. https://doi.org/10.1016/j.autcon.2014.02.009.

[11] Delff Andersen, P., Iversen, A., Madsen, H., Rode, C., 2014. Dynamic modeling of presence of occupants using inhomogeneous markov chains. Energy and Buildings doi=https://doi.org/10.1016/j.enbuild.2013.10.001.

[12] Desarbo, W., Cron, W., 1988. A conditional mixture maximum likelihood methodology for clusterwise linear regression. Journal of Classification 5, 249–282. https://doi.org/10.1007/BF01897167.

[13] Devijver, E., Goude, Y., Poggi, J.M., 2015. Clustering electricity consumers using high-dimensional regression mixture models. Applied Stochastic Models in Business and Industry 36. https://doi.org/10.1002/asmb.2453.

[14] Djatouti, Z., Waeytens, J., Chamoin, L., Chatellier, P., 2020. Thermal behavior of a two-story concrete building under controlled winter and heat wave scenarios in the sense-city equipment through temperature, flux and energy consumption dataset. Data in Brief 33, 106458. https://doi.org/10.1016/j.dib.2020.106458.

[15] El Assaad, H., Samé, A., Govaert, G., Aknin, P., 2016. A variational expectation–maximization algorithm for temporal data clustering. Computational Statistics Data Analysis 103, 206–228. https://doi.org/10.1016/j.csda.2016.05.007.

[16] Firth, S., Kane, T., Dimitriou, V., Hassan, T., Fouchal, F., Coleman, M., Webb, L., 2017. Refit smart home dataset. https://doi.org/10.17028/rd.lboro.2070091.v1.

[17] Hietaharju, P., Ruusunen, M., Leiviskä, K., 2018. A dynamic model for indoor temperature prediction in buildings. Energies 11. URL: https://www.mdpi.com/1996-1073/11/6/1477. https://doi.org/10.3390/en11061477.

[18] Jordan, M.I., Ghahramani, Z., Jaakola, T.S., Saul, L.K., 1999. An introduction to variational methods for graphical models. Machine Learning 37, 183–233. https://doi.org/10.1023/A:1007665907178.

[19] Kane, T., Firth, S., Hassan, T., Dimitriou, V., 2017. Heating behaviour in english homes: An assessment of indirect calculation methods. Energy and Buildings 148. https://doi.org/10.1016/j.enbuild.2017.04.059.

[20] Liisberg, J., Møller, J., Bloem, H., Cipriano, J., Mor, G., Madsen, H., 2016. Hidden markov models for indirect classification of occupant behaviour. Sustainable Cities and Society 27, 83–98. https://doi.org/10.1016/j.scs.2016.07.001.

[21] Mclachlan, G., Krishnan, T., 2007. The EM Algorithm and Extensions (Wiley Series in Probability and Statistics). https://doi.org/10.1002/9780470191613.

[22] McLachlan, G., Peel, D., 2004. Finite Mixture Models. Wiley Series in Probability and Statistics, Wiley. Url: https://books.google.fr/books?id=c2_fAox0DQoC.

[23] Rodrigues, F., Ferreira, B., 2016. Product recommendation based on shared customer's behaviour. Procedia Computer Science 100, 136–146. https://doi.org/10.1016/j.procs.2016.09.133.

[24] Ruiz, L., Pegalajar Jiménez, M.d.C., Arcucci, R., Molina-Solana, M., 2020. A time-series clustering methodology for knowledge extraction in energy consumption data. Expert Systems with Applications 160, 113731. https://doi.org/10.1016/j.eswa.2020.113731.

[25] Samé, A., Chamroukhi, F., Govaert, G., 2009. Modèle à processus latent et algorithme em pour la régression non linéaire. 41e Journée de Statistique, SFdS https://hal.inria.fr/inria-00386702/file/p132.pdf.

[26] Wang, S., Chaganty, A.T., Liang, P.S., 2015. Estimating mixture models via mixtures of polynomials. Advances in Neural Information Processing Systems 28, 487–495. https://proceedings.neurips.cc/paper/2015/file/8dd48d6a2e2cad213179a3992c0be53c-Paper.pdf.

[27] Wedel, M., Desarbo, W., 1995. A mixture likelihood approach for generalized linear models. Journal of Classification 12, 21–55. https://doi.org/10.1007/BF01202266.

[28] Widèn, J., Wäckelgård, E., 2010. A high-resolution stochastic model of domestic activity patterns and electricity demand. Applied Energy https://doi.org/10.1016/j.apenergy.2009.11.006.