

# Non-stationary data segmentation with hidden evidential semi-Markov chains

Clément Fernandes <sup>a</sup> and Wojciech Pieczynski <sup>b\*</sup>

<sup>a</sup>Department Automobiles, Segula Matra Automotive, Zone d'activité Pissaloup, 8 Av. Jean d'Alembert, 78190 Trappes, France

<sup>b</sup>SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France

\* Corresponding author. Tel.: +33 1 6 60 76 44 25

E-mail address: [Wojciech.Pieczynski@telecom-sudparis.eu](mailto:Wojciech.Pieczynski@telecom-sudparis.eu)

## Abstract

Hidden Markov chains (HMCs) are widely used in unsupervised Bayesian hidden discrete data restoration. They are very robust and, in spite of their simplicity, they are sufficiently efficient in many cases. However, in complex situations, extensions of HMCs models are of interest. In particular, when sojourn time in hidden states is not geometrical, hidden semi-Markov chains (HSMCs) may work better. Besides, hidden evidential Markov chains (HEMCs) showed its interest in non-stationary situations. In this paper, we propose a new model simultaneously extending HSMCs and HEMCs. Based on triplet Markov chains (TMCs), it is used in an unsupervised framework, parameters being estimated with the Expectation-Maximization (EM) algorithm. We validate its interest through some experiments on hand-drawn images noised with artificial noises.

*Keywords:* Hidden Markov chains, hidden semi-Markov chains, non-stationary data, theory of evidence, triplet Markov chains, unsupervised Bayesian segmentation.

## 1. Introduction

The problem dealt with by this article is to estimate a hidden realization of a random sequence from an observed one. This general problem is recurrent in many fields. One of them is signal segmentation, where the hidden sequence corresponds to labels assigned to each observation. One key aspect of the problem is the modelization of the probabilistic links between hidden and observed data. Indeed, the modelization chosen must be rich enough, to best reflect the reality that we try to model, but also simple enough, to ensure that the computations associated to the estimation and segmentation processes are feasible and relatively fast. One model that has been particularly popular in this context is the hidden Markov chain (HMC). The primary reason is that HMCs permit the computation of Bayesian solutions to the problem in a fast manner [1], [2], [3], [4], [5], among others. In particular, in the case of a discrete hidden sequence, HMCs allow Bayesian restoration through Maximum Posterior Mode (MPM), which consists in estimating each hidden label as the mode of the marginal law of each component of the hidden sequence, conditionally to the observations. This is the method we consider in this paper. Despite their simplicity, hidden Markov chains turn out to be very robust and are sufficiently efficient in many situations. Therefore, authors applied them in many fields such as image analysis [6], handwritten recognition [7], analysis of genome structure [8], transportation forecasting [9], weather [10] and financial forecasting [11], or still speech recognition [5] and synthesis [12]. However, in some complex situations, models that are more sophisticated may be of interest. Therefore, authors have extended HMCs in different directions. Among extensions, hidden semi-Markov chains (HSMCs) can be very useful as they allow the modelling of any sojourn time in a given class, when it is necessarily geometric in HMCs [13], [14], [15], [16]. Hidden bivariate Markov chains [17], [18], double Markov chains [19], or still pairwise Markov chains (PMCs) [20], [21], [22], [23] are other extensions. This paper is related to “triplet Markov chains” (TMCs), which is an extension of PMCs consisting in considering a third stochastic sequence, which might or might not have a practical signification, along the hidden sequence to be estimated and the sequence of observations, and assuming the joint Markovianity of the three sequences. The probabilistic dependencies of TMC, as well as those of HMC and PMC for comparison, are represented in Figure 1.

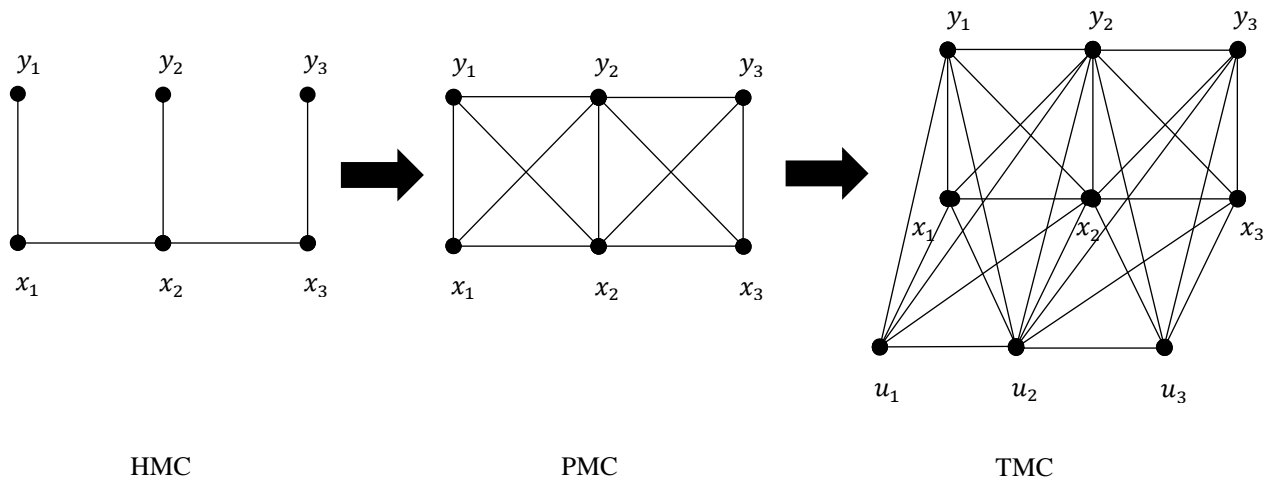


Figure 1. Probabilistic dependencies of, successively, hidden Markov chains (HMC), pairwise Markov chains (PMC) and triplet Markov chains (TMC). The sequence  $(y_1, y_2, y_3)$  is the sequence of observations,  $(x_1, x_2, x_3)$  corresponds to the hidden sequence that we want to estimate, and finally  $(u_1, u_2, u_3)$  is a third stochastic sequence that we add, when constructing a TMC.

To develop further, let us consider two stochastic sequences, one that correspond to hidden values  $X^N = (X_1, \dots, X_N)$  and one that models the observed values  $Y^N = (Y_1, \dots, Y_N)$ . For  $n = 1, \dots, N$ ,  $X_n$  takes its values in the set of classes  $\Omega = \{1, \dots, K\}$ , and  $Y_n$  takes its values in  $\mathbb{R}$ . With these notations, the problem is to estimate the hidden  $X^N = x^N$  from the observed  $Y^N = y^N$ . The “triplet Markov chains” (TMCs) consists in adding a third stochastic chain  $U^N = (U_1, \dots, U_N)$  and assuming that the triplet  $T^N = (X^N, U^N, Y^N)$  is Markovian. TMCs are very rich as  $U^N$  is arbitrary. However, when each  $U_n$  take its values in a finite set, MPM is still computable in a similar way it is done in HMCs and PMCs. TMCs have been successively used in image segmentation [24], [25], [26], normalized difference vegetation index modelling [27], activity classification [28], repayment of consumer loan modelling [29], non-stationary fuzzy data segmentation [30], [31], or still respiratory signal analysis [32]. As  $U^N$  is arbitrary in the general case, it may not have a practical interpretation. However, in practical applications, one often particularizes  $U^N$  such as it is at least partially interpretable. It is the case in the two TMCs we are interested in, which are discussed in the following.

The first TMC is constructed by considering Theory of Evidence (TE) of Dempster-Shafer [33], [34], [35], [36], [37], which is of interest in numerous situations. It is based on “plausibility” notion that can be seen, in some sense, as extension of probability. TE can be used in a Markovian context [38], [39], and in some situations introducing TE in the Markov chain framework results in a TMC, which turns out to be particularly interesting when dealing with non-stationary data [40], [41], [42]. We call the related model “hidden evidential Markov chain” (HEMC). In the latter, the variables  $U_n$  are particular random sets, taking their values in the power set  $2^\Omega$  of  $\Omega$ . Their joint law, which is Markovian, corresponds to prior defined on the power set of  $\Omega^N$  by a “basic belief assignment”. The deep reasons of the interest of these evidential models are not easy to understand and there is no theoretical results, at our knowledge, until now. There is only a partial result obtained in the simple case of independent variables given in [41], that we recall in paragraph 2.4.

The second TMC is constructed by noticing that hidden semi-Markov chains (HSMCs), which are well-known extensions of HMCs, can be considered as particular TMC. In usual HSMCs variables  $U_n$  take their values in  $\mathbb{N}$  and designate the remaining sojourn time in a given state. Taking as value of  $U_n$  the minimal remaining sojourn time instead of the exact remaining sojourn time, allows considering a finite set of values for variables  $U_n$ , and gives a TMC allowing MPM. The difference between such HSMC [43] considered here, and the classical approach is that when the sojourn time is up the state chain can continue remaining in the same state, while it must leave in the classical presentation. Thus, the set of possible durations is  $\mathbb{N}$  in both models.

Finally, one can consider TMCs with multivariate  $U^N$ . Then  $U^N = (U^{N,1}, \dots, U^{N,m})$ , with each  $U^{N,i} = (U_1^i, \dots, U_N^i)$  modelling some property of  $(X^N, Y^N)$ . For example, in [28] and [43] one considers  $U^N = (U^{N,1}, U^{N,2})$ , with  $U^{N,2}$  modeling the semi-Markovianity of  $X^N$ , while  $U^{N,1}$  models a noise with mixtures in [28], and models the presence of switches in [43].

In the present paper we consider  $U^N = (U^{N,1}, U^{N,2})$ , where  $U^{N,1} = (U_1^1, \dots, U_N^1)$  models the evidential nature of  $X^N$  with TE, while  $U^{N,2} = (U_1^2, \dots, U_N^2)$  extends to the semi-Markovianity the Markov chain  $(X^N, U^{N,1})$ . Thus, the new model, called “hidden

evidential semi-Markov chain” (HESMC) simultaneously extends HEMCs [41], [44] and HSMCs [13], [16]. As HEMCs and HSMCs showed their interest, our aim is thus to study whether the proposed new model can still simultaneously improve the efficiency of both of them. Of course, it is theoretically true as it extends both of them; however, the problem is to see whether this remains true in practice, in an unsupervised framework. Indeed, the new model is more complex than both HEMCs and HSMCs, implying the rapid growth of the number of parameters; then the problem is to verify that they are estimated robustly enough to ensure an efficient unsupervised segmentation. We show through experiments that the proposed adaptation of the well-known Expectation-Maximization (EM [45]) algorithm, which is frequently used in HMC [46], works with these triplets. By considering different experiments on hand-drawn images noisy with stationary or non-stationary noise, we show that the unsupervised segmentation based on HESMC can improve the results obtained by both hidden evidential Markov chain and by hidden semi-Markov chains. The practical interest of our model is potentially important, as it is an extension of both HEMC and HSMC, which are in turns two extensions of HMC. As the latter can be used in many applications, so can our model, and replacing HMC by HESMC in any application can potentially improve the results quite significantly, more so if the data considered are non-stationary.

The paper is organized as follows. We present the new model with related processing in section II. Section III is devoted to the parameter estimation algorithm, while section IV contains experiments. We present some conclusions and perspectives in the last section V.

## 2. Hidden Evidential semi-Markov Chain

### 2.1. Non stationarity and triplet Markov chains

Let us specify what “non-stationarity” means in this paper, and why it is of importance in image segmentation. We will consider two kinds of non-stationarity. In the first, described in [43], the model parameters switch at stochastic moments, while taking their value in a finite set of possible parameters. As they are stationary between two switches, such models are “piecewise stationary”. The second one is more general, the parameters can stochastically evolve in a continuous manner, or even in a hybrid manner.

Let us consider the classic hidden Markov chain model  $(X^N, Y^N)$ , where each  $X_n$  takes its values in  $\Omega = \{1, \dots, K\}$ , and  $Y_n$  takes its values in  $\mathbb{R}$ . Then the couple  $(X^N, Y^N)$  is Markovian, with distribution  $p(x^N, y^N)$  defined by  $p(x_1, y_1)$  and transitions of the form

$$p(x_{n+1}, y_{n+1} | x_n, y_n) = p(x_{n+1} | x_n) p(y_{n+1} | x_{n+1}) \quad (1)$$

To set these ideas on a simple example, let us assume  $p(y_n | x_n)$  Gaussian. The model is said “homogenous” when neither  $p(x_{n+1} | x_n)$  nor  $p(y_{n+1} | x_{n+1})$  (for  $n = 1, \dots, N - 1$ ) depend on  $n$ . Then the number of parameters defining  $p(x^N, y^N)$  is limited: there is one transition matrix of size  $K \times K$  for  $p(x_{n+1} | x_n)$ ,  $K$  means and  $K$  variances for Gaussian  $p(y_{n+1} | x_{n+1})$ , and, to define  $p(x_1, y_1)$ , there are  $K$  parameters defining  $p(x_1)$ , and  $K$  means and  $K$  variances defining  $p(y_1 | x_1)$  (note that  $p(y_1 | x_1)$  is possibly different from other  $p(y_{n+1} | x_{n+1})$ ). In homogeneous HMC the marginal distributions from  $p(x_n, y_n)$  can vary and can all be different. Stationary HMC are homogeneous HMCs such that  $p(x_n, y_n)$  do not depend on  $n = 1, \dots, N$ . Then the distribution  $p(x^N, y^N)$  of a stationary HMC is defined by  $p(x_1, y_1, x_2, y_2)$ , assumed equal to all  $p(x_n, y_n, x_{n+1}, y_{n+1})$ ,  $n = 2, \dots, N - 1$ . As  $p(x_1, y_1, x_2, y_2) = p(x_1, x_2) p(y_1 | x_1) p(y_2 | x_2)$ , we see that it is defined by a matrix of size  $K \times K$  giving  $p(x_1, x_2)$  - with identical margins  $p(x_1)$ ,  $p(x_2)$  -,  $K$  means and  $K$  variances giving Gaussian  $p(y_n | x_n)$ . To simplify, let us consider a stationary HMC, with the set of parameters denoted with  $\theta$ . Let us now imagine that there are two stationary HMCs given with parameters  $\theta_1, \theta_2$ , respectively. Let us consider HMC with  $p(x^N, y^N)$  defined by  $\theta_1$  for  $(x_1, y_1), \dots, (x_m, y_m)$ , and by  $\theta_2$  for  $(x_{m+1}, y_{m+1}), \dots, (x_N, y_N)$ . Such a HMC is called “switching” HMC. We can easily extend it to a finite number of switches, which we will assume in what follows. Such situations are usual in image segmentation. Let us consider two examples:

(i) Let  $y^N$  be a line of an optic aerial image with two classes “water” and “forest”, and let  $x^N$  be the line of hidden classes. Because of the clouds, the sun lights a part of the line, and another part is in the shade. Then the  $p(y_n | x_n)$  are different in lit or not lit parts, and thus we have two different  $\theta_1, \theta_2$ . Such a situation is thus a “noise switching” one [47].

A line of such an image is presented in Figure 2, (a).

(ii) Let  $y^N$  be as above, with two classes “urban area” and “trees”.  $N = 500$ , and the pixels in the image are of size 25 meters. There is a town in the image, with its center containing parks for  $(x_1, y_1), \dots, (x_{40}, y_{40})$ , its suburbs for  $(x_{41}, y_{41}), \dots, (x_{150}, y_{150})$ , a forest for  $(x_{151}, y_{151}), \dots, (x_{490}, y_{490})$ , and a village for  $(x_{491}, y_{491}), \dots, (x_{500}, y_{500})$ . Obviously, the  $p(x_{n+1} | x_n)$  are different on these four pieces of HMC  $(X^N, Y^N)$ . Such a situation is thus a “class transitions switching” one.

A line of such an image is presented in Figure 2 (b).

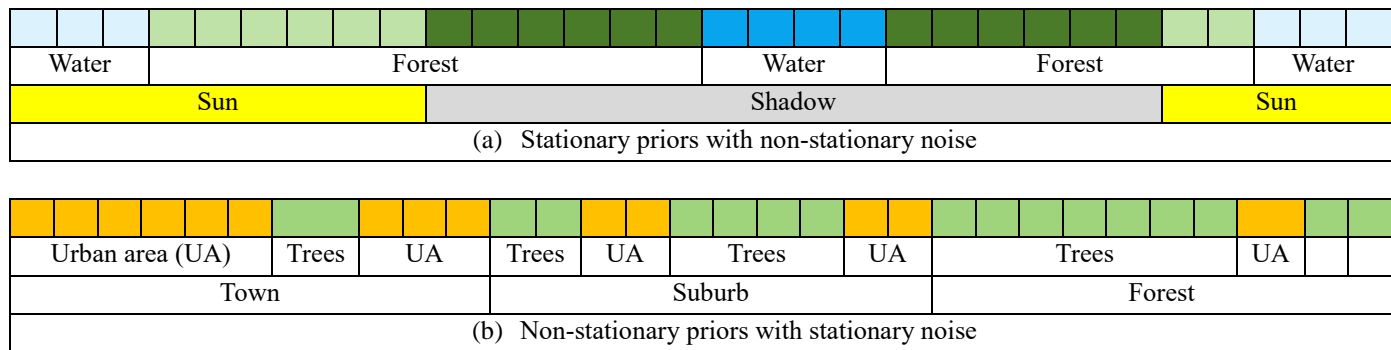


Figure 2. (a) line of non-stationary image partly in full sun and partly in shade. (b) line of image with non-stationary priors

Note that it is possible to deal with both situations simultaneously. When we know switches, we have a classic non-stationary HMC, which is piecewise stationary. The interesting case is when we do not know the switches. To find them by a probabilistic method, we have to define a random sequence, say  $U^N$ , taking its values in the set of switches  $\Theta = \{\theta_1, \dots, \theta_M\}$ , and its distribution  $p(u^N)$ . Usual way is setting it Markovian. Then the distribution of the triplet  $(X^N, U^N, Y^N)$  is  $p(x^N, u^N, y^N) = p(u^N)p(x^N, y^N|u^N)$ , with  $p(u^N)$  Markovian and  $p(x^N, y^N|u^N)$  a HMC (1). An example of such a model, in which, in addition  $(X^N, U^N, Y^N)$  is Markovian, is given by the following equation:

$$p(x^N, u^N, y^N) = p(u_1)p(x_1, y_1|x_1), \prod_{n=1}^{N-1} p(u_{n+1}|u_n)p(x_{n+1}|x_n, u_{n+1})p(y_{n+1}|x_{n+1}) \quad (2)$$

The interesting point is that (2) can be used to estimate  $(x^N, u^N)$  from  $y^N$ ; indeed, as described in the next subsection, this is possible because  $(X^N, U^N, Y^N)$  is a triplet Markov chain. Another interesting point is that  $(X^N, U^N, Y^N)$  is stationary, and thus, setting  $V^N=(X^N, U^N)$ ,  $(V^N, Y^N)$  is a stationary HMC. Then the parameters of  $p(x^N, u^N, y^N)$  can be estimated by the usual methods, for example of the Expectation-Maximization (EM) family ones. We see how a switching HMC, which is non-stationary conditionally on switches (but remains piecewise stationary), can be converted into a stationary TMC, allowing non-supervised (with parameters estimated from the sole  $y^N$ ) estimation of both  $x^N$  and switches  $u^N$ . It is worth noticing that in switching  $(X^N, U^N, Y^N)$  the couple  $(X^N, Y^N)$  is not Markovian in general, while it is always Markovian conditionally on  $U^N$ .

The second kind of non-stationarity is more general in that the parameters  $\theta$  are not necessarily in a finite set of values and can evolve continuously. For example, in the first example above, if the nature of clouds is such that they produce more or less shadows, the situations “lit” and “not lit” evolve continuously, with the shade more or less important. Then the parameters of  $p(y_{n+1}|x_{n+1})$  evolve in a continuous manner. Similarly, in the second example above, one can imagine that passing from the town center to the suburb implies continuous modification of  $p(x_{n+1}|x_n)$ . We can still model such situations with a triplet  $(X^N, U^N, Y^N)$  as above; however, for  $U^N$  taking its values in a continuous set like  $\mathbb{R}$ , analytical estimation of  $(x^N, u^N)$  from  $y^N$  is no longer computable in general. As recalled in subsection 2.4 below, one can deal with such situations using the so-called “hidden evidential Markov chains”, see [41] and references therein. On the contrary to the switching non-stationarity above, interpreting  $U^N$  is not immediate, and the reason why it allows improving non-stationary HMC segmentation is not clearly established. There is a very partial theoretical justification in [42], page 3093, example 3.3, in the simple case of two classes and independent variables  $(X_1, Y_1), \dots, (X_N, Y_N)$ .

## 2.2. MPM restoration in general Triplet Markov chain

Let  $T^N = (T_1, \dots, T_N)$  be a triplet Markov chain (TMC) as introduced above, with each  $U_n$  taking its values in  $\Lambda = \{1, \dots, L\}$ . Let us note  $p(t_1)$  the distribution of  $T_1$ , and  $p(t_{n+1}|t_n)$  distributions of  $T_{n+1}$  conditional on  $T_n = t_n$ . In the whole paper, we will consider a particular TMC, verifying

$$\begin{aligned} p(t_1) &= p(x_1, u_1)p(y_1|x_1), \\ p(t_{n+1}|t_n) &= p(x_{n+1}, u_{n+1}|x_n, u_n)p(y_{n+1}|x_{n+1}), \end{aligned} \quad (3)$$

Thus (3) is a simplified TMC, obtained by assuming  $p(x_{n+1}, u_{n+1} | x_n, u_n, y_n) = p(x_{n+1}, u_{n+1} | x_n, u_n)$ , and  $p(y_{n+1} | x_n, u_n, y_n, x_{n+1}, u_{n+1}) = p(y_{n+1} | x_{n+1})$ . The graphical probabilistic dependencies of such a model are represented in Figure 3.

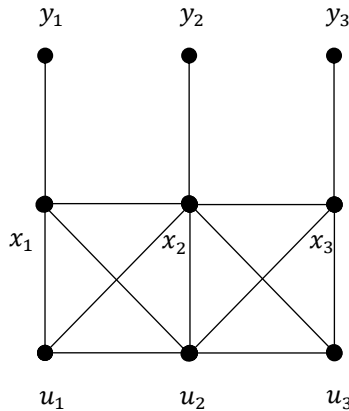


Figure 3. Probabilistic dependencies of a TMC verifying  $p(x_{n+1}, u_{n+1} | x_n, u_n, y_n) = p(x_{n+1}, u_{n+1} | x_n, u_n)$ , and  $p(y_{n+1} | x_n, u_n, y_n, x_{n+1}, u_{n+1}) = p(y_{n+1} | x_{n+1})$ . The sequence  $(y_1, y_2, y_3)$  is the sequence of observations,  $(x_1, x_2, x_3)$  corresponds to the hidden sequence that we want to estimate, and finally  $(u_1, u_2, u_3)$  is a third stochastic sequence that we add, when constructing a TMC.

An important point is that the Bayesian Maximum Posterior Mode (MPM) method  $\hat{S}_{MPM}$ , defined by

$$[\hat{S}_{MPM}(y^N) = \hat{x}^N] \Leftrightarrow \left[ p(\hat{x}_n | y^N) = \max_{x_n \in \Omega} p(x_n | y^N) \right], \quad (4)$$

is computable with a complexity linear in  $N$ . Indeed, the posterior marginal distributions  $p(x_n | y^N)$  are computable as follows:

(i) compute “forward” probabilities  $\alpha_n(x_n, u_n) = p(x_n, u_n, y_1, \dots, y_n)$  with

$$\alpha_1(x_1, u_1) = p(x_1, u_1, y_1);$$

$$\alpha_{n+1}(x_{n+1}, u_{n+1}) = \sum_{(x_n, u_n)} p(x_{n+1}, u_{n+1} | x_n, u_n) p(y_{n+1} | x_{n+1}) \alpha_n(x_n, u_n); \quad (5)$$

(ii) compute “backward” probabilities  $\beta_n(x_n, u_n) = p(y_{n+1}, \dots, y_N | x_n, u_n)$  with

$$\beta_N(x_N, u_N) = 1;$$

$$\beta_n(x_n, u_n) = \beta_{n+1}(x_{n+1}, u_{n+1}) \sum_{(x_n, u_n)} p(x_{n+1}, u_{n+1} | x_n, u_n) p(y_{n+1} | x_{n+1}); \quad (6)$$

(iii) compute  $p(x_n, u_n | y^N)$  with

$$p(x_n, u_n | y^N) = \frac{\alpha_n(x_n, u_n) \beta_n(x_n, u_n)}{\sum_{(x_n, u_n) \in \Omega \times \Lambda} \alpha_n(x_n, u_n) \beta_n(x_n, u_n)}; \quad (7)$$

(iv) compute  $p(x_n | y^N)$  with

$$p(x_n | y^N) = \sum_{u_n \in \Lambda} p(x_n, u_n | y^N). \quad (8)$$

The computation of the forward and backward probabilities in a triplet Markov chain verifying (3) are summarized in respectively algorithm 1 and algorithm 2. The overall procedure to calculate the MPM estimator in a triplet Markov chain verifying (3) is presented in algorithm 3.

**Inputs:** A sequence of observations  $\mathbf{y}^N = (y_1, \dots, y_N)$ ,  $|\Omega| |\Lambda|$  initial probabilities  $p(x_1, u_1)$ ,  $|\Omega| |\Lambda| \times |\Omega| |\Lambda|$  transition laws  $p(x_{n+1}, u_{n+1} | x_n, u_n)$ ,  $|\Omega|$  noise densities  $p(y_n | x_n)$ .

**Results:** A matrix  $|\Omega| |\Lambda| \times N$  of forward values

1. For each value of  $(x_1, u_1) \in \Omega \times \Lambda$ , compute  $\alpha_1(x_1, u_1) = p(x_1, u_1) p(y_1 | x_1)$
2. For n from N to 1:

For each value of  $(x_n, u_n) \in \Omega \times \Lambda$ , compute  $\alpha_{n+1}(x_{n+1}, u_{n+1})$  with (5)

**end**

**Algorithm 1:** Computation of forward probabilities in a triplet Markov chain verifying (3),  $|\cdot|$  is the cardinal function.

**Inputs:** A sequence of observations  $\mathbf{y}^N = (y_1, \dots, y_N)$ ,  $|\Omega||\Lambda|$  initial probabilities  $p(x_1, u_1)$ ,  $|\Omega||\Lambda| \times |\Omega||\Lambda|$  transition laws  $p(x_{n+1}, u_{n+1}|x_n, u_n)$ ,  $|\Omega|$  noise densities  $p(y_n|x_n)$ .  $|\cdot|$  is the cardinal function.

**Results:** A matrix  $|\Omega||\Lambda| \times N$  of backward values

3. **For each** value of  $(x_N, u_N) \in \Omega \times \Lambda$ , set  $\beta_N(x_N, u_N) = 1$

4. **For n** from N to 1:

For each value of  $(x_n, u_n) \in \Omega \times \Lambda$ , compute  $\beta_n(x_n, u_n)$  with (7)

**end**

**Algorithm 2:** Computation of backward probabilities in a triplet Markov chain verifying (3),  $|\cdot|$  is the cardinal function.

**Inputs:** A sequence of observations  $\mathbf{y}^N = (y_1, \dots, y_N)$ ,  $|\Omega||\Lambda|$  initial probabilities  $p(x_1, u_1)$  of size,  $|\Omega||\Lambda| \times |\Omega||\Lambda|$  transition laws  $p(x_{n+1}, u_{n+1}|x_n, u_n)$ ,  $|\Omega|$  noise densities  $p(y_n|x_n)$ .

**Results:** A sequence of estimated values  $\hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_N)$

1. Compute forward probabilities with **algorithm 1**

2. Compute backward probabilities with **algorithm 2**

3. **For n** from 1 to N:

**For each** value of  $(x_n, u_n) \in \Omega \times \Lambda$ , compute  $p(x_n, u_n|y^N)$  with (7)

**For each** value of  $x_n \in \Omega$ , compute  $p(x_n|y^N)$  with (8)

Set  $\hat{x}_n = \max_{x_n \in \Omega} p(x_n|y^N)$

**end**

**Algorithm 3:** Computation of the mode of posterior marginals estimator of  $x^N$  in a triplet Markov chain verifying (3),  $|\cdot|$  is the cardinal function.

### 2.3. MPM restoration with hidden semi-Markov chains

In stationary (thus also homogeneous) Markov chains  $p(x^N)$ , the sojourn time in a given class obeys a geometrical distribution. Indeed, knowing that the chain is in class  $k$  at  $n$  ( $x_n = k$ ), the probability that it remains in  $k$  exactly  $r$  times is  $p(x_{n+1} = x_{n+2} = \dots = x_{n+r} = k, x_{n+r+1} \neq k | x_n = k) = [p(x_{n+1} = k | x_n = k)]^r [1 - p(x_{n+1} = k | x_n = k)]$  (recall that the transitions  $p(x_{n+1} = \cdot | x_n = k)$  do not depend on  $n$ ). It means that this probability decreases “rapidly” when  $r$  increases. In images with large size class areas this can be a serious drawback. Of course, one can choose  $p(x_{n+1} = k | x_n = k)$  very close to 1, but then it will be poorly adapted to the parts of the images containing small size class areas. To remedy this, Markov chains have been extended to semi-Markov chains. The idea is to introduce a random variable  $U$  valued in  $\mathbb{N}$  with any distribution, non-necessarily geometric, to model the exact sojourn time in a given class. Then for  $x_n = k$  and  $x_{n+1} \neq k$ , the realization  $U = r$  means that  $x_{n+1} = x_{n+2} = \dots = x_{n+r} = k, x_{n+r+1} \neq k$ . To avoid  $\mathbb{N}$  and replace it by a finite set, allowing introduction of triplet Markov chains, we consider here a slightly different semi-Markov model [28], [43]. In the model considered  $U_n = u_n$  is the minimal remaining sojourn time instead of the exact remaining sojourn time.

More precisely, let  $(X^N, Y^N)$  be random sequences as above, with  $X^N$  hidden and  $Y^N$  observed. We introduce a third random sequence  $U^N = (U_1, \dots, U_N)$  taking its values in  $\Lambda = \{0, 1, \dots, D - 1\}$ , such that  $U_n = d \neq 0$  means that  $X_n$  will remain in the same state *at least*  $d$  times. Therefore  $(X_n, U_n) = (x_n, d)$  implies  $x_n = x_{n+1} = x_{n+2} = \dots = x_{n+d}$ . Thus for  $d = 0$  we can have  $x_{n+1} = x_n$  or not:  $d$  is the minimum remaining time of sojourn in the current state, not the exact one as it is usual in classic semi-Markov models. This allows considering a finite  $\Lambda$ . For  $a, b, c, d$  let us set  $\delta_{(a,b)}(c, d) = 1$  if  $(a, b) = (c, d)$ , and  $\delta_{(a,b)}(c, d) = 0$  if  $(a, b) \neq (c, d)$ . Then the distribution of the hidden semi-Markov  $(X^N, Y^N)$  considered is the marginal distribution of the TMC  $(X^N, U^N, Y^N)$  distribution defined by  $p(x_1, u_1, y_1)$  and transitions (3) with

$$p(x_{n+1}, u_{n+1} | x_n, u_n) = \begin{cases} \delta_{(x_{n+1}, u_{n+1})}(x_n, u_n - 1) & \text{if } u_n > 0 \\ p^*(x_{n+1}, u_{n+1} | x_n, u_n = 0) & \text{if } u_n = 0 \end{cases}; \quad (9)$$

$$p(y_{n+1} | x_{n+1}, u_{n+1}) = p(y_{n+1} | x_{n+1}).$$

(9) means that for  $u_n = 0$ , the next  $(x_{n+1}, u_{n+1})$  is driven by  $p^*(x_{n+1}, u_{n+1} | x_n, u_n = 0)$  and for  $u_n \neq 0$ ,  $x_{n+1} = x_n$  and  $u_{n+1} = u_n - 1$ . The graphical probabilistic dependencies of a hidden semi-Markov chain are the same than the general model presented in subsection 2.2, and therefore can also be observed on Figure 3. Finally,  $(X^N, U^N, Y^N)$  is a TMC verifying (3), and thus (4) is computable with (5)-(8).

#### 2.4. MPM restoration with hidden evidential Markov chains

Let us consider a finite set  $\Omega = \{1, \dots, K\}$  and  $2^\Omega$  its power set. In theory of evidence one considers “basic belief assignments” - or “masses” -  $m$  which are functions from  $2^\Omega$  to  $[0, 1]$  verifying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Omega} m(A) = 1$ . Such masses are extensions of probability distribution; indeed,  $m$  null outside singletons are probabilities. Basic belief assignments are very useful in merging heterogeneous information sources, which is achieved by the following “Dempster-Shafer fusion” (DS fusion).

For two masses  $m_1, m_2$ , their DS fusion  $m = m_1 \oplus m_2$  is defined with

$$m(A) = (m_1 \oplus m_2)(A) = C \sum_{A_1 \cap A_2 = A} m_1(A_1) m_2(A_2) \quad (10)$$

The sum in (10) is taken over all couples  $(A_1, A_2)$  in  $2^\Omega \times 2^\Omega$  such that  $A_1 \cap A_2 = A$ , and  $C$  is the normalizing constant making the sum of  $m(A)$  over  $2^\Omega$  equal to 1.

##### Example 1

For  $p$  probability on  $\{1, \dots, K\}$ ,  $m$  mass on  $\{\{1\}, \dots, \{K\}, \{1, \dots, K\}\}$ ,  $p \oplus m$  is the probability on  $\{1, \dots, K\}$  given with  $(p \oplus m)(\{k\}) = Cp(\{k\})(m(\{k\}) + m(\{1, \dots, K\}))$ , with  $\frac{1}{c} = \sum_{k=1}^K p(\{k\})(m(\{k\}) + m(\{1, \dots, K\}))$ .

One can see that DS fusion of a probability – i.e. mass null outside singletons – with any other mass is a probability. In this paper we will use in experiments the masses null outside singletons and  $\{1, \dots, K\}$  from Example 1. We show in Example 2 below how replacing erroneous priors by such masses can improve Bayesian segmentation in non-stationary case of independent data  $(X_1, Y_1), \dots, (X_N, Y_N)$ . This simple example is a kind of justification for the use of evidential Markov models; however, the interest of the latter models for segmentation of non-stationary data is observed experimentally, and, to our knowledge, has not yet been demonstrated.

##### Example 2

Let us consider the following example extracted from [42].  $(X_1, Y_1), \dots, (X_N, Y_N)$  is a sequence of independent random variables, with each  $X_n$  taking its values in  $\Omega = \{\omega_1, \omega_2\}$ , and each  $Y_n$  taking its values in  $[0, 1]$ . Let us assume that  $p_n = P[X_n = \omega_1]$  depends on  $n$ , but the two densities of the distributions  $f_1(y) = 2(1 - y)$  and  $f_2(y) = y$  – which models  $p(y_n | x_n)$  – do not depend on  $n$ . When using the true parameter  $p_n$ , the Bayesian restoration corresponding to the classical “0–1” loss function is  $d(y) = \omega_1$  if  $y \leq p_n$ , and  $d(y) = \omega_2$  if  $y \geq p_n$ . This gives the error probability  $Er(p_n) = p_n \int_{p_n}^1 2(1 - y)dy + (1 - p_n) \int_0^{p_n} 2ydy = p_n(1 - p_n)$ . When using a false  $r$  instead of  $p_n$ , the error probability, given with similar calculus, becomes  $Er(p_n, r) = p_n \int_r^1 2(1 - y)dy + (1 - p_n) \int_0^r 2ydy = p_n(1 - p_n) + (p_n - r)^2$ . Finally, according to what will be done in the Markov context below, we replace the false  $r = P[X_n = \omega_1]$ ,  $1 - r = P[X_n = \omega_2]$  with a basic belief assignment  $m(\{\omega_1\}) = r - t$ ,  $m(\{\omega_2\}) = 1 - r - t$ ,  $m(\{\omega_1, \omega_2\}) = 2t$  obtained by “weakening” the probability  $r$ ,  $1 - r$ . The Dempster-Shafer fusion  $m \oplus q$ , where  $q(\omega_1) =$

$\frac{f_1(y)}{f_1(y)+f_2(y)}, q(\omega_2) = \frac{f_2(y)}{f_1(y)+f_2(y)}$ , is then a probability extending the posterior probability, which is found again for  $t = r - p_n$ . Using the latter probability  $m \oplus q$  to perform the restoration - that is to say, putting  $d^*(y) = \omega_1$  if  $m \oplus q(\{\omega_1\}) \leq m \oplus q(\{\omega_2\})$ , and  $d^*(y) = \omega_2$  if  $m \oplus q(\{\omega_1\}) \geq m \oplus q(\{\omega_2\})$  -, gives  $d^*(y) = \omega_1$  if  $y \leq \frac{r+t}{1+2t}$ , and  $d^*(y) = \omega_2$  if  $y \geq \frac{r+t}{1+2t}$ . With a calculation similar to the one above, this leads to the error probability  $Er(p_n, r, t) = p_n(1 - p_n) + (p_n - \frac{r+t}{1+2t})^2$ . Therefore, the problem is to know whether a  $t > 0$  such that  $Er(p_n, r, t) < Er(p_n, r)$  - which is equivalent to  $(p_n - \frac{r+t}{1+2t})^2 < (p_n - r)^2$  - does exist. In other words, is it possible to decrease the error probability by introducing a basic belief assignment  $m$  when having a false ? The response is positive in the following context. As  $p_1, \dots, p_N$  are not known and  $p_n$  can vary with  $n$ , let us assume that they do vary with  $n$  and that they are realizations of a random variable  $W$ , with  $E[W] = 0.5$ . The problem is then to see whether the error expectation  $E[Er(W, r, t)] = 0.5 - \frac{r+t}{1+2t} + (\frac{r+t}{1+2t})^2$  decreases when using the basic belief assignment  $m$  instead of  $r$  or, in other words, when  $t$  starts from 0. A classical calculus leads to  $\frac{d}{dt}(E[Er(W, r, t)])(0) = -(1 - 2r)^2$ , which shows that the “mean” error decreases, and thus, for an  $N$  large enough, the error also decreases when using  $m$  instead of  $r$ .

Let  $(X^N, Y^N)$  be random sequences as above with  $X^N$  hidden and  $Y^N$  observed. We wish estimate  $X^N$  from  $Y^N$  using the classic hidden Markov model. However,  $X^N$  in non-stationary, we don't know its distribution, and only  $Y^N$  is observed. If we consider  $X^N$  as stationary with margins  $p(x_n)$  equal to some  $p^*$ , we will have the problem  $p(x_n) \neq p^*$  mentioned above for each  $n$ . The idea is to try applying the idea of Example 2 in Markov context. To do so, one possibility is to consider an “evidential” Markov model, in which marginal distributions  $p(x_n)$  would be extended to evidential masses  $m(A_n)$  on  $2^\Omega$ , similarly to what is made in Example 2 above. The constraint to be respected is that when  $m(A_n) = 0$  outside singletons, then the model gives again the classic hidden Markov chain. We propose to use the following Evidential Markov Chain. First, the Markov distribution  $p(x^N)$  is extended to an evidential mass on  $[2^\Omega]^N$ :

$$m(u^N) = m(u_1) \prod_{i=1}^{N-1} m(u_{i+1}|u_i). \tag{11}$$

*Example 3*

Let set  $\Omega = \{1,2,3\}$ . Consider that each  $U_n$  takes values in the following subset of  $2^\Omega$ :  $A = \{\{1\}, \{2\}, \{1,3\}, \{1,2,3\}\}$ . Setting  $A = \{a_1, a_2, a_3, a_4\}$ ,  $m(u^N)$  simply is a classic Markov chain defined on  $A^N$ .

For  $\Omega = \{1, \dots, K\}$ , if  $m(u_1) = 0$  and  $m(u_{n+1}|u_n) = 0$  outside singletons  $\{1\}, \dots, \{K\}$ ,  $m$  is equivalent to a classic Markov chain. Then the evidential Markov chain (EMC) is defined as a couple  $(X^N, U^N)$  of random variables valuable in  $\Omega^N \times [2^\Omega]^N$ , whose probabilistic distribution is defined such as:

$$p(x^N, u^N) = m(u_1)p(x_1|u_1) \prod_{i=1}^{N-1} m(u_{i+1}|u_i) p(x_{i+1}|u_{i+1}), \tag{12}$$

with

$$p(x_{n+1}|u_{n+1}) = \frac{1_{[x_{n+1} \in u_{n+1}]}}{|u_{n+1}|}, \tag{13}$$

*Example 4*

Let us assume that  $m(u^N)$  is a classic Markov chain defined on  $A^N$ , with  $A = \{\{1\}, \{2\}, \{1,3\}, \{1,2,3\}\}$  as in Exemple 3. Then, according to (13), each  $(X_n, U_n)$  takes its values in

$$B = \{(1, \{1\}), (1, \{1,3\}), (1, \{1,2,3\}), (2, \{2\}), (2, \{1,2,3\}), (3, \{3\}), (3, \{1,3\}), (3, \{1,2,3\})\} = \{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\},$$

and thus  $p(x^N, u^N)$  is a classic Markov chain on  $B^N$ . Its transitions are defined with  $m(u_{n+1}|u_n)$  - which are estimated with EM

in experiments section - and transitions  $p(x_{n+1}|u_{n+1})$ , which are, accordingly to (13), of the form  $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$ .

Let us notice:



(i) EMC (12) considered in this paper has the structure of the very classic HMC, where  $x^N$  would be “observed” and  $u^N$  “hidden”; more complex EMC are possible [41],

(ii) It is well known that in HMC the observed chain is not Markov in general, so that EMC is a couple  $(X^N, U^N)$  with  $X^N$  non-Markov in general. However, important is that the couple  $(X^N, U^N)$  is Markov.

The couple  $(X^N, U^N)$  being Markov, we define hidden EMC as a classic HMC with  $(X^N, U^N)$  hidden and  $Y^N$  observed. More precisely, we consider

$$p(x^N, u^N, y^N) = m(u_1)p(x_1|u_1)p(y_1|x_1) \prod_{i=1}^{N-1} m(u_{n+1}|u_n) p(x_{n+1}|u_{n+1})p(y_{n+1}|x_{n+1}), \quad (14)$$

We can notice that (14) is a particular HMC in which  $p(y_{n+1}|u_{n+1}, x_{n+1}) = p(y_{n+1}|x_{n+1})$ .

Note that is is very easy to sample realizations of such a hidden EMC: one samples the classical Markov chain  $U^N = u^N$  according to (11), then  $X^N = x^N$  is sampled – using  $u^N$  - accordingly to (13), and finally  $Y^N = y^N$  is sampled according to Gaussian distributions  $p(y_1|x_1), \dots, p(y_N|x_N)$ .

Important is that when if  $m(u_1) = 0$  and  $m(u_{n+1}|u_n) = 0$  outside singletons  $\{1\}, \dots, \{K\}$ , we have, according to (13),  $p(x_{n+1}|u_{n+1}) = 1$ , and thus  $x_{n+1} = u_{n+1}$ , and (14) become the classic hidden Markov chain (1).

Now, as (14) is a TMC,  $p(x_n, u_n|y^N)$  are computable with (5)-(7), which allows computation of  $p(x_n|y^N)$  with (8), which is here

$$p(x_n|y^N) = \sum_{u_n \in 2^\Omega} p(x_n, u_n|y^N), \quad (15)$$

making the use of MPM (4) possible. The overall procedure of computing the MPM estimator is the same than presented in algorithm 3, with  $p(x_1, u_1) = m(u_1) \frac{1_{[x_1 \in u_1]}}{|u_1|}$  and  $p(x_{n+1}, u_{n+1}|x_n, u_n) = m(u_{n+1}|u_n) \frac{1_{[x_{n+1} \in u_{n+1}]}}{|u_{n+1}|}$  and  $\Lambda = 2^\Omega$ . In addition, is it possible to use parameter estimation methods like EM, which makes the MPM unsupervised.

Let us notice that one can also estimate the realization of latent  $U^N$ . Indeed, similarly to (15) giving  $p(x_n|y^N)$ , we can compute  $p(u_n|y^N)$ :

$$p(u_n|y^N) = \sum_{x_n \in \Omega} p(x_n, u_n|y^N). \quad (16)$$

Then realization  $u^N$  of  $U^N$  can be estimated with MPM. Such estimates are presented in [41], page 1608. They are somewhat difficult to interpret. Let us notice that understanding relationship between estimated  $u^N$  and  $x^N$  is probably the key to understand the deep reasons of the efficiency of hidden evidential Markov chains in unsupervised non-stationary data segmentation.

Given the intended application, we will consider Markovian mass (11) of the simple form:  $m(u_1) = 0$  and  $m(u_{n+1}|u_n) = 0$  for  $u_1$  and  $u_{n+1}$  outside  $\{\{1\}, \dots, \{K\}, \{1, \dots, K\}\}$ . Then considering marginal distributions and  $K = 2$  we find again, under independence hypothesis, example 1. However, model (12)-(14) remains valid for any  $m(u^N)$  of form (12), with potential applications to any problem involving non-stationary hidden data processed using the classic hidden Markov (1).

Finally, the overall procedure for setting up unsupervised segmentation of data  $y^N = (y_1, \dots, y_N)$  with non-stationary priors is as follows. The set of classes  $\{1, \dots, K\}$  and probabilities  $p(y_n|x_n)$ , independent from  $n$ , are given.

- consider  $U^N = (U_1, \dots, U_N)$ , each  $U_n$  taking its values in  $\{\{1\}, \dots, \{K\}, \{1, \dots, K\}\}$ ;
- consider  $p(x^N, u^N, y^N)$  defined with (14) as a classic hidden Markov chain  $p(v^N, y^N)$ , with  $v^N = (x^N, u^N)$  Markov chain, with each  $(x_n, u_n)$  in  $\{1, \dots, K\} \times \{\{1\}, \dots, \{K\}, \{1, \dots, K\}\}$ ;
- apply classic EM to estimate the parameters of  $(x^N, u^N, y^N)$ ;
- apply (5)-(7) to compute  $p(x_n, u_n|y^N)$ ;
- apply (8) to compute  $p(x_n|y^N)$ ;
- apply (4) to compute MPM.

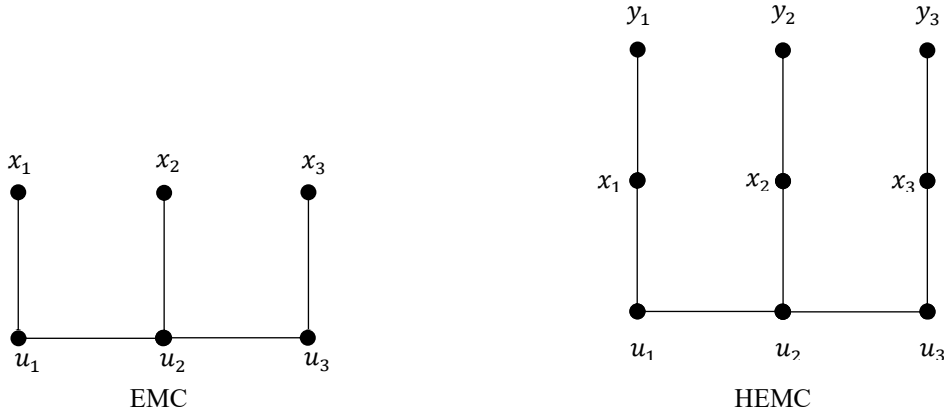


Figure 4. Probabilistic dependencies of a hidden evidential Markov chain. The sequence  $(y_1, y_2, y_3)$  is the sequence of observations,  $(x_1, x_2, x_3)$  corresponds to the hidden sequence that we want to estimate, and finally  $(u_1, u_2, u_3)$  is a third stochastic sequence that we add, when constructing a TMC.

### 2.5. MPM restoration with hidden evidential semi-Markov chains

Let  $(X^N, Y^N)$  be random sequences as in the previous sections, with  $X^N$  hidden and  $Y^N$  observed. We recalled in subsection 2.4 above the construction of an evidential Markov chain  $V^N = (X^N, U^{N,1})$ , with  $U^{N,1} = (U_1^1, \dots, U_N^1)$  and each  $U_n^1$  taking its values in the power set  $2^\Omega$ . The triplet  $(V^N, Y^N) = (X^N, U^{N,1}, Y^N)$  obtained with (2) is then a hidden Markov chain. To obtain the new hidden evidential semi-Markov chains (HESMC) we propose, we apply the introduction of the semi-Markovianity described in subsection 2.3 to the HMC  $(V^N, Y^N)$ . Thus, we introduce a second random sequence  $U^{N,2} = (U_1^2, \dots, U_N^2)$  taking its values in  $\Lambda = \{0, 1, \dots, D-1\}$  such as (9) be verified with  $X^N$  replaced by  $V^N$ , and  $U^N$  replaced by  $U^{N,2}$ . So that the distribution of the HESMC  $T^N = (V^N, U^{N,2}, Y^N) = (X^N, U^{N,1}, U^{N,2}, Y^N)$  is defined by

$$\begin{aligned}
 p(t_1) &= p(u_1^1, u_1^2)p(x_1|u_1^1)p(y_1|x_1); \\
 p(t_{n+1}|t_n) &= p(u_{n+1}^1, u_{n+1}^2|u_n^1, u_n^2)p(x_{n+1}|u_{n+1}^1)p(y_{n+1}|x_{n+1}),
 \end{aligned} \tag{17}$$

with

$$\begin{aligned}
 p(u_{n+1}^1, u_{n+1}^2|u_n^1, u_n^2) &= \begin{cases} \delta_{(u_{n+1}^1, u_{n+1}^2)}(u_n^1, u_n^2 - 1) & \text{if } u_n^2 > 0 \\ p^*(u_{n+1}^1, u_{n+1}^2|u_n^1, u_n^2 = 0) & \text{if } u_n^2 = 0 \end{cases} \\
 p(x_n|u_n^1) &= \frac{1_{[x_n \in u_n^1]}}{|u_n^1|} \text{ for } n = 1, \dots, N.
 \end{aligned}$$

Finally, setting  $U^N = (U^{N,1}, U^{N,2})$ ,  $(X^N, U^N, Y^N)$  is a TMC verifying (3), and thus (4) can be computed with (5)-(8). The graphical probabilistic dependencies of a hidden evidential semi-Markov chain are presented in Figure 5.

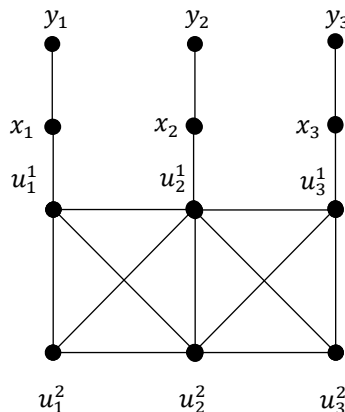


Figure 5. Probabilistic dependencies of a hidden evidential semi-Markov chain. The sequence  $(y_1, y_2, y_3)$  is the sequence of observations,  $(x_1, x_2, x_3)$  corresponds to the hidden sequence that we want to estimate,  $(u_1^1, u_2^1, u_3^1)$  is a stochastic sequence added to model the evidential part, and  $(u_1^2, u_2^2, u_3^2)$  is a stochastic sequence added to model the semi-Markovianity of  $(u_1^1, u_2^1, u_3^1)$

## 2.6. Comparison with previous semi-Markov non stationary chains

There are two other papers published with similar key words; indeed, both deal with non-stationary data segmentation [43], [48] using hidden semi-Markov chains and Dempster-Shafer fusion. Let us specify what the novelties of the present paper are with respect to them. As specified in subsection 2.1 above, [43] deals with “piece-wise” stationarity of the HSMCs, which is different from the “general” non-stationarity considered in this paper. On the contrary, the model discussed in [48] considers semi-Markov chains, non-stationary data, and theory of evidence, likely the model proposed in the present paper. However, the two models are fundamentally different.

More precisely the difference lies in the following. In both models there are a hidden chain  $X^N$  and an observed chain  $Y^N$ . In this paper we first consider an evidential Markov chain  $V^N = (X^N, U^{N,1})$ . Then we extend the Markov chain  $V^N$  to a semi-Markov chain, by introducing the random variables  $U^{N,2}$  modelling the minimal remaining sojourn time. In the final step, the semi-Markov chain  $(V^N, U^{N,2})$  is considered as hidden. Thus, joined to the observation process  $Y^N$ , it gives the hidden evidential semi-Markov chain  $(V^N, U^{N,2}, Y^N)$  (HESMC) proposed in the present paper. Here, variables  $X_n, U_n^1, U_n^2, R_n$  take their values in  $\Omega = \{1, \dots, K\}$ ,  $2^\Omega$ ,  $\Lambda = \{0, 1, \dots, D - 1\}$  and  $\mathbb{R}$ , respectively.

In [48] we first consider a semi-Markov chain  $X^N$ , whose distribution is the marginal distribution of a Markov chain  $W^N = (X^N, U^{N,2})$  valued in  $\Omega \times \Lambda$ .  $U^{N,2}$  models the minimal remaining sojourn time as above, however, the set of classes is here  $\Omega$ , while it is  $2^\Omega$  in the case of HESMC above. Then  $W^N$  is extended to an evidential Markov chain  $(W^N, U^{N,1})$ . Adding the observation chain  $Y^N$ , we finally obtain hidden semi-Markov evidential chain (HSMEC)  $(W^N, U^{N,1}, Y^N)$ . Thus, in the model discussed in [48] variables  $X_n, U_n^1, U_n^2, R_n$  take their values in  $\Omega = \{1, \dots, K\}$ ,  $2^{\Omega \times \Lambda}$ ,  $\Lambda = \{0, 1, \dots, D - 1\}$ , and  $\mathbb{R}$ , respectively.

We see that the variables  $U_n^1$  take their values in different spaces, which is  $2^\Omega$  for the present model HESMC, and  $2^{\Omega \times \Lambda}$  for the HSEMC proposed in [48]. This implies that the two models - we call “models” distributions of  $(X^N, Y^N)$  - are necessarily different in general.

For example, let us consider  $\Omega = \{1, 2\}$ ,  $\Lambda = \{0, 1, 2\}$ . Concerning HESMC, let  $U_n^1$  takes its values in  $\{\{1\}, \{2\}, \{1,2\}\}$ . Then, according to Dempster-Shafer fusion rule,  $(X_n, U_n^1)$  takes its values in  $A = \{(1, \{1\}), (1, \{1,2\}), (2, \{2\}), (2, \{1,2\})\}$ . Thus  $(X^N, U^{N,1}, U^{N,2})$  takes its values in  $A \times \Lambda$ , and thus the related HESMC  $(X^N, U^{N,1}, U^{N,2}, Y^N)$  can be considered as a classical HMC with 12 classes. Concerning HSMEC, we first consider  $(X^N, U^{N,2})$  valued in  $\Omega \times \Lambda$ , which contains 6 elements. Then  $(X^N, U^{N,1}, U^{N,2})$  takes its values in  $2^{\Omega \times \Lambda}$  containing 56 elements, so that HSEMC can be considered as a classical HMC with 56 classes.

Let us notice the following:

- HESMC simultaneously extends both HEMC and HSMC. Indeed, taking the bba  $m$  null outside singletons, HESMC is a HSMC; taking  $p(u_n^2 \neq 0) = 0$  for each  $n$ , HESMC is a HEMC.

- According to the dependence graph in Figure 5, the considered HESMC is a quite particular HESMC; the most general one is obtained by applying the semi-Markovianity to  $V^N = (X^N, U^{N,1})$ , not only to  $U^{N,1}$ . Of course, parameter estimation would be more difficult, but if EM continues its excellent behavior, extending the model used in the experiments is likely to further improve its efficiency.

- TMC  $T^N = (X^N, U^N, Y^N)$  (3) considered in the paper is a very particular one:  $(X^N, U^N)$  is markovian, and  $p(y_n | x^N, u^N) = p(y_n | x_n)$ . However,  $X^N$  is not necessarily Markovian and thus  $(X^N, Y^N)$  is not necessarily a HMC.

## 3. HESMC Parameter Estimation with EM

We will say that HESMC is “Gaussian” if  $p(y_{n+1} | x_{n+1})$  in (17) are Gaussian. Such a Gaussian HESMC can be considered as a particular classic hidden Markov chain  $(W^N, Y^N)$ , with  $W^N = (X^N, U^{N,1}, U^{N,2})$ . Then one can estimate all parameters with the classic version of Expectation-Maximization (EM) algorithm, whose great efficiency in HMCs is well known.

Let us set  $\Omega = \{1, \dots, K\}$ ,  $\Lambda = \{0, 1, \dots, D - 1\}$ , and finally let us number each element of  $2^\Omega$  from 1 to  $L$ . This means that  $2^\Omega$  is equivalent to a set  $\{1, \dots, L\}$ , with  $L = 2^{|\Omega|}$ . Let  $T^N = (X^N, U^{N,1}, U^{N,2}, Y^N)$ , be a homogeneous HESMC verifying (17), with  $p(y_n | x_n)$  Gaussian. Let  $\mu_j$  and  $\sigma_j^2$  denote mean and variance of  $p(y_n | x_n = j)$ . Thus parameters to be estimated are, for  $i, j = 1, \dots, L$ ;  $k, l = 1, \dots, K$ ,  $q, r = 0, \dots, D - 1$ :  $\pi_{lr} = p(u_1^l = l, u_1^r = r)$ ,  $p_{klqr} = p(u_2^l = l, u_2^r = r | u_1^k = k, u_1^q = q)$ ,  $\mu_j$ , and  $\sigma_j^2$ . Let

us notice that we must ensure that  $p(t_{n+1}|t_n) = p(u_{n+1}^1, u_{n+1}^2|u_n^1, u_n^2)p(x_{n+1}|u_{n+1}^1)$ , with  $p(x_n|u_n^1) = \frac{1_{[x_n \in u_n^1]}}{|u_n^1|}$  throughout the EM algorithm, for our model to stay a HESMC. Let us note

$$\begin{aligned} p_{jlr,n} &= p(x_1 = j, u_1^1 = l, u_1^2 = r|y^N), \\ p_{ijklqr,n,n+1} &= p(x_{n+1} = j, u_{n+1}^1 = l, u_{n+1}^2 = r, x_n = i, u_n^1 = k, u_n^2 = q|y^N) \\ p_{ijklqr,n} &= p(x_{n+1} = j, u_{n+1}^1 = l, u_{n+1}^2 = r|x_n = i, u_n^1 = k, u_n^2 = q, y^N) \end{aligned}$$

Let us specify that  $p_{ijklqr,n,n+1}$  may be computed, for  $n = 1, \dots, N - 1$  with we set  $z_n = (x_n, u_n^1, u_n^2)$ :

$$p(x_{n+1}, u_{n+1}^1, u_{n+1}^2, x_n, u_n^1, u_n^2|y^N) = p(z_n, z_{n+1}|y^N) = \frac{\alpha_n(z_n)p(z_{n+1}|z_n)p(y_{n+1}|x_{n+1})\beta_{n+1}(z_{n+1})}{\sum_{z_n \in \Omega \times \Lambda} \alpha_n(z_n)p(z_{n+1}|z_n)p(y_{n+1}|x_{n+1})\beta_{n+1}(z_{n+1})} \quad (18)$$

and  $p_{jlr,n}$  can be computed with (7).

Let  $\theta$  be the vector of all parameters. EM is an iterative method producing a sequence  $\theta^0, \theta^1, \dots, \theta^u, \dots$ . After having determined the initial value  $\theta^0$  in some way,  $\theta^{s+1}$  is determined from  $\theta^s$  and  $Y^N = y^N$  as follows:

$$\pi_{lr}^{(s+1)} = \frac{1}{N} \sum_j p_{jlr,1}^{(s)} + \dots + p_{jlr,N}^{(s)}; \quad (19)$$

$$p_{klqr}^{(s+1)} = \frac{\sum_i \sum_j [p_{ijklqr,1,2}^{(s)} + \dots + p_{ijklqr,N-1,N}^{(s)}]}{\sum_j [p_{jlr,1}^{(s)} + \dots + p_{jlr,N}^{(s)}]}; \quad (20)$$

$$\mu_j^{(s+1)} = \frac{\sum_{n=1}^N \sum_r \sum_{l=1}^L p_{jlr,n}^{(s)} y_n}{\sum_{n=1}^N \sum_r \sum_{l=1}^L p_{jlr,n}^{(s)}}; \quad (21)$$

$$\sigma_j^{2,(s+1)} = \frac{\sum_{n=1}^N \sum_r \sum_{l=1}^L p_{jlr,n}^{(s)} (y_n - \mu_j^{(s+1)})^2}{\sum_{n=1}^N \sum_r \sum_{l=1}^L p_{jlr,n}^{(s)}}. \quad (22)$$

Thus, at each iteration  $\theta^s$  is used to compute  $p_{klqr}^{(s)}$  with (7) and  $p_{ijklqr,n,n+1}^{(s)}$  with (18), then (19)-(22) are used to find  $\theta^{s+1}$ .

To initialize the EM, we estimate  $p(x_{n+1} = j, x_n = i)$  from  $Y^N = y^N$  using a K-means method, then we set  $(u_2^1 = \{j\}, u_1^1 = \{i\}) = (1 - 0.25) \times p(x_2 = j, x_1 = i)$ ,  $p(u_2^1 = \{j\}, u_1^1 = \Omega) = p(u_2^1 = \Omega, u_1^1 = \{i\}) = p(u_2 = \Omega, u_1 = \Omega) = 0.25 / (2 \times |\Omega| + 1)$ , then  $p(u_2^1 = l|u_1^1 = k, u_2^1 = 0) = p(u_2^1 = l|u_1^1 = k)$ . Finally, we set  $p(u_2^2|u_1^1 = k)$  and  $p(u_2^2|u_1^1 = k, u_2^1 = l, u_1^2 = 0)$  uniform on  $\{0, 1, \dots, D - 1\}$ .

## 4. Experiments

### 4.1 Synthetic image segmentations

In this section, we propose a segmentation study to answer two questions: could HESMC improve simultaneously both HSMC and HEMC, how HESMC compare to the NSHSMC from [43]. We applied the four models mentioned before plus a HMC for reference to the unsupervised segmentation of three hand-drawn noisy images.

To use mono-dimensional chains considered in the paper, we convert the bi-dimensional set of pixels to a mono-dimensional sequence using Hilbert–Peano scan, which gave interesting results in image segmentation in similar situations [43], [30], [31]. For the bba  $M$  of the HEMC and HESMC, we choose  $M$  to be null outside of the singletons and  $\Omega$ . We estimate the parameters by the EM algorithm from the previous section. For HSMC, HESMC and NSHSMC the maximum of minimal sojourn time chosen is  $D = 10$ . For NSHSMC,  $U^{N,2}$  is a two-value random chain. On the choice of the images, we can note several things. The first image whose segmentations are presented in Figure 6, contain large size homogenous areas, which should favor the HSMC. However, as HMC performs already very well on such image, we decided to noise it with a very high independent gaussian noise  $N(0,1)$  and  $N(0.3,1)$ , to increase the influence of the prior law in the segmentation. The second image, whose segmentations are presented in Figure 7, contains very fine details which are quite challenging to segment for algorithms such as HMC, but we know from [42] that HEMC can improve HMC quite significantly in this case, provided that the noise is not too strong. Therefore, we noised it with a relatively low independent gaussian noise  $N(0,1)$  and  $N(2,1)$ . The third image, whose segmentations are presented in Figure 8, is a three-class image which was designed to combine both fine details and large size homogenous areas. We noised it with an original non-stationary noise obtained in the following way. For each  $n = 1, \dots, N$ , we first sample from uniform law on intervals  $[a, b]$  of length 2 centered on a value related to each class. We use the values obtained as the means of the Gaussian law, with

variance 1, for noising the pixel in question. We will denote such noises - which are not Gaussian - with  $N(U[a, b], 1)$ . This noise was designed to increase the influence of the prior law, as it is not gaussian (and the five models assume a gaussian law for the noise), while not completely losing the fine details with a noise whose means would be too close. Estimated noise parameters and error ratio for the three segmentations in Figure 6, Figure 7 and Figure 8 are presented in TABLE I, TABLE II and TABLE III respectively. Additionally, the masses estimated by the EM algorithm for HEMC and HESMC, for the third experiment, namely the segmentation in Figure 8, are presented in TABLE IV. The masses for the HEMC are estimated using a simplified version of the EM method presented in section 3, replacing  $p_{klqr} = p(u_2^1 = l, u_2^2 = r | u_1^1 = k, u_1^2 = q)$  with  $m_{kl} = m(u_2 = l, |u_1 = k)$  [42]. The masses for HESMC are obtained from (20) with  $m(u_2 = l, |u_1 = k) = \sum_{(q,r)} p_{klqr}$ , moreover these masses are not Markovian but rather semi-Markovian. The presented results and some other ones non reported allow the following general conclusions:

(i) HSMCs are of interest when images have large size homogenous areas and very strong noise such as in Figure 6, where they outperform HMC quite significantly. We can explain it partially because the prior law of a HSMC seems more adapted to large size homogenous areas than the one from HMC, which assumes that sojourn time in a given class obeys a geometrical distribution. We can also note that the HESMC performs very close to HSMC in this case.

(ii) HEMCs are of interest when the class image is not very homogeneous and simultaneously contains areas of large size and small details such as in Figure 7. We already know this from [42] but the key point here is that HESMC performs as well as HEMC in this case. Moreover, we can see that NSHSMC, while also designed to manage non-stationarity, is unable to restore very fine details in this case. We can infer that the combination of areas of large size and really small details in image segmentation is better managed with models that are built for the second kind of non-stationarity (nonfinite, continuous non-stationarity) from section 2.1, than models that are built for the first one (switching non-stationarity).

(iii) HESMC can simultaneously improve HSMC and HEMC by up to 40%. This seems to be the case on images containing large size homogenous areas and fine details with a noise with continuous non-stationarity such as  $N(U[a, b], 1)$  (Figure 8). We suppose that it is because this kind of images allows HESMC to use both its semi-Markov and evidential part.

(iv) HESMC can improve NSHSMC segmentation by up to 10% in the case of on images containing simultaneously large size homogenous areas and fine details with a noise with continuous non-stationarity such as  $N(U[a, b], 1)$  (Figure 8). If the noise is stationary however, the performance gap is reduced between the two models and sometimes NSHSMC can even perform better than HESMC. This is somewhat consistent with the theory as NSHSMC combine the semi-Markovianity with a switching triplet Markov chain, better suited for the first kind of non-stationarity (switching non-stationarity) presented in section 2.1, while HESMC combine the semi-Markovianity with an evidential triplet Markov chain, better suited for the second kind of non-stationarity (nonfinite, continuous non-stationarity).

(v) Parameters estimation by EM method turns out to be of an extraordinary efficiency. As mentioned above, to differentiate the four models, we consider very high-level noises on purpose, and according to TABLE I, TABLE II and TABLE III, parameters estimation remains quite correct. It is even more surprising considering that the data is not likely to follow of the four models; indeed, the pixel chain obtained by using Hilbert-Peano scan have a very complex structure. In addition, in the segmentation from Figure 3, all models consider the noise Gaussian, when it is not. On the estimation of HESMC specifically, we can see that when HSMC only seems to be of interest, parameter estimation can “reduce” the HESMC to a HSMC, while when HEMC only seems to be of interest, parameter estimation can “reduce” the HESMC to a HEMC. In the third experiment (TABLE III, Figure 8) we can see that when HSMC and HEMC improve HMC, HESMC is able to still improve both of them. Those results are close to the theoretical fact that HESMC is more general than both HSMC and HEMC, which confirms that EM is efficient enough for the greater generality of HESMC to play its role.

Finally, we can discuss a little about the computational complexity of the models presented here. If we look at MPM restoration presented in section 2, all the models have a time complexity of the same form than in an HMC, i.e.,  $O(S^2T)$  (in space  $O(ST)$ ), with  $S$  the number of hidden states (including  $X^N$  and all the  $U^N$  considered), and  $T$  the length of the sequence. For the parameter estimation by EM, the time complexity for one iteration of the algorithm is also  $O(S^2T)$  (in space  $O(ST)$ ). The key change between HMC and our models (HESMC in particular) is that in the latter the number of hidden states is much larger. Fortunately, the operations for each hidden state can be run in parallel, which makes the actual time of execution of our algorithms competitive with classical HMC if this property is exploited in the code.

	Estimates			Error ratio
		Class 1	Class 2	
True	M, V	0, 1	0.3, 1	
HMC	M, V	0.17, 0.95	0.35, 0.98	0.42
HSMC	M, V	0.03, 0.99	0.28, 1.01	0.20
HEMC	M, V	0.17, 0.96	0.27, 1.00	0.32
<b>HESMC</b>	<b>M, V</b>	<b>0.06, 0.98</b>	<b>0.26, 1.00</b>	<b>0.21</b>
NSHSMC	M, V	0.03, 1.01	0.28, 1.01	0.21

TABLE I Estimated parameters and error ratio. M: mean, V: variance, for the segmentation presented in Figure 1.

	Estimates			Error ratio
		Class 1	Class 2	
True	M, V	0, 1	2, 1	
HMC	M, V	0.99, 2.02	1.99, 1.01	0.23
HSMC	M, V	0.99, 2.02	1.99, 1.02	0.23
HEMC	M, V	-0.01, 1.01	2.00, 1.00	0.08
<b>HESMC</b>	<b>M, V</b>	<b>0.09, 1.14</b>	<b>2.00, 1.02</b>	<b>0.08</b>
NSHSMC	M, V	0.99, 2.02	1.99, 1.02	0.23

TABLE II Estimated parameters and error ratio. M: mean, V: variance, for the segmentation presented in Figure 2.

	Estimates				Error ratio
		Class 1	Class 2	Class 3	
True	M, V	0, 1.33	2, 1.33	4, 1.33	
HMC	M, V	-0.19, 1.26	0.05, 1.38	2.55, 2.65	0.29
HSMC	M, V	0.01, 1.38	1.92, 1.48	2.73, 3.62	0.16
HEMC	M, V	-0.52, 1.11	0.20, 1.30	2.59, 2.54	0.17
<b>HESMC</b>	<b>M, V</b>	<b>-0.02, 1.34</b>	<b>2.02, 1.58</b>	<b>4.12, 1.21</b>	<b>0.09</b>
NSHSMC	M, V	-0.01, 1.33	1.96, 1.40	3.74, 1.69	0.10

TABLE III Estimated parameters and error ratio. M: mean, V: variance, for the segmentation presented in Figure 3.

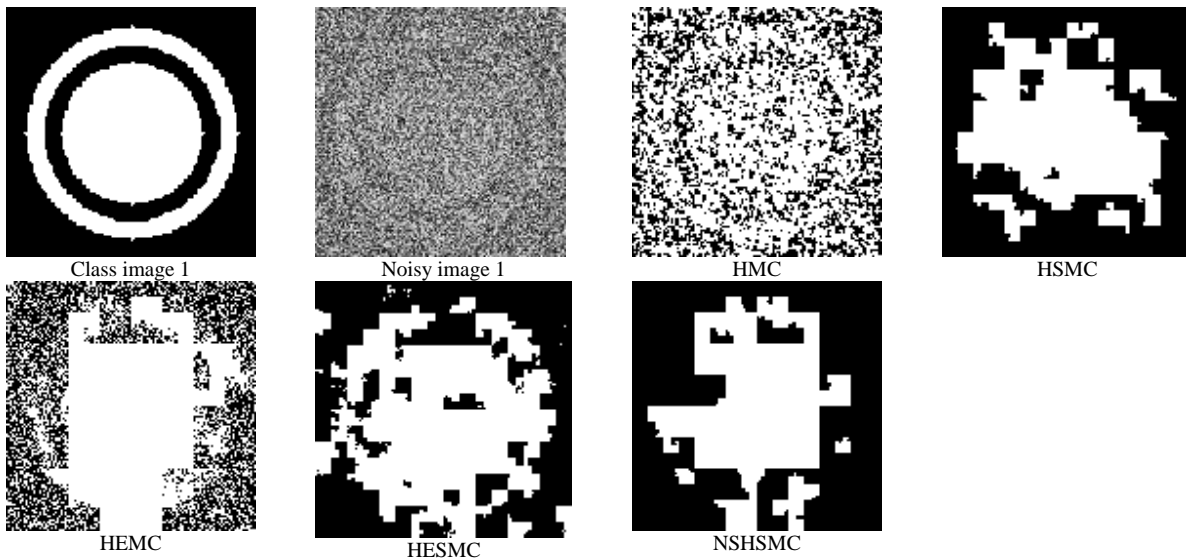


Figure 6. Segmentation of two classes hand-drawn image (Class image 1) corrupted by non-stationary noise  $N(0,1)$ ,  $N(0.3,1)$  (Noisy image 1). HMC, HSMC, HEMC, HESMC and NSHSMC: unsupervised segmentation results with the related models.

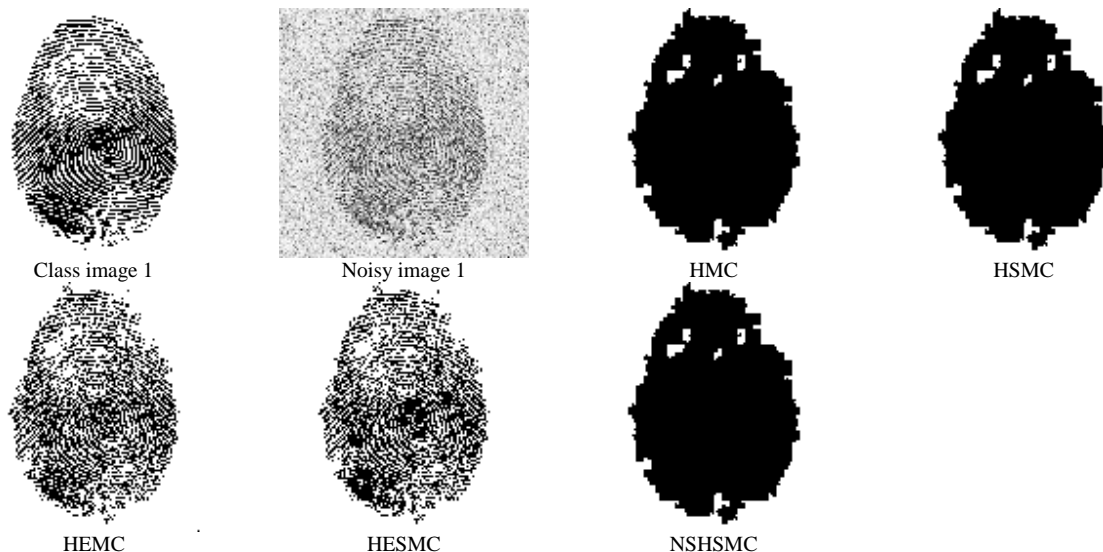


Figure 7. Segmentation of two classes hand-drawn image (Class image 1) corrupted by non-stationary noise  $N(0,1)$ ,  $N(2,1)$  (Noisy image 1). HMC, HSMC, HEMC, HESMC and NSHSMC: unsupervised segmentation results with the related models.

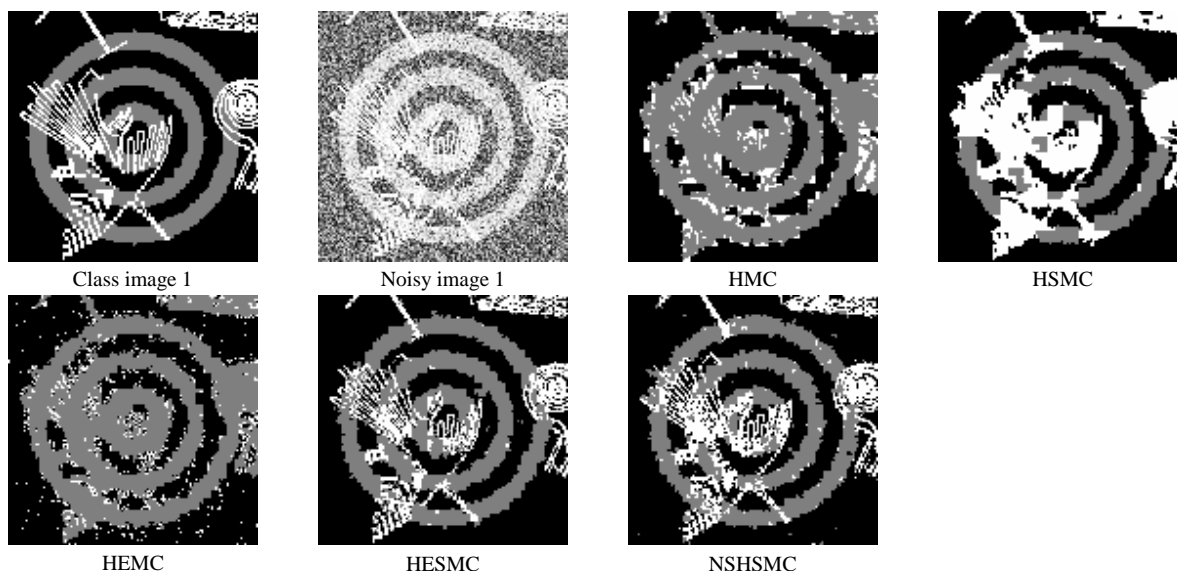


Figure 8. Segmentation of three classes hand-drawn image (Class image 1) corrupted by non-stationary noise  $N(U([-1,1]),1)$ ,  $N(U([1,3]),1)$ , and  $N(U([3,5]),1)$  (Noisy image 1). HMC, HSMC, HEMC, HESMC and NSHSMC: unsupervised segmentation results with the related models.

#### 4.2 Real image segmentations

In this section, we consider one real gray scale SAR image, and we segmented it by HMC, HEMC, HSMC, HESMC and NSHSMC. The settings of the experiment are the same than in the previous paragraph, thus the image is converted to a chain using the Peano scan, the parameters are estimated by the EM algorithm and the models considered are exactly the same. The results obtained are presented in Figure. 9. As we have no ground truth it is difficult to draw general conclusions. However, both HESMC and NSHSMC based segmentations seem clearly of better quality than the HMC, HEMC and HSMC based ones, mainly because some details seem better preserved by the first two models, as demonstrated by the parts of the images framed in yellow. Thus, both HESMC and NSHSMC can be of interest when replacing HMC, HSMC or HEMC. If we compare HESMC and NSHSMC, the difference is less striking, but NSHSMC segmentation seems a bit less clean for the “white” class. Finally, we are in an interesting case where both HEMC and HSMC don’t seem of interest compared to HMC, but HESMC is still able to “improve” the segmentation of HMC.

HEMC	$m(u_n, u_{n+1})$	{1}	{2}	{3}	{1,2,3}	
		{1}	0.28	0	0.06	0.01
		{2}	0	0.36	0	0.02
		{3}	0.06	0	0.04	0
		{1,2,3}	0.01	0.02	0	0.14
HESMC	$m(u_n^1, u_{n+1}^1)$	{1}	{2}	{3}	{1,2,3}	
		{1}	0.24	0.01	0	0
		{2}	0.01	0.51	0	0.01
		{3}	0	0	0.05	0.02
		{1,2,3}	0	0	0.02	0.13

TABLE IV Masses estimated by the EM algorithm, for HEMC and HESMC, for the third segmentation presented in Figure 8.

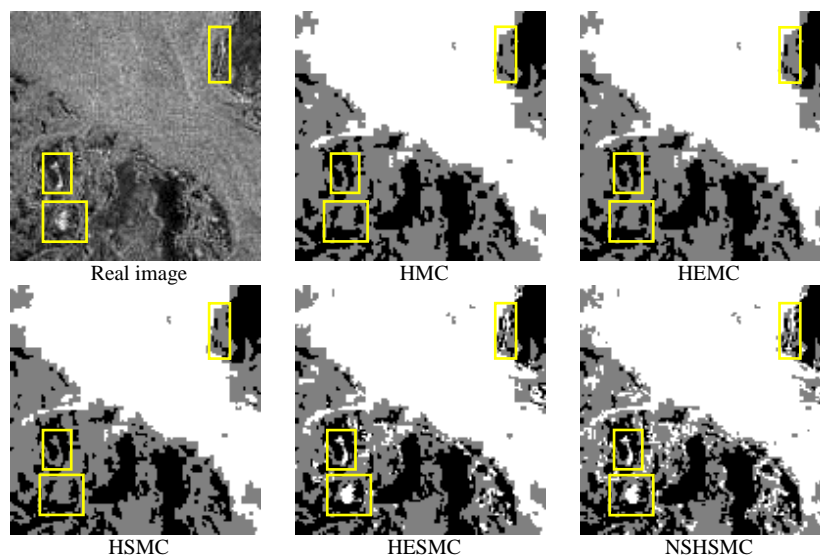


Figure 9. Segmentation of a real gray scale SAR image (Real image). HMC, HSMC, HEMC, HESMC and NSHSMC: unsupervised segmentation results with the related models.

## 5. Conclusion

We dealt in this paper with unsupervised segmentation of complex non-stationary data. The tool used was a new triplet Markov chain model called “hidden evidential semi-Markov chain” (HESMC). We obtain it by the introduction of two auxiliary chains: the first one is based on theory of evidence to manage the lack of stationarity [41], while the second one uses semi-Markovianity, to manage the sojourn time in a given class with more flexibility than Markov models do [13]. When unsupervised image segmentation is concerned, HESMC may be of strong interest on images containing simultaneously large homogenous areas, fine details, and non-stationary noise, where it can significantly improve hidden evidential Markov chain (HEMC [13]), hidden semi-Markov chain (HSMC [43]), and “non-stationary hidden semi-Markov chain” (NSHSMC [43]). Furthermore, as shown in the experiments, HESMC can perform almost as well as HSMC when the latter is of interest, and we can say the same for HEMC. Considering that the actual time of execution of MPM and parameter estimation for HESMC is competitive with HSMC and HEMC if the operations for each hidden state are run in parallel, this model can allow someone not to choose between HSMC and HEMC modelization by using a model that include both.

Let us notice that if data are well suited to HESMC, its use must give better or equivalent results compared to those obtained with HEMC or HSMC, simply because it extends them. However, this is no longer necessarily true in unsupervised framework where results also depend on the effectiveness of parameter estimation. Of course, this is even truer when data match no one of the models HESMC, HEMC or HSMC, as considered in experiments. Thus, it is of importance to notice that in spite of the complexity of HESMC, the EM considered is efficient enough to improve unsupervised HEMC or HSMC based processing, at least in considered experiments framework. Finally, the generality of the HESMC and the efficiency of its parameter estimation by the EM algorithm imply that this model could be interesting in a lot of real-world applications. Indeed, as HESMC is an extension of both



HEMC and HSMC, which are in turns extensions of HMC. Therefore HESMC could replace HMC in any application it is considered. HMC's real-world applications are very popular and continue to be widely used. The most interesting conclusion of the article is that the new HESMC is likely to improve the results of any of these applications.

As perspective for designing more complex models and related unsupervised processing, let us note the possibility to extend HESMC by adding others random sequences in a way that the model remains a TMC. Another perspective for extension is the simplicity of the considered modelization of the noise in the HESMC presented. It is Gaussian, and its link with the hidden data is very simple, while these conditions are not necessary ones for the MPM works. Considering further extensions of different models considered is realistic because of the excellent behavior of the EM noticed in the presented studies.

## References

- [1] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains", *Annals of Mathematical Statistics*, vol. 41, no. 1, pp. 164-171, February 1970.
- [2] O. Cappé, E. Moulines, and T. Ryden, "Inference in Hidden Markov Models", Springer, Series in Statistics, 2005.
- [3] Y. Ephraim, "Hidden Markov processes", *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp.1518–1569, June 2002.
- [4] T. Koski, "Hidden Markov Models for Bioinformatics", Kluwer Academic Publishers, 2001.
- [5] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257- 286, Feb. 1989.
- [6] K. Aas, L. Eikvil, and R.B. Huseby, "Applications of hidden Markov chains in image analysis", *Pattern Recognition*, vol. 32, no. 4, pp. 703-713, 1999.
- [7] M.-Y. Chen, A. Kundu, and J. Zhou, "Off-Line handwritten word recognition using a hidden Markov model type stochastic network," *IEEE Trans. Pattern Analysis and Machine Intelligence*", vol. 16, no. 5, pp. 481- 496, May 1994.
- [8] G.A. Churchill, "Hidden Markov chains and the analysis of genome structure", *Computers and Chemistry*, vol. 16, no. 2, pp. 107-115, 1992.
- [9] J.F. Zaki, A. Ali-Eldin, S.E. Hussein, S.F. Saraya, F.F. Areed, Traffic congestion prediction based on Hidden Markov Models and contrast measure, *Ain Shams Eng J*, 2020.
- [10] D. Khiatani and U. Ghose, "Weather forecasting using Hidden Markov Model", 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), pp. 220-225, 2017.
- [11] Nystrup, P., Madsen, H., & Lindström, E. Stylised facts of financial time series and hidden Markov models in continuous time. *Quantitative Finance*, 15(9), 1531-1541, 2015.
- [12] Kayte, Sangramsing N. & Mal, Monica & Gujrathi, Jayesh. Hidden Markov Model based Speech Synthesis: A Review. *International Journal of Computer Applications*. 130. 975-8887, 2015.
- [13] V. S. Barbu, N. Limnios, "Semi-Markov Chains and hidden semi-Markov models toward applications: Their use in reliability and DNA analysis" Springer Science & Business Media, New York, USA, vol. 191, 2009.
- [14] S. Z. Yu and H. Kobayashi, "An efficient forward-backward algorithm for an explicit-duration hidden Markov model", *IEEE Signal Processing Letters*, vol. 10, no. 1, pp. 11-14, 2003.
- [15] Y. Guédon, Exploring the state sequence space for hidden Markov and semi-Markov chains, *Computational Statistics & Data Analysis*, Volume 51, Issue 5, 2007.
- [16] S. Z. Yu, "Hidden Semi-Markov Models: Theory, Algorithms and Applications", Morgan Kaufmann: Amsterdam, The Netherlands, 2015.
- [17] Y. Ephraim, B. L. Mark, "Causal recursive parameter estimation for discrete-time hidden bivariate Markov chains", *IEEE Trans. on Signal Processing*, Vol. 63, No. 8, pp. 2108-2117, August 2015.
- [18] Y. Sun, B. L. Mark, and Y. Ephraim, "Collaborative spectrum sensing via online estimation of hidden bivariate Markov models", *IEEE Trans. on Wireless Communications*, vol. 15, no. 8, pp. 5430-5439, 2016.
- [19] M. Fitzpatrick and D. Marchev, "Efficient Bayesian estimation of the multivariate Double Chain Markov Model", *Statistics and Computing*, vol. 23, no. 4, pp. 467-480, 2013.
- [20] M. Ameer, N. Idrissi, and C. Daoudi, "Color images segmentation using Pairwise Markov Chain", *International Conference on Multimedia Computing and Systems*, Rabat, Marocco, May 10-12, 2018.
- [21] A. K. Atiampo and G. L. Loum, "Unsupervised image segmentation with Pairwise Markov Chains based on nonparametric estimation of copula using orthogonal polynomials", *Journal of Image and Graphics*, vol. 16, no. 4, pp. 2526-2541, 2016.
- [22] S. Le Cam, F. Salzenstein, and C. Collet, "Fuzzy pairwise Markov chain to segment correlated noisy data", *Signal Processing*, vol. 88, no. 10, pp. 2526-2541, 2008.
- [23] S. Derrode, W. Pieczynski, Unsupervised data classification using pairwise Markov chains with automatic copulas selection, *Volume 63*, 2013.
- [24] M. Ameer, N. Idrissi, and C. Daoudi, "Triplet Markov chain in images segmentation", *International Conference on Intelligent Systems and Computer Vision (ISCV)*, Fez, Marocco April 02-04, 2018.
- [25] S. Brick, C. Collet, J-P. Armspach, "Triplet Markov chain for 3D MRI brain segmentation using a probabilistic atlas", *3rd IEEE ISBI*, Arlington, United States, April 06-09, 2006.
- [26] H. Gangloff, K. Morales, Y. Petetin, Deep parameterizations of pairwise and triplet Markov models for unsupervised classification of sequential data, *Computational Statistics & Data Analysis*, Volume 180, 2023.
- [27] A. Ben Abbes, I. R. Farah, and V. Barra, "A non-stationary NDVI time series modelling using triplet Markov chain", *International Journal of Information and Decision Sciences*, vol. 11, no. 2, pp. 163-179, 2019.
- [28] H. Li, S. Derrode, and W. Pieczynski, "Adaptive on-line lower limb locomotion activity recognition of healthy individuals using semi-Markov model and single wearable inertial sensor", *Sensors*, 19, 4242, 2019.

- [29] S. Chen and X. Jiang, "Modeling repayment behavior of consumer loan in portfolio across business cycle: a triplet Markov model approach", *Complexity*, Article ID 5458941, 2020.
- [30] C. Carincotte, S. Derrode, and S. Bourennane, "Unsupervised change detection on SAR images using fuzzy hidden Markov chains", *IEEE Trans. on Geosciences and Remote Sensing*, vol. 44, no. 2, pp. 432-441, 2006.
- [31] F. Salzenstein, C. Collet, S. Le Cam, and M. Hatt, "Non-stationary fuzzy Markov chain", *Pattern Recognition*, vol. 28, no. 16, pp. 2201-2208, 2007.
- [32] S. Le Cam, C. Collet, and F. Salzenstein, "Acoustical respiratory signal analysis and phase detection", *ICASSP 08 Proc.*, pp. 3629-3632, 2008.
- [33] T. Denoeux, 40 years of Dempster-Shafer theory, *International Journal of Approximate Reasoning*, vol. 79, pp. 1-6, 2016.
- [34] J. Guan and D.A. Bell, "Evidence theory and its applications", North-Holland, Amsterdam, 1991.
- [35] G. Shafer, "A mathematical theory of evidence", Princeton University Press, Princeton, 1976.
- [36] P. Smets, "Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem", *International Journal of Approximate Reasoning*, vol. 9, pp. 1-35, 1993.
- [37] P. Smets and R. Kennes, "The transferable belief model", *Artificial Intelligence*, vol. 66, pp. 191-234, 1994.
- [38] E. Ramasso and T. Denoeux, "Making use of partial knowledge about hidden states in HMMs: an approach based on belief functions", *IEEE Trans. on Fuzzy Systems*, vol. 22, no. 2, pp. 1102-1114, 2017.
- [39] E. Ramasso, "Inference and learning in evidential discrete latent Markov models", *IEEE Trans. on Fuzzy Systems*, vol. 25, no. 5, pp. 395-405, 2014.
- [40] W. Pieczynski, Multisensor triplet Markov chains and theory of evidence, *International Journal of Approximate Reasoning*, Vol. 45, No. 1, pp. 1-16, 2007.
- [41] M. Y. Boudaren and W. Pieczynski, "Dempster-Shafer fusion of evidential pairwise Markov chains", *IEEE Trans. on Fuzzy Systems*, Vol. 24, No. 6, pp. 1598-1610, December 2016.
- [42] P. Lanchantin and W. Pieczynski, Unsupervised restoration of hidden non stationary Markov chain using evidential priors, *IEEE Trans. on Signal Processing*, vol. 53, no. 8, pp. 3091-3098, 2005.
- [43] J. Lapuyade-Lahorgue and W. Pieczynski, "Unsupervised segmentation of hidden semi-Markov non stationary chains", *Signal Processing*, vol. 92, no. 1, pp. 29-42, January 2012.
- [44] M. Y. Boudaren and W. Pieczynski, Unified representation of sets of heterogeneous Markov transition matrices, *IEEE Trans. on Fuzzy Systems*, Vol. 24, No. 2, pp. 497-503, April 2016.
- [45] G. J. McLachlan and T. Krishnan, "EM algorithm and extensions", Wiley, Series in Probability and Statistics, 1997.
- [46] Brian G. Leroux, Maximum-likelihood estimation for hidden Markov models, *Stochastic Processes and their Applications*, Volume 40, Issue 1, 1992, Pages 127-143.
- [47] M. Y. Boudaren, E. Monfrini, and W. Pieczynski, Unsupervised segmentation of random discrete data hidden with switching noise Distributions, *IEEE Signal Processing Letters*, Vol. 19, No. 10, pp. 619-622, October 2012.
- [48] W. Pieczynski, "Modeling non stationary hidden semi-Markov chains with triplet Markov chains and theory of evidence", *SSP2005*, Bordeaux, France, July 17-20, 2005.