Available online at www.sciencedirect.com
ScienceDirect
Signal Processing 88 (2008) 1134–1151
SIGNAL PROCESSING
www.elsevier.com/locate/sigpro

Unsupervised segmentation of triplet Markov chains hidden with long-memory noise

Pierre Lanchantin, Jérôme Lapuyade-Lahorgue, Wojciech Pieczynski*

GET/INT, CITI Department, CNRS UMR 5157, 9, rue Charles Fourier, 91000 Evry, France

Received 16 November 2006; received in revised form 1 August 2007; accepted 21 October 2007
Available online 20 November 2007

Abstract

The hidden Markov chain (HMC) model is a couple of random sequences (X,Y), in which X is an unobservable Markov chain, and Y is its observable noisy version. Classically, the distribution $p(y|x)$ is simple enough to ensure the Markovianity of $p(x|y)$, that enables one to use different Bayesian restoration techniques. HMC model has recently been extended to "pairwise Markov chain" (PMC) model, in which one directly assumes the Markovianity of the pair $Z=(X,Y)$, and which still enables one to recover X from Y. Finally, PMC has been extended to "triplet Markov chain" (TMC) model, which is obtained by adding a third chain U and considering the Markovianity of the triplet $T=(X,U,Y)$. When U is not too complex, X can still be recovered from Y. Then U can model different situations, like non-stationarity or semi-Markovianity of (X,Y). Otherwise, PMC and TMC have been extended to pairwise "partially" Markov chains (PPMC) and triplet "partially" Markov chains (TPMC), respectively. In a PPMC $Z=(X,Y)$ the distribution $p(x|y)$ is a Markov distribution, but $p(y|x)$ may not be and, similarly, in a TPMC $T=(X,\ U,\ Y)$ the distribution $p(x,u|y)$ is a Markov distribution, but $p(y|x,u)$ may not be. However, both PPMC and TPMC can enable one to recover X from Y, and TPMC include different long-memory noises. The aim of this paper is to show how a particular Gaussian TPMC can be used to segment a discrete signal hidden with long-memory noise. An original parameter estimation method, based on "Iterative Conditional Estimation" (ICE) principle, is proposed and some experiments concerned with unsupervised segmentation are provided. The particular unsupervised segmentation method used in experiments can also be seen as identification of different stationarities in fractional Brownian noise, which is widely used in different problems in telecommunications, economics, finance, or hydrology.

Keywords: Hidden Markov chains; Pairwise Markov chains; Triplet Markov chains; Long-memory noise; Iterative conditional estimation; Unsupervised Bayesian segmentation

1. Introduction

Let $X=(X_n)_{1\leqslant n\leqslant N}$ and $Y=(Y_n)_{1\leqslant n\leqslant N}$ be two stochastic processes, where X is hidden and Y is observable. Each X_n takes its values in a finite set of classes $\Omega=\{\omega_1,\ldots,\omega_K\}$ and each Y_n takes its values in R. The problem of estimating X from Y, which occurs in numerous applications, can be solved with

*Corresponding author. Tel.: +33 1 60 76 44 25; fax: +33 1 60 76 44 33.
E-mail address: wojciech.pieczynski@int-evry.fr (W. Pieczynski).

0165-1684/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2007.10.015

Bayesian methods once one has chosen some accurate distribution $p(x,y)$ for $Z=(X,Y)$. The hidden Markov chain (HMC) model is the simplest and most well-known model [1–3]. Its applications cover numerous fields, such as acoustics [4], biosciences [3,5], climatology [6], ecology [7,8], control [9], communications [10,11], econometrics and finance [12,13], handwriting and text recognition [34], image processing and computer vision [14–18], signal processing [1], etc. We only cite one or two recent publications on each subject, each of them containing a rich bibliography. Moreover, a rich bibliography on classical HMC can also be found in [1–3]. This model has been extended to pairwise Markov chains model (PMC [19,38]) and then to triplet Markov chains model (TMC [20–23]). The PMC and TMC models, on the other hand, have then been extended to pairwise partially Markov chains (PPMC) and triplet partially Markov chains (TPMC) [24,25]. The interest of PPMC is that the distribution of the noise—in other words, the distribution $p(y|x)$ of Y conditionally on X—is not necessarily a Markov chain. One possible application, addressed in this paper, is to consider a "long-memory" noise, also called "long-range dependence noise", which occurs in numerous situations [26–29] and which cannot be taken into account neither via classical HMC, nor recent PMC.

Finally, developing some first ideas proposed in [30], the aim of this paper is to study the possibilities of defining models including discrete random chain hidden with long-memory noise, and the interest of using them in unsupervised hidden discrete signal segmentation. We propose a new parameter estimation method based on the general "Iterative Conditional Estimation" (ICE, already successfully used in [16,17,19,31–33]) principle, and show how the Bayesian maximum of the posterior marginals (MPM) method based on ICE estimated parameters can improve the classical HMC- and ICE-based results.

Let us notice that the particular model proposed in Section 3.3 can be interpreted as a switching model of some phenomenon modelled with long-memory process. The switching process is modelled as a Markov chain, and thus one directly arrives at a "Markov chains hidden with a long-memory noise", which is a generalization of the classical hidden Markov chains, without any reference to PMC or TMC. In particular, long-memory processes found numerous applications into telecommunications, economics, hydrology, or turbulence (see Part III, pp. 369–524 in [29]). Therefore using the proposed model makes possible to find unknown discrete states modelling different homogeneities of observed sequences.

The paper is organized as follows. Section 2 is devoted to the classical hidden Markov chains, pairwise and triplet Markov chains, and Bayesian segmentation. Partially PMC and TMC are recalled and the new models are introduced in Section 3. The parameter estimation problem is addressed in Section 4, where a new estimation method well-suited to Gaussian cases is proposed. Different experiments showing the interest of the new models and the related unsupervised processing are presented in Section 5, while the last Sixth section is devoted to conclusions and perspectives.

2. Hidden Markov chains and extensions

2.1. Hidden Markov chains and Bayesian segmentation

Let $X=(X_n)_{1\leqslant n\leqslant N}$ and $Y=(Y_n)_{1\leqslant n\leqslant N}$ be two stochastic processes as specified above; the problem is to estimate $X=x$ from $Y=y$. In the whole paper, we will use the notations $x=(x_1,\ldots,x_N)$ and $y=(y_1,\ldots,y_N)$; however, we will also possibly write $x_1^N=(x_1,\ldots,x_N)$ and $y_1^N=(y_1,\ldots,y_N)$ to re-emphasize the starting and the ending indices. More generally, we will set $x_i^j=(x_i,\ldots,x_j)$ (for $1\leqslant i<j\leqslant N$), and the same for y_i^j. Considering the classical HMC with independent noise (HMC-IN) consists of considering the distribution $p(x,y)$ of (X,Y) of the form

$$p(x,y)=p(x_1)p(y_1|x_1)p(x_2|x_1)\times p(y_2|x_2)\ldots p(x_N|x_{N-1})p(y_N|x_N). \quad (2.1)$$

The hidden chain $X=(X_n)_{1\leqslant n\leqslant N}$ is then a Markov chain with the distribution $p(x)=p(x_1)p(x_2|x_1)\ldots p(x_N|x_{N-1})$, and the distribution of $Y=(Y_n)_{1\leqslant n\leqslant N}$ conditional on $X=(X_n)_{1\leqslant n\leqslant N}$, which can be considered as modeling the "noise", is given by $p(y|x)=p(y_1|x_1)p(y_2|x_2)\ldots p(y_N|x_N)$. This particular form of $p(y|x)$ is equivalent to the following two hypotheses:

(i) the random variables $(Y_n)_{1\leqslant n\leqslant N}$ are independent conditionally on X;
(ii) for each $n=1,\ldots,N$, the distribution of Y_n conditional on $X=(X_n)_{1\leqslant n\leqslant N}$ is equal to its distribution conditional on X_n: $p(y_n|x)=p(y_n|x_n)$.

Let us notice that such a model is usually said "HMC", and "IN", which means "independent noise" because of (ii), is omitted. However, we will keep "HMC-IN" for the classical model (2.1) in this paper, and we will call HMC every distribution $p(x,y)$ of (X,Y) such that the hidden chain X is a Markov one.

One of the nice properties of HMC-IN we are interested in this paper is that all the posterior marginal distributions $p(x_n|y)$ are calculable, even for very large N. One introduces the "Forward" probabilities $\alpha^n(x_n) = p(x_n,y_1,\ldots,y_n)$, and the "Backward" probabilities $\beta^n(x_n) = p(y_{n+1},\ldots,y_N|x_n)$, which are both calculated by the following classical forward and backward recursions ([1,2]):

$$\begin{aligned}&\alpha^1(x_1) = p(x_1,y_1);\\ &\alpha^{n+1}(x_{n+1}) = \sum_{x_n\in\Omega}\alpha^n(x_n)p(x_{n+1}|x_n)p(y_{n+1}|x_{n+1}),\end{aligned} \tag{2.2}$$

$$\begin{aligned}&\beta^N(x_N) = 1,\\ &\beta^n(x_n) = \sum_{x_{n+1}\in\Omega}\beta^{n+1}(x_{n+1})p(x_{n+1}|x_n)p(y_{n+1}|x_{n+1}).\end{aligned} \tag{2.3}$$

The marginal posterior distributions of the hidden state can then be calculated by ("$\propto$" means "proportional to"):

$$p(x_n|y) \propto \alpha^n(x_n)\beta^n(x_n). \tag{2.4}$$

Having $p(x_n|y)$, we can use the following Bayesian maximum posterior marginals (MPM) segmentation method

$$\begin{aligned}&\hat{s}_{\mathrm{MPM}}(y_1,\ldots,y_N) = (\hat{x}_1,\ldots,\hat{x}_N),\\ &\text{with } \hat{x}_n = \arg\max_{x_n\in\Omega} p(x_n|y),\end{aligned} \tag{2.5}$$

whose interest lies in the following Bayesian optimality property. Let $x = (x_1,\ldots,x_N)$ be the true unknown sequence, and let $x' = (x'_1,\ldots,x'_N)$ be another one. Considering x' instead of x has a cost, which is modeled by a cost function $L(x',x)$. The Bayesian estimator $\hat{s}_L$ linked with L then minimizes the mean cost: $E[L(\hat{s}_L(Y),X] = \min_{\hat{s}} E[L(\hat{s}(Y),X]$. Now, the estimator (2.5) is the Bayesian estimator defined by the cost function

$$L_1(x',x) = \frac{\delta(x'_1,x_1)+\cdots+\delta(x'_N,x_N)}{N},$$

where $\delta(x'_n,x_n) = 1$ if $x'_n = x_n$ and $\delta(x'_n,x_n) = 0$ if $x'_n \neq x_n$, which is simply the proportion of errors in the sequence $x' = (x'_1,\ldots,x'_N)$. In other words, for N large enough, estimator (2.5) minimizes the proportion of wrongly classified points. Let us also mention that HMC-IN also makes possible the use of the loss function $L_2(x',x) = \delta(x',x)$, which gives the Bayesian "maximum a posteriori" (MAP) estimator $\hat{s}(y) = \arg\max_{x\in\Omega^N} p(x|y)$, calculable by the classical Viterbi algorithm.

Finally, the HMC-IN (2.1) has been widely applied in different areas mentioned in Introduction and, in spite of the simplicity (open to criticism) of the hypotheses (i) and (ii), it generally gives satisfying results.

2.2. Pairwise and triplet Markov chains

Let $X = (X_n)_{1\leqslant n\leqslant N}$ and $Y = (Y_n)_{1\leqslant n\leqslant N}$ be two stochastic processes as above; considering a TMC consists in choosing a third stochastic process $U = (U_n)_{1\leqslant n\leqslant N}$ such that $T = (X,U,Y) = ((X_n, U_n, Y_n))_{1\leqslant n\leqslant N}$ is a Markov chain. In this paper, we assume that each U_n takes its values in a finite set $\Lambda = \{\lambda_1,\ldots,\lambda_M\}$. To simplify, let us introduce $V = (V_n)_{1\leqslant n\leqslant N} = (X_n,U_n)_{1\leqslant n\leqslant N}$. Therefore each V_n takes its values in $\Omega\times\Lambda$ and (V,Y) is a Markov chain.

Let us recall that $X = (X_n)_{1\leqslant n\leqslant N}$ is the process of interest, whose realization is hidden, and $Y = (Y_n)_{1\leqslant n\leqslant N}$ is the observed process, and thus both of them usually have precise physical meaning. For example, let us assume that the points $(1,\ldots,N)$ are pixels of a line of a digital image in which there are two classes "forest" and "water". Then each X_n takes its values in $\Omega = \{\omega_1,\omega_2\}$, where ω_1 is "forest" and ω_2 is "water". Otherwise, each Y_n takes its values in R and thus $Y = (Y_1,\ldots,Y_N) = (y_1,\ldots,y_N)$ is the observed line of the observed digital image. Then $(y_1,\ldots,y_N)$ can be seen as a "noisy" version of $(x_1,\ldots,x_N)$; however, the sense of the word "noise" is here very general. For instance, there is a natural variability of the class "forest", whose aspect—and thus the value of y_n—varies with n. Concerning the chain $U = (U_n)_{1\leqslant n\leqslant N}$ things are different and its main interest is to enrich the family of possible distributions $p(x,y)$. However, it can admit some useful precise interpretations. For example, it can be used to model different stationarities in a non-stationary distribution $p(x,y)$, as studied in [35]. Another use is to model semi-Markov chains X; in fact, the distribution $p(x)$ of a semi-Markov chain X can be defined as the marginal distribution of the distribution $p(x,u)$ of a Markov chain (X,U). Moreover, these both chains can be used

simultaneously: as described in [36], non-stationary hidden semi-Markov chains can be seen as stationary TMC (X, U, Y), with $U=(U^1, U^2)$, where U^1 models the semi-Markovianity of X, and U^2 models its different stationarities. More recently, the model proposed in [36] has been generalized to a "long-memory noise" model [37], in which the noise can be seen as a simplified version of the noise considered in this paper.

Let us notice the great generality of the family of TMCs. In fact, only $T=(X, U, Y)$ is assumed to be Markov and thus none of the six chains X, U, Y, (X, U), (U, Y), (X, Y) is necessarily Markov: see some conditions under which some of these chains are Markov in [38,39]. However, in spite of this great generality, the posterior marginal distributions $p(x_n|y)$ are still calculable in TMC and thus the Bayesian MPM estimator (2.5) is applicable.

More precisely, we introduce the "Forward" probabilities $\alpha^n(v_n)=p(v_n,y_1\ldots,y_n)$, and the "Backward" probabilities $\beta^n(v_n)=p(y_{n+1},\ldots,y_N|v_n,y_n)$, which are both calculated with

$$\alpha^1(v_1)=p(v_1,y_1),$$
$$\alpha^{n+1}(v_{n+1})=\sum_{v_n\in\Omega\times\Lambda}\alpha^n(v_n)p(v_{n+1},y_{n+1}|v_n,y_n), \quad (2.6)$$

$$\beta^N(v_N)=1,$$
$$\beta^n(v_n)=\sum_{v_{n+1}\in\Omega\times\Lambda}\beta^{n+1}(v_{n+1})p(v_{n+1},y_{n+1}|v_n,y_n). \quad (2.7)$$

The marginal posterior distributions of the hidden state can be calculated by

$$p(v_n|y)\propto\alpha^n(v_n)\beta^n(v_n) \quad (2.8)$$

which gives

$$p(x_n|y)=\sum_{u_n\in\Lambda}p(v_n|y). \quad (2.9)$$

Let us notice that $p(u_n|y)$ is also calculable with

$$p(u_n|y)=\sum_{x_n\in\Omega}p(v_n|y) \quad (2.10)$$

and can have some interesting meaning.

Finally, let us notice that $p(v_n, v_{n+1}|y)$, which will be needed in the parameter estimation method discussed in Section 4, is also calculable and is given by

$$p(v_n,v_{n+1}|y)\propto\alpha^n(v_n)p(v_{n+1},y_{n+1}|v_n,y_n)\beta^{n+1}(v_{n+1}) \quad (2.11)$$

and thus the transitions of the Markov chain $p(v|y)$ are:

$$p(v_{n+1}|v_n,y)=p(v_{n+1},y_{n+1}|v_n,y_n)\frac{\beta^{n+1}(v_{n+1})}{\beta^n(v_n)}. \quad (2.12)$$

Finally, PMCs can be seen as particular TMC, where $U=X$; formulas (2.6)–(2.11) remains then valid with v replaced by x. Of course, PMC can be introduced directly by considering that $Z=(X,Y)$ is a Markov chain.

3. Pairwise and triplet partially Markov chains

3.1. General pairwise partially Markov chains

Let $X=(X_n)_{1\leqslant n\leqslant N}$ and $Y=(Y_n)_{1\leqslant n\leqslant N}$ be two stochastic processes as specified in the Introduction.

Definition 3.1. The pairwise chain $Z=(X,Y)$ is called "pairwise partially Markov chain" (PPMC) if its distribution $p(z)$ verifies for each $n=1,\ldots,N-1$:

$$p(z_{n+1}|z^n)=p(z_{n+1}|z_n,y^{n-1}). \quad (3.1)$$

We see that "partially Markov" comes from the fact that $Z=(X,Y)$ is Markovian with respect to the variables X, but is not necessarily Markovian with respect to the variables Y.

Remark 3.1. In the classical HMC-IN considered in Section 2, we have $p(z_{n+1}|z^n)=p(x_{n+1}|x_n)p(y_{n+1}|x_{n+1})$. Writing $p(z_{n+1}|z_n,y^{n-1})=p(x_{n+1}|z_n,y^{n-1})p(y_{n+1}|z_n,x_{n+1},y^{n-1})$, we see that a PPMC $Z=(X,Y)$ is a classical HMC-IN if $p(x_{n+1}|z_n,y^{n-1})=p(x_{n+1}|x_n)$ and $p(y_{n+1}|z_n,x_{n+1},y^{n-1})=p(y_{n+1}|x_{n+1})$.

Let $Z=(X,Y)$ be a PPMC. According to (3.1) we have

$$p(z)=p(z_1)\prod_{n=1}^{N-1}p(z_{n+1}|z_n,y^{n-1}). \quad (3.2)$$

Let us show that $p(x|y)$ is a Markov chain with $p(x_1|y)$ and the transitions given by

$$p(x_1|y)=\frac{p(x_1,\,y_1)\,\beta_1(x_1)}{\sum_{x_1'}p(x_1',\,y_1)\,\beta_1(x_1')},$$
$$p(x_{n+1}|x_n,y)=\frac{p(z_{n+1}|z_n,y^{n-1})\beta_{n+1}(x_{n+1})}{\beta_n(x_n)}, \quad (3.3)$$

where $\beta^n(x_n)$ are calculated by the following "backward" recursions

$$\beta^N(x_N)=1,$$
$$\beta^n(x_n)=\sum_{x_{n+1}} p(z_{n+1}|z_n,y^{n-1})\beta^{n+1}(x_{n+1})$$
$$\text{for } n=N-1,\ldots,1. \tag{3.4}$$

Let us notice that one could show that $\beta^n(x_n)=p(y_{n+1}^N|x_n,y_1^n)$—which extends to long correlation noise the classical definition—but this is not essential here; important is that $\beta^n(x_n)$ are computable.

The proof of (3.3) and (3.4) consists of direct application of the following lemma proved in [39]:

Lemma 3.1. *Let $X=(X_1,\ldots,X_N)$ be a random chain, each X_n taking its values in the same finite set Ω. Then X is a Markov chain if and only if there exist $N-1$ positive functions $q_1,\ldots,q_{N-1}$ such that the law of X is proportional to the product $q_1(x_1,x_2)\times\cdots\times q_{N-1}(x_{N-1},x_N)$:*

$$p(x)\propto q_1(x_1,x_2)\times\cdots\times q_{N-1}(x_{N-1},x_N). \tag{3.5}$$

If (3.5) is verified, $p(x_1)$ and the transitions $p(x_n|x_{n-1})$ of the Markov chain X are given by

$$p(x_1)=\frac{\beta_1(x_1)}{\sum_{x_1'}\beta_1(x_1')},$$
$$p(x_n|x_{n-1})=\frac{q_{n-1}(x_{n-1},x_n)\beta_n(x_n)}{\beta_n(x_{n-1})}$$
$$\text{for } 2\leqslant n\leqslant N, \tag{3.6}$$

where $\beta_1(x_1),\ldots,\beta_N(x_N)$ are calculated from $q_1,\ldots,q_{N-1}$ by the recursive formulas

$$\beta_N(x_N)=1$$
and
$$\beta(x_{n-1})=\sum_{x_n\in\Omega}\beta_n(x_n)q_{n-1}(x_{n-1},x_n)\text{ for }2\leqslant n\leqslant N. \tag{3.7}$$

In fact, (3.2) is of the form (3.5), with $q_1(x_1,x_2)=p(z_1)p(z_2|z_1),\ldots,q_{N-1}(x_{N-1},x_N)=p(z_N|z_{N-1},y^{n-2})$.

Having $p(x_1|y)$ and the transitions $p(x_{n+1}|x_n,y)$, the marginal distributions $p(x_n|y)$ are classically calculated by the recursive formulas

$p(x_1|y)$ given
and
$$p(x_{n+1}|y)=\sum_{x_n\in\Omega}p(x_n|y)p(x_{n+1}|x_n,y)\text{ for }1\leqslant n\leqslant N-1. \tag{3.8}$$

Finally, an important point is the following. Similarly to the classical case of HMC-IN, the posterior transitions $p(x_{n+1}|x_n,y)$ and the posterior marginal distributions $p(x_n|y)$ are calculable once the transitions $p(z_{n+1}|z_n,y^{n-1})$ considered in (3.3) are calculable for every $1\leqslant n\leqslant N-1$. We will see in the next subsection that this is feasible in a particular PPMC case.

3.2. Long-memory noise

Let $W=(W_1, W_2, \ldots, W_n, \ldots)$ be a real random stationary chain with zero mean and covariance function $\gamma_m=E[W_iW_{i+m}]$. "Long memory" (which is also called "long correlation" or still "long-range dependence"), occurs when the covariances tend to zero like a power function and so slowly that their sum diverges. One possible definition we adopt [29, p. 14]) is $\gamma_m\sim m^{-\alpha}L(m)$ ($g(m)\sim h(m)$ means here that $g(m)/h(m)\underset{x\to+\infty}{\longrightarrow}1$), where L is a *slowly varying* function at infinity, which means that L is bounded on a finite interval and for all $\beta>0$, $L(\beta x)/L(x)\underset{x\to+\infty}{\longrightarrow}1$ (constants and logarithms are examples of slowly varying functions). Thus a Markov chain, in which the covariance decays exponentially, is not a long-memory chain (it is sometimes called "short memory" or "short-range dependence" chain). The long-memory chains have been shown to be useful in numerous situations, in which Markov chains turn out to be little efficient. In particular, one of the interests of the fractional Brownian motion (fBm) $Y=(Y_t)_{t\in R}$, which is a classical widely used stochastic process, is that it has stationary Gaussian increments $W_{n+1}=Y_{n+1}-Y_n$ which form a long-memory chain called "fractional Gaussian noise" (fGn). Such processes are well suited to model different phenomena like data network traffic [29, p. 373], macroeconomics and finance [29, p. 373], or hydrology [29, p. 462]. Let us imagine that there are several possible fGn to model a phenomenon—for example, there are two possible states of the data traffic: "high traffic" and "low traffic"—and that the system can switch from one state to another during the time. Then an interesting question is to find, in an automated way, in which state the system lies. We will solve this problem in the next subsection, by modelling the hidden switching sequence by a Markov chain.

3.3. Markov chains hidden with Gaussian long-memory noise

The aim here is to propose a PPMC in which the transitions $p(z_{n+1}|z_n,y^{n-1})$ considered in (3.3) are calculable for every $1\leqslant n\leqslant N-1$. Let us consider a particular PPMC in which

(i) $p(x_{n+1}|z_n,y^{n-1})=p(x_{n+1}|x_n)$;
(ii) $p(y|x)$ are Gaussian.

Hypothesis (i) implies that X is a Markov chain. To see that, we replace in (3.2) $p(z_{n+1}|z_n,y^{n-1})$ by $p(x_{n+1}|z_n,y^{n-1})p(y_{n+1}|z_n,x_{n+1},y^{n-1})$; then integrating (3.2) with respect to $y_N,\ldots,y_1$, we find $p(x)=p(x_1)\Pi_{n=1}^{N-1}p(x_{n+1}\,|\,z_n,y^{n-1})=p(x_1)\Pi_{n=1}^{N-1}p(x_{n+1}|x_n)$. Of course, hypothesis (ii) is restrictive because of the Gaussianity; however, it remains relatively general for $p(y|x)$ are of any form.

On the one hand, according to (i) we have $p(z_{n+1}|z_n,y^{n-1})=p(x_{n+1}|x_n)\,p\,(y_{n+1}\,|z_n,x_{n+1},y^{n-1})$, and thus the problem is to calculate the transitions $p(y_{n+1}|z_n,x_{n+1},y^{n-1})$, which are also written $p(y_{n+1}|x_n,x_{n+1},y_1^n)$. On the other hand, according to (ii), these transitions are also Gaussian and thus they can be recursively calculated using the following classical "Property 1":

Property 1. Let $W=W_1^N$ be a real Gaussian chain with, for each $1\leqslant n\leqslant N$, $M_1^n=(M_i)_{i=1}^n$ the mean vector and $\Gamma^n=(\gamma_{kl})_{k\leqslant n,l\leqslant n}$ the covariance matrix of $W_1^n=(W_i)_{i=1}^n$. For each n, the Gaussian density $p(y_1^n)$ of the distribution of W_1^n, can be written $p(y_1^n)=p(y_1^{n-1})p(y_n|y_1^{n-1})$, where $p(y_n|y_1^{n-1})$ is Gaussian with mean $M_n+(A^n)^T(\Gamma^{n-1})^{-1}(y_1^{n-1}-M_1^{n-1})$ and variance $\gamma_{nn}-(A^n)^t(\Gamma^{n-1})^{-1}A^n$, where $A^n=\left((\gamma_{i,n})_{i=1}^{n-1}\right)^t$. So, $p(y^n)$ is calculated from $p(y^{n-1})$, which is calculated from $p(y^{n-2})$ …, and so on. Having $p(y^1)$, …, $p(y^N)$, we also have $p(y_n|y^{n-1})$ for every $2\leqslant n\leqslant N$. We will say that $p(y_n|y^{n-1})$ are calculated by a "forward recursion".

The idea is then to apply this property K^2 times (remember that K is the number of possible values for each x_n, and thus (x_n,x_{n+1}) can have K^2 different values). More precisely, y_1^N is fixed and for each (ω_i,ω_j) in Ω^2, Property 1 is used to calculate recursively (for n varying from 1 to $N-1$) the $N-1$ transitions $t_{ij}^{n,n+1}=p(y_{n+1}|x_n=\omega_i,x_{n+1}=\omega_j,y^n)$. Important is that knowing these transitions permits the use of (3.4) to calculate $\beta^n(x_n)$.

Finally, Gaussian PPMC makes possible the calculation of the marginal distributions $p(x_n|y)$ in the following way:

(1) Calculate all transitions $p(z_{n+1}|z_1^n)=p(x_{n+1}|x_n)$ $p(y_{n+1}|x_n,x_{n+1},y_1^n)$ using $p(y_{n+1}|x_n,x_{n+1},y_1^n)$ computed by K^2 forward recursions according to the Property 1 .
(2) Calculate $\beta^n(x_n)$ by backward recursions (3.4) and deduce $p(x_{n+1}|x_n,y)$ and $p(x_1|y)$ with (3.3).
(3) Calculate $p(x_n|y)$ by the classical forward recursions (3.8).

We see that the points (2) and (3) are classical and used in HMC-IN, while the point (1) is new and is due to the "partially" Markov aspect of the model.

Let us consider the problem of calculation of the Gaussian distributions $p(y|x)=p(y_1,\ldots,y_N|x_1,\ldots,x_N)$, which will be useful, in particular, in the parameter estimation problem considered in the next section. As we are going to see, the main difference with the classical models is that $p(y_n|x_1,\ldots,x_n)$ does depend on all $x_1,\ldots,x_n$.

We will use the following classical properties of Gaussian vectors (see, for example, [40]):

Property 2. For $p(w)$ the density of a Gaussian random vector W of mean M and convariance matrix Γ, we will use the notation $p(w)\sim N(M,\Gamma)$.

Let W^1, W^2 be two Gaussian vectors, and let $W=(W^1,\ W^2)$. We have:

(i) $p(w^1,\ w^2)\sim N\left(\begin{bmatrix}M^1\\M^2\end{bmatrix},\begin{bmatrix}\Gamma_1 & \Gamma_{12}\\ \Gamma_{21} & \Gamma_2\end{bmatrix}\right)$ implies $p(w^2|w^1)\sim N(M^2+\Gamma_{21}\Gamma_1^{-1}(w^1-M^1),\Gamma_2-\Gamma_{21}\Gamma_1^{-1}\Gamma_{12})$;
(ii) $p(w^1)\sim N(M^1,\ \Gamma_1)$ and $p(w^2|w^1)\sim N(Aw^1+B,\Gamma_{2/1})$ imply $p(w^1,w^2)\sim N\left(\left[M^1,AM^1+B\right],\begin{bmatrix}\Gamma_1 & \Gamma_1A^{\mathrm{T}}\\ A\Gamma_1 & \Gamma_{2/1}+A\Gamma_1A^{\mathrm{T}}\end{bmatrix}\right)$.

We use these classical properties to establish a relation between $p(y_1^n|x_1^n)$ and $p(y_1^{n+1}|x_1^{n+1})$, which will allow one to compute the latter from the former, and thus to compute $p(y|x)=p(y_1^N|x_1^N)$. We have $p(y_1^{n+1}|x_1^{n+1})=p(y_1^n|x_1^n)p(y_{n+1}|x_1^{n+1},y_1^n)$ with, according to the model, $p(y_{n+1}|x_1^{n+1},y_1^n)=$

$p(y_{n+1}|x_n, x_{n+1}, y_1^n)$. Finally

$$p(y_1^{n+1}|x_1^{n+1}) = p(y_1^n|x_1^n)p(y_{n+1}|x_n, x_{n+1}, y_1^n). \quad (3.9)$$

For $(x_n, x_{n+1}) = (\omega_i, \omega_j)$, $p(y_{n+1}|x_n, x_{n+1}, y_1^n)$ is Gaussian and is obtained by applying (i) to the Gaussian law

$$p(y_1^{n+1}) = p(y_1^n, y_{n+1}) \sim N\left(\begin{bmatrix} M_{ij}^n \\ M_{ij} \end{bmatrix}, \begin{bmatrix} \Gamma_1^{ij} & \Gamma_{12}^{ij} \\ \Gamma_{21}^{ij} & \gamma_{ij} \end{bmatrix}\right)$$

(we have $(w^1, w^2) = (y_1^n, y_{n+1})$). Thus

$$p(y_{n+1}|x_n, x_{n+1}, y_1^n) \sim N(M_{ij} + \Gamma_{21}^{ij}(\Gamma_1^{ij})^{-1}(y^n - M_{ij}^n), \gamma_{ij} - \Gamma_{21}^{ij}(\Gamma_1^{ij})^{-1}\Gamma_{12}^{ij}). \quad (3.10)$$

Let $p(y_1^n|x_1^n) \sim N(M^{x^n}, \Gamma_1^{x^n})$. As the mean in (3.10) is $\Gamma_{21}^{ij}(\Gamma_1^{ij})^{-1}y_1^n + M_{ij} - \Gamma_{21}^{ij}(\Gamma_1^{ij})^{-1}M_{ij} = Ay_1^n + B$, applying (ii) to $p(y_1^n|x_1^n)$ and $p(y_{n+1}|x_n, x_{n+1}, y_1^n)$ gives:

$$p(y_1^{n+1}|x_1^{n+1}) \sim N\left(\begin{bmatrix} M^{x^n} \\ M_{ij} + \Gamma_{21}^{ij}(\Gamma_1^{ij})^{-1}(M^{x^n} - M_{ij}^n) \end{bmatrix}, \begin{bmatrix} \Gamma_1^{x^n} & \Gamma_1^{x^n}[\Gamma_{21}^{ij}(\Gamma_1^{ij})^{-1}]^{\mathrm{T}} \\ \Gamma_{21}^{ij}(\Gamma_1^{ij})^{-1}\Gamma_1^{x^n} & \gamma_{ij} - \Gamma_{21}^{ij}(\Gamma_1^{ij})^{-1}[\Gamma_{12}^{ij} - \Gamma_1^{x^n}[\Gamma_{21}^{ij}(\Gamma_1^{ij})^{-1}]^{\mathrm{T}}] \end{bmatrix}\right). \quad (3.11)$$

In particular, we see according to (3.11) that the mean $m^{x^{n+1}}$ and the variance $\gamma^{x^{n+1}}$ of the Gaussian marginal distributions $p(y_{n+1}|x_1, \ldots, x_{n+1})$ depend on all $x_1, \ldots, x_n, x_{n+1}$ and verify

$$m^{x^{n+1}} = M_{ij} + \Gamma_{21}^{ij}(\Gamma_1^{ij})^{-1}(M^{x^n} - M_{ij}^n),$$
$$\gamma^{x^{n+1}} = \gamma_{ij} - \Gamma_{21}^{ij}(\Gamma_1^{ij})^{-1}[\Gamma_{12}^{ij} - \Gamma_1^{x^n}[\Gamma_{21}^{ij}(\Gamma_1^{ij})^{-1}]^{\mathrm{T}}]. \quad (3.12)$$

3.4. Triplet partially Markov chains

Let $T = (X, U, Y)$ be a triplet random chain, where $X = (X_n)_{1 \leqslant n \leqslant N}$ and $Y = (Y_n)_{1 \leqslant n \leqslant N}$ are two stochastic processes as above, and $U = (U_n)_{1 \leqslant n \leqslant N}$ is a third random chain, each U_n taking its values in a finite set $\Lambda = \{\lambda_1, \ldots, \lambda_M\}$. As in Section 2.2, let us introduce $V = (V_n)_{1 \leqslant n \leqslant N} = (X_n, U_n)_{1 \leqslant n \leqslant N}$ with each V_n taking its values in $\Omega \times \Lambda$.

Definition 3.2. The triplet chain $T = (X, U, Y)$ is called "TPMC" if the pairwise chain $T = (V, Y)$, where $V = (X, U)$, is a PPMC defined in Definition 3.1.

Therefore all the results discussed in Sections 3.1 and 3.2 remain valid once X has been replaced with V. Then we arrive at a more general model; in particular, considering the model developed in the previous subsection in which the chain X has been replaced with V, we see that V is a Markov chain, but X can no longer be Markov. Otherwise, in such TPMC models U can have an intuitive meaning. For example, U can model the semi-Markovianity of X as indicated in [23,36], which leads to a "semi-Markov chain hidden with long-memory noise". Another use of U could be to model different stationarities of $Z = (X, Y)$ [35,41], resulting in a "non-stationary PPMC". Such TPMC can then be used in Bayesian segmentation as indicated in Section 2; furthermore, the parameter estimation method proposed in the next section can be extended to such TPMC leading to unsupervised segmentation methods.

4. Parameter estimation

4.1. Iterative conditional estimation

Let us consider two random processes (X, Y) whose distribution depends on a parameter $\theta = (\theta_1, \ldots, \theta_m) R^m$. The problem is to estimate θ from Y. The most known and used method is the so-called "expectation–maximization" (EM [42]) method, whose aim is to iteratively maximize the likelihood $p(y|\theta)$. The "iterative conditional estimation" (ICE [43]) principle we propose to use is somewhat different from EM and is often easier to perform in complex situations. The intuitive reason behind ICE is the following. To simplify, let $\theta \in R$. In general, one can estimate all parameters from complete data (X, Y) with some estimator $\hat{\theta}(X, Y)$, whose efficiency is often measured by the mean square error $E_\theta[(\theta - \hat{\theta}(X, Y))^2]$. As X is not available, the idea is to approximate $\hat{\theta}(X, Y)$ by some function of Y. The best approximation, in the sense of mean square error, is the conditional expectation $\tilde{\theta}(Y) = E_\theta[\hat{\theta}(X, Y)|Y]$. Thus, on the one hand, $\hat{\theta}(X, Y)$ is close to θ in the mean square error sense and, on the other hand, $\tilde{\theta}(Y)$ is close to $\hat{\theta}(X, Y)$ according to the same criterion. In other words, the possible good "mean square error" properties of $\hat{\theta}(X, Y)$ are saved as far as possible by using $\tilde{\theta}$. Of course, $\tilde{\theta}$ is no longer an estimator because it does depend on θ. However, this leads to an iterative method given by (i) and (ii) below, which is expected to keep, at least when the current θ^q is close to θ, the good properties of $\hat{\theta}(X, Y)$.

Finally, ICE is an iterative method based on the following principle. Let $\hat{\theta}(x,y)$ be an estimator of θ from complete data $(X,Y)=(x,y)$ and let us assume that we can sample realizations of X according to $p(x|y)$. ICE runs as follows:

(i) initialize θ^0;
(ii) using $Y=y$ and the current value of the parameter θ^q, compute $\theta_i^{q+1}=E[\hat{\theta}_i(X,Y)|\,Y=y,\theta^q)$. for the components θ_i for which this computation is feasible;
(iii) for other components θ_i, simulate $x_1^q,\ldots,x_l^q$ independent realizations of X according to $p(x|y,\theta^q)$ and set $\theta_i^{q+1}=(\hat{\theta}(x_1^q,y)+\cdots+\hat{\theta}(x_l^q,y))/l$.

Let us notice that in (iii) one simply approximates, using the law of large numbers, the expectation by the empirical mean. In principle, the greater is l the better is the approximation; however, in practice taking small l, or even $l=1$, can have little influence on the final estimation results. Otherwise, we will see that in the problem we are concerned with in this paper, the point (ii) can be applied to the components θ_i defining the distribution $p(x)$, while we have to use the point (iii) for the components θ_i defining $p(y|x)$.

We see that ICE is applicable under two very slight hypotheses: existence of an estimator $\hat{\theta}(x,y)$ from the complete data, and the ability of simulating X according to $p(x|y)$. The first hypothesis is not really a constraint because if we are not able to estimate θ from complete data (x,y), there is no point in searching an estimator from incomplete ones given by y. The second hypothesis is always verified for PPMC $Z=(X,Y)$ (or TPMC $T=(X,U,Y)$); in fact, $p(x|y)$ (or $p(x,u|y)$) is a Markov chain distribution.

Let us notice that since its introduction in [43] ICE has been successfully applied in many problems of unsupervised statistical signal or image segmentation [15,16,19], sometimes using complex models, like hidden fuzzy (also said "mixed-states") Markov fields [33], hidden evidential Markov fields [32], or triplet Markov fields [31]. Otherwise, some relationships between ICE and EM has been specified in [44], and some preliminary results concerning the asymptotic behaviour of ICE in the case of independent data have been proposed in [45].

4.2. ICE in Markov chains hidden with Gaussian long-memory noise

The results presented in Section 3 above remain valid for any forms of the K^2 Gaussian chains (given by K^2 mean vectors M_1^N and K^2 variance–co-variance matrices $\Gamma^N=(\gamma_{kl})_{1\leqslant k\leqslant N,1\leqslant l\leqslant N}$). In this subsection we will consider a particular case. The first simplification consists of taking $p(y_{n+1}|x_n,x_{n+1},y^n).=p(y_{n+1}|x_{n+1},y^n)$, which means that $p(y_{n+1}|z_n,x_{n+1},y^{n-1})$ depends on x_{n+1} but not on x_n. Thus we have K Gaussian chains instead of K^2. This is not a very significant simplification and the parameter estimation method proposed in this subsection remains valid in the general case; we make it just because such a simpler model is used in experiments below. Furthermore, for each Gaussian chains considered the means vector M_1^N will have all its components equal, and the variance–covariance matrix will have the following form:

$$\gamma_{kl}=\gamma(|k-l|+1)^{-\alpha}\ \text{ with }\ \alpha\in]0,1[. \tag{4.1}$$

Otherwise, the Markov chain X will be assumed stationary, which means that $p(x_n,x_{n+1})$ does not depend on $n=1,\ldots,N-1$.

Such a model will be called "HMC with long-memory noise" (HMC-LMN). Let us specify the real parameters defining, for K classes, such a model. First, the distribution $p(x)$ of the stationary Markov chain X is defined by K^2 parameters $p_{ij}=p(x_1=\omega_i,\ x_2=\omega_j),\ 1\leqslant i,j\leqslant K$. Second, each of the K Gaussian chains included in the model is defined by three parameters (M,γ,α), where M is the common value of the components of M_1^N, and (γ,α) are the parameters defining the covariance matrix with (4.1). Finally, we have $3K$ parameters (M_1,γ_1,α_1), …, (M_K,γ_K,α_K). Therefore the number of components of θ, which is the global number of parameters, is K^2+3K. According to the notations of the previous subsection, θ^q will designate the current parameters in ICE procedure, which are the parameters obtained after q iterations.

According to the ICE principle, we have to consider an estimator $\hat{\theta}(X,Y)$ from complete data (X,Y). Concerning the parameters p_{ij}, their estimation from the complete data can be ensured by the classical estimator

$$\hat{p}_{ij}(x,y)=\frac{1}{N-1}\sum_{n=1}^{N-1}1_{[x_n=\omega_i,x_{n+1}=\omega_j)}. \tag{4.2}$$

Knowing that for a given set A the expectation of the function 1_A is the probability of A (the same is true for conditional expectation and conditional probability), we see that the conditional expectation in the point (ii) of the previous subsection applied to (4.2) gives

$$p_{ij}^{q+1} = E[\hat{p}_{ij}(X,Y)|\,Y=y,\theta^q) = \frac{1}{N-1}\sum_{n=1}^{N-1} p(x_n=\omega_i, x_{n+1}=\omega_j|y,\theta^q). \quad (4.3)$$

It is important to note that each $p(x_n=\omega_i, x_{n+1}=\omega_j|y)$ is computable; in fact, we have $p(x_n, x_{n+1}|y) = p(x_n|y)p(x_{n+1}|x_n, y)$, where all $p(x_{n+1}|x_n, y)$ are computable using (3.3), and all $p(x_n|y)$ are computable using (3.8). Therefore, p_{ij}^{q+1} are computable. Finally, the conditional expectation in (ii) in the previous subsection is computable for the components of θ of the form p_{ij}.

Concerning the parameters (M_i,γ_i,α_i), which are the remaining components of θ, things are more complicated because their estimation from the complete data is not immediate. To see the difficulty, let us consider the following example.

Example 4.1. Let $K=2$ and $N=10$. Thus we observe x_1^{10}, y_1^{10} and the problem is to estimate (M_1,γ_1,α_1) and (M_2,γ_2,α_2). According to (4.1), we have two Gaussian chains defined by the distributions

$$N\left(\begin{bmatrix}M_1\\M_1\end{bmatrix},\begin{bmatrix}\gamma_1 & \rho_1\\ \rho_1 & \gamma_1\end{bmatrix}\right) \text{ and } N\left(\begin{bmatrix}M_2\\M_2\end{bmatrix},\begin{bmatrix}\gamma_2 & \rho_2\\ \rho_2 & \gamma_2\end{bmatrix}\right),$$

where $\rho_1 = 2^{-\alpha_1}\gamma_1$ and $\rho_2 = 2^{-\alpha_2}\gamma_2$, and thus the problem is to estimate the means M_1, M_2, the variances, γ_1, γ_2 and the covariances ρ_1, ρ_2 from x_1^{10}, y_1^{10}. As an example, let us consider $x_1^{10} = (\omega_1,\omega_1,\omega_2,\omega_2,\omega_1,\omega_1,\omega_2,\omega_2,\omega_2,\omega_2)$. The classical hypothesis (as, for example, in the classical HMC-IN) is to assume that $p(y_i|x_1^{10}) = p(y_i|x_i)$. Under such hypothesis the problem is quite simple; in fact, we can classically set

$$\hat{M}_1(x_1^{10},y_1^{10}) = \frac{y_1+y_2+y_5+y_6}{4},$$
$$\hat{M}_2(x_1^{10},y_1^{10}) = \frac{y_3+y_4+y_7+y_8+y_9+y_{10}}{6}, \quad (4.4)$$

and similar formula for $\hat{\Gamma}_2(x_1^{10},y_1^{10})$, with $\hat{M}_1$ replaced by $\hat{M}_2$ and (y_1, y_2), (y_5, y_6) replaced by (y_3, y_4), (y_7, y_8), (y_9, y_{10}) (there would be three terms in the sum defining $\hat{\Gamma}_2(x_1^{10},y_1^{10})$). Using formula (4.5) is possible in classical models because $p(y_i, y_{i+1}|x_1,\ldots,x_{i+1}) = p(y_i, y_{i+1}|x_i, x_{i+1})$ and the difficulty comes from the fact that the latter equality is no longer true in the HMC-LMN considered in this paper. In fact, returning to our example and considering $x_1^4 = (\omega_1,\omega_1,\omega_2,\omega_2)$, $y_1^4 = (y_1,y_2,y_3,y_4)$ extracted from x_1^{10}, y_1^{10} above, we see that

$$\Gamma_1 = \begin{bmatrix}\gamma_1 & \rho_1\\ \rho_1 & \gamma_1\end{bmatrix}$$

is the covariance matrix of $p(y_1,y_2|x_1,x_2)$, but

$$\Gamma_2 = \begin{bmatrix}\gamma_2 & \rho_2\\ \rho_2 & \gamma_2\end{bmatrix}$$

is not the covariance matrix of $p(y_3,y_4|x_1^4)$. In fact, to obtain the covariance matrix of the latter distribution, we have to use the Property 2 in Section 3. The searched covariance matrix Γ^* of $p(y_3,y_4|x_1^4)$ is a sub-matrix of the covariance matrix

$$\Gamma^{1,4} = \begin{bmatrix}\Gamma^{1,2} & \Delta\\ \Delta & \Gamma^*\end{bmatrix}.$$

of $p(y_1,y_2,y_3,y_4|x_1^4)$.

Let

$$\Gamma = \begin{bmatrix}\gamma_2 & 2^{-\alpha_2}\gamma_2 & 3^{-\alpha_2}\gamma_2 & 4^{-\alpha_2}\gamma_2\\ 2^{-\alpha_2}\gamma_2 & \gamma_2 & 2^{-\alpha_2}\gamma_2 & 3^{-\alpha_2}\gamma_2\\ 3^{-\alpha_2}\gamma_2 & 2^{-\alpha_2}\gamma_2 & \gamma_2 & 2^{-\alpha_2}\gamma_2\\ 4^{-\alpha_2}\gamma_2 & 3^{-\alpha_2}\gamma_2 & 2^{-\alpha_2}\gamma_2 & \gamma_2\end{bmatrix} = \begin{bmatrix}\Gamma_2 & A_2\\ A_2^{\mathrm{T}} & \Gamma_2\end{bmatrix} \quad (4.6)$$

be the covariance matrix of the distribution of the Gaussian chain corresponding to $x_3 = x_4 = \omega_2$. According to the point (i) of the Property 2 (with $M^1 = M^2 = \begin{bmatrix}M_2\\M_2\end{bmatrix}$, $w^1 = y_1^2$ and $w^2 = y_3^4$), the distribution $p(y_3,y_4|x_3^4,y_1,y_2) = p(y_3^4|x_3^4,y_1^2)$ is Gaussian with mean $M^2 + A_2^{\mathrm{T}}\Gamma_2^{-1}(y_1^2 - M^2) = A_2^{\mathrm{T}}\Gamma_2^{-1}y_1^2 + M^2 - A_2^{\mathrm{T}}\Gamma_2^{-1}M^2 = Ay_1^2 + B$ and covariance matrix $\Gamma_2 - A_2\Gamma_2^{-1}A_2^{\mathrm{T}}$. The searched $p(y_1^4|x_1^4)$ is then

$$\hat{\Gamma}_1(x_1^{10},y_1^{10}) = \frac{(y_1-\hat{M}_1, y_2-\hat{M}_1)\begin{pmatrix}y_1-\hat{M}_1\\ y_2-\hat{M}_1\end{pmatrix} + (y_5-\hat{M}_1, y_6-\hat{M}_1)\begin{pmatrix}y_5-\hat{M}_1\\ y_6-\hat{M}_1\end{pmatrix}}{2} \quad (4.5)$$

obtained by applying the point (ii) of the Property 2, with $w^1 = y_1^2$, $w^2 = y_3^4$, $p(w^1) = p(y_1^2|x_1^2)$, $p(w^2|w^1) = p(y_3^4|x_3^4, y_1^2)$, $M^1 = \begin{bmatrix} M_1 \\ M_1 \end{bmatrix}$, $A = A_2^{\mathrm{T}}\Gamma_2^{-1}$, $B = M^2 - A_2^{\mathrm{T}}\Gamma_2^{-1}M^2$, and $\Gamma_{2/1} = \Gamma_2 - A_2^{\mathrm{T}}\Gamma_2^{-1}$, A_2. Recalling that we search the mean M^* and the covariance matrix Γ^* of $p(y_3, y_4|x_1^4)$, we have

$$M^* = M^2 + A_2^{\mathrm{T}}\Gamma_2^{-1}(M^1 - M^2), \tag{4.7}$$

$$\Gamma^* = \Gamma_2 - A_2^{\mathrm{T}}\Gamma_2^{-1}A_2 + A_2^{\mathrm{T}}\Gamma_2^{-1}\Gamma_1[A_2^{\mathrm{T}}\Gamma_2^{-1}]^{\mathrm{T}}. \tag{4.8}$$

Finally, $p(y_3^4|x_1^4) \sim N(M^*, \Gamma^*)$ and we see that it is not possible to apply (4.4) and (4.5) as in classical models. Then the idea is to find a linear transformation $y_3^4 \to (y_3^4)'$ in such a way that $p((y_3^4)'|x_1^4) \sim N(M^2, \Gamma_2)$, which will make possible the use of classical estimation. Considering C and D such that $\Gamma_2 = CC^{\mathrm{T}}$ and $\Gamma^* = DD^{\mathrm{T}}$, we verify that

$$(y_3^4)' = CD^{-1}(y_3^4 - M^*) + M^2 \tag{4.9}$$

is the searched transformation. Of course, C and D do depend on unknown parameters and thus doing so does not provide an estimator. However, we will see below that such a way can be used inside ICE, where at each iteration q the matrices C and D will be computed using the current value of parameters θ^q.

Let us return to the general situation and let us generalize the calculations presented in Example 4.1 above. Let y_1^N be the observed data, and let $x_1^{N,q}$ be the sampled x_1^N according to $p(x|y,\theta^q)$, as specified in the point (iii) of the definition of ICE. To simplify, $x_1^{N,q}$ will be denoted by x_1^N, or even by x, in the following. Our aim is to find a linear transformation $y_1^N \to (y_1^N)'$, depending on the current parameters θ^q, in such a way that for each $i = 1,\ldots,K$, if $x_j = x_{j+1} = \omega_i$, then

$$p(y_j', y_{j+1}'|x_1^{j+1}) \sim N\left(\begin{bmatrix} M_i \\ M_i \end{bmatrix}, \begin{bmatrix} \gamma_i & 2^{-\alpha_i}\gamma_i \\ 2^{-\alpha_i}\gamma_i & \gamma_i \end{bmatrix}\right) = N(M^i, \Gamma_i).$$

Then the estimation of $(M_1,\gamma_1,\alpha_1), \ldots, (M_K,\gamma_K,\alpha_K)$ from $(X,\ Y') = (x,\ y')$, which will give the next parameters θ^{q+1} in ICE, will be similar to the estimation specified in Example 4.1, which is quite a classical one.

Thus let us consider the sampled $x = (x_1,\ldots,x_N)$ and a class ω_i. Let us set $J(i)$ the set of j such that $x_j = x_{j+1} = \omega_i$ and let us assume that $J(i)$ is not empty and contains r elements $n_1 < n_2 < \ldots < n_r$. Knowing that $p(y_1^{n_1+1}|x_1^{n_1+1}), \ldots, p(y_1^{n_r+1}|x_1^{n_r+1})$ are Gaussian and are computable as specified in Section 3.3, we obtain $p(y_{n_1}^{n_1+1}|x_1^{n_1+1}) \sim N(M^{1,*}, \Gamma^{1,*}), \ldots, p(y_{n_r}^{n_r+1}|x_1^{n_r+1}) \sim N(M^{r,*}, \Gamma^{r,*})$ as marginal distributions of $p(y_1^{n_1+1}|x_1^{n_1+1}), \ldots, p(y_1^{n_r+1}|x_1^{n_r+1})$. Then the transformation described in Example 4.1 is applied to each $y_{n_1}^{n_1+1}, \ldots, y_{n_r}^{n_r+1}$. More precisely, let us consider C_i such that $\Gamma_i = C_i(C_i)^{\mathrm{T}}$ and $D_i^{1,*}, \ldots, D_i^{r,*}$ such that $\Gamma^{1,*} = D_i^{1,*}(D_i^{1,*})^{\mathrm{T}}, \ldots, \Gamma^{r,*} = D_i^{r,*}(D_i^{r,*})^{\mathrm{T}}$ and let us set

$$(y_{n_1}^{n_1+1})' = C_i(D_i^{1,*})^{-1}(y_{n_1}^{n_1+1} - M^{1,*}) + M^i, \ldots,$$
$$(y_{n_r}^{n_r+1})' = C_i(D_i^{r,*})^{-1}(y_{n_r}^{n_r+1} - M^{r,*}) + M^i. \tag{4.10}$$

Then we verify that all random vectors $(Y_{n_1}^{n_1+1})', \ldots, (Y_{n_r}^{n_r+1})'$ are Gaussian with mean

$$M^i = \begin{bmatrix} M_i \\ M_i \end{bmatrix},$$

and covariance matrix

$$\Gamma_i = \begin{bmatrix} \gamma_i & 2^{-\alpha_i}\gamma_i \\ 2^{-\alpha_i}\gamma_i & \gamma_i \end{bmatrix},$$

which makes possible the use of classical estimators.

Finally, the parameters (M_i,γ_i,α_i) (for $i = 1,\ldots,K$) are estimated from $(x,\ y) = (x_1,y_1,\ \ldots,\ x_N,\ y_N)$, where $x = (x_1,\ldots,x_N)$ is a $x_1^{N,q}$ sampled according to $p(x|y,\theta^q)$ in the following way:

(i) for $n = 2, \ldots, N$, calculate $p(y_1^n|x_1^n)$ using (3.11);
(ii) consider $J(i) = \{n_1,\ldots,n_r\}$, with $n_1 < \ldots < n_r$, the set of j such that $x_j = x_{j+1} = \omega_i$ and consider the corresponding sample $(y_{n_1}, y_{n_1+1}), \ldots, (y_{n_r}, y_{n_r+1})$. Using (i) find $p(y_{n_1}^{n_1+1}|x_1^{n_1+1}) \sim N(M^{1,*}, \Gamma^{1,*}), \ldots, p(y_{n_r}^{n_r+1}|x_1^{n_r+1}) \sim N(M^{r,*}, \Gamma^{r,*})$;
(iii) calculate C_i such that $\Gamma_i = C_i(C_i)^{\mathrm{T}}$ and $D_i^{1,*}, \ldots, D_i^{r,*}$ such that $\Gamma^{1,*} = D_i^{1,*}(D_i^{1,*})^{\mathrm{T}}, \ldots, \Gamma^{r,*} = D_i^{r,*}(D_i^{r,*})^{\mathrm{T}}$;
(iv) calculate $(y_{n_1}^{n_1+1})', \ldots, (y_{n_r}^{n_r+1})'$ using (iii) and (4.10);
(v) calculate

$$\hat{M}_i = \frac{1}{r}\left(\begin{bmatrix} y'_{n_1+1} \\ y'_{n_1} \end{bmatrix} + \cdots + \begin{bmatrix} y'_{n_r+1} \\ y'_{n_r} \end{bmatrix}\right), \tag{4.11}$$

$$\hat{\Gamma}_2^i = \begin{bmatrix} \hat{\gamma}_i & \hat{\gamma}_{1i} \\ \hat{\gamma}_{1i} & \hat{\gamma}_i \end{bmatrix} = \frac{1}{r}\left[\left(\begin{bmatrix} y'_{n_1+1} \\ y'_{n_1} \end{bmatrix} - \hat{M}_i\right) \times \left(\begin{bmatrix} y'_{n_1+1} \\ y'_{n_1} \end{bmatrix} - \hat{M}_i\right)^{\mathrm{T}} + \cdots + \left(\begin{bmatrix} y'_{n_r+1} \\ y'_{n_r} \end{bmatrix} - \hat{M}_i\right) \times \left(\begin{bmatrix} y'_{n_r+1} \\ y'_{n_r} \end{bmatrix} - \hat{M}_i\right)^{\mathrm{T}}\right] \quad (4.12)$$

giving $\hat{M}_i$, $\hat{\gamma}_i$, and $\hat{\alpha}_i = -\mathrm{Log}(\hat{\gamma}_i)/\mathrm{Log}(2)$, which are thus the next values of the parameters in ICE procedure.

Remark 4.1. The ICE procedure we propose is somewhat more complicated that the classical version. In fact, in classical application the current value of the parameter θ^q is used to sample x^q according to $p(x|y,\theta^q)$, and the next value θ^{q+1} is given by $\theta^{q+1} = \hat{\theta}(x^q, y)$. In the application described above θ^q is used twice. First, x^q is sampled just like in the classical case. Second, both θ^q and x^q are used to calculate y', and then $\theta^{q+1} = \hat{\theta}(x^q, y')$.

5. Experiments

We provide below three series of experiments. In the first series, we show that the information contained in the sole "long memory" correlation can be sufficient to perform a viable segmentation. In the second series we test the new model and the related parameter estimation method in the general case and, in particular, we compare it to the classical hidden Markov chains model. Finally, a real SAR image is segmented with the classical HMC-IN and with the new HMC-LMN model.

5.1. Common means and variances

In this subsection, we will assume that the means and the variances in the Gaussian chains are equal. To illustrate the results visually, we will consider a four classes image presented in Fig. 1. The two-dimensional set of pixels is transformed into a one-dimensional set of indices via the so-called "Hilbert–Peano" scan, and thus the class image $X = x$ is viewed as a realization of a mono-dimensional stationary Markov chain X, whose parameters are estimated from $X = x$. Then four Gaussian long-memory noises are considered as described in the previous section. The four means are equal, and it is the same for the four variances, all equal to one. Thus the correlations are the only discriminating parameters. All the four autocorrelations are of the form $\rho(\tau_{ij}) = |\tau_{ij}+1|^{-\alpha}$, where $\tau_{ij} = |j-i|$. Then we consider the following four parameters: $\alpha_{\omega_1} = 0.99$ for the class "black", $\alpha_{\omega_2} = 0.3$ for the class "dark grey", $\alpha_{\omega_3} = 0.05$ for the class "bright grey", and $\alpha_{\omega_4} = 0.01$ for the class "white". Finally, the model parameters are: the distribution $p(x_1,x_2)$ on $\Omega^2 = \{\omega_1, \omega_2, \omega_3, \omega_4\}^2$, which is estimated from the chain $X = x$, and α_{ω_1}, α_{ω_2}, α_{ω_3}, α_{ω_4}, which are known. The class image, its noisy version, and the MPM segmentation are presented in Fig. 1; we can see that the noise is rather strong and the human eye cannot distinguish the four classes. However, the segmentation result enables one to have a relatively good idea about their position in the image. The misclassified pixels' ratio is

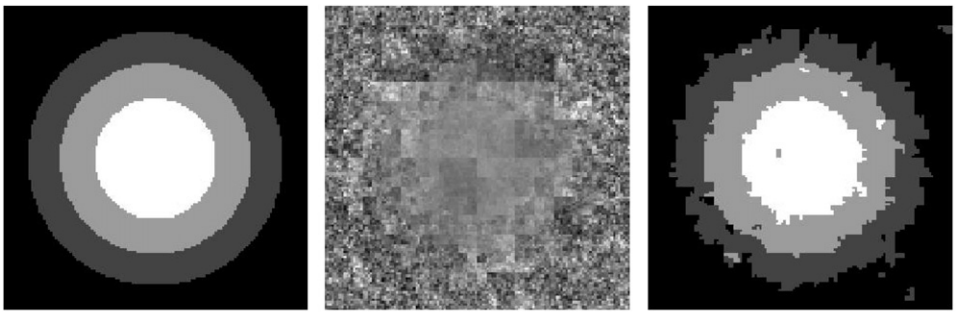

Fig. 1. MPM segmentation of a four class image noisy with a long-memory Gaussian noise. The means and the variances of the noise are equal for the four classes.

equal to 6.9%. This show that the information contained in the sole long-memory correlations can be rich enough to enable a viable segmentation with the proposed HMC-LMN model. This is all the more interesting that the realization $X = x$ can be hardly considered as a stationary Markov chain, which indicates that the HMC-LMN used is robust with respect to the stationarity of the hidden chain X.

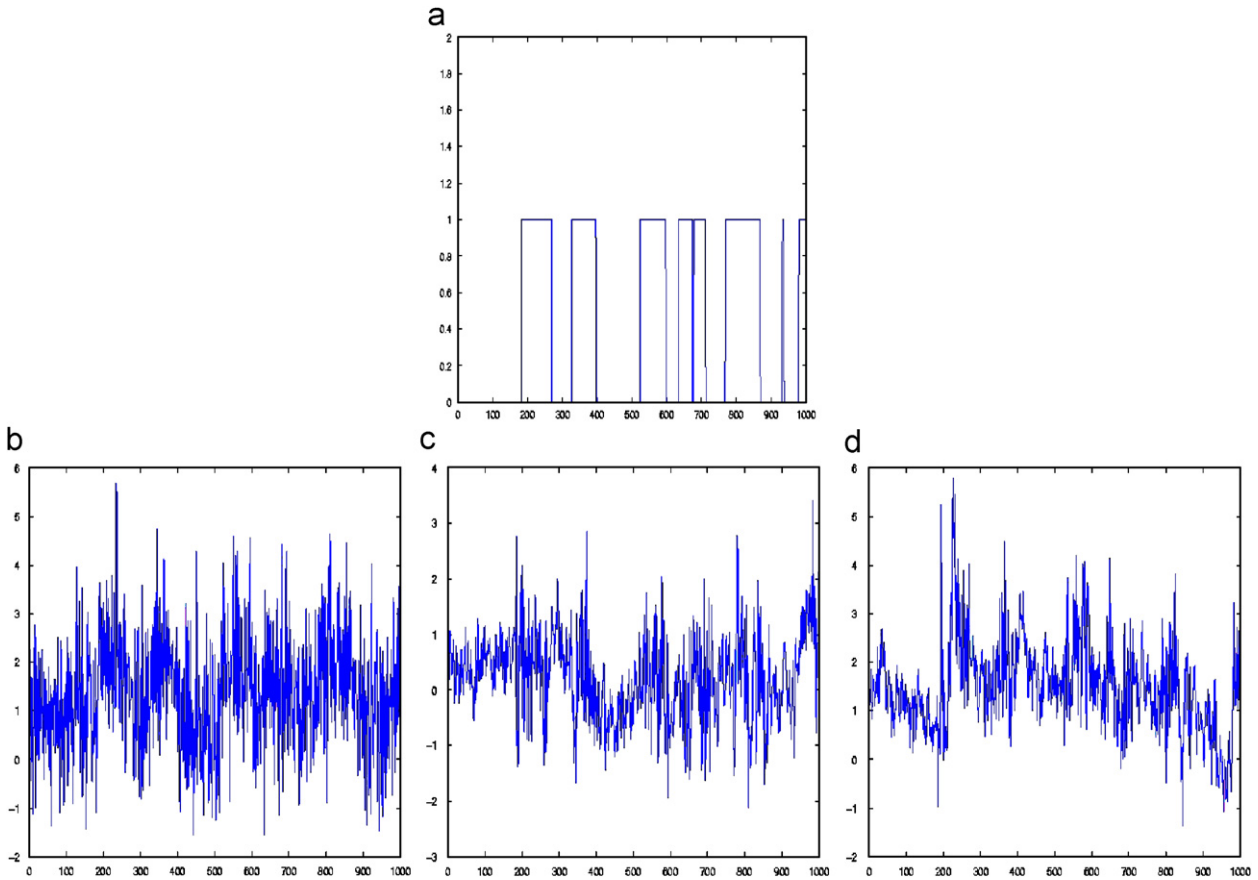

Fig. 2. (a) A two classes process; (b) the noisy version with independent noise; (c) the noisy version with long-memory noise (α is the only discriminating parameter) and (d) noisy version with long-memory noise (same variances).

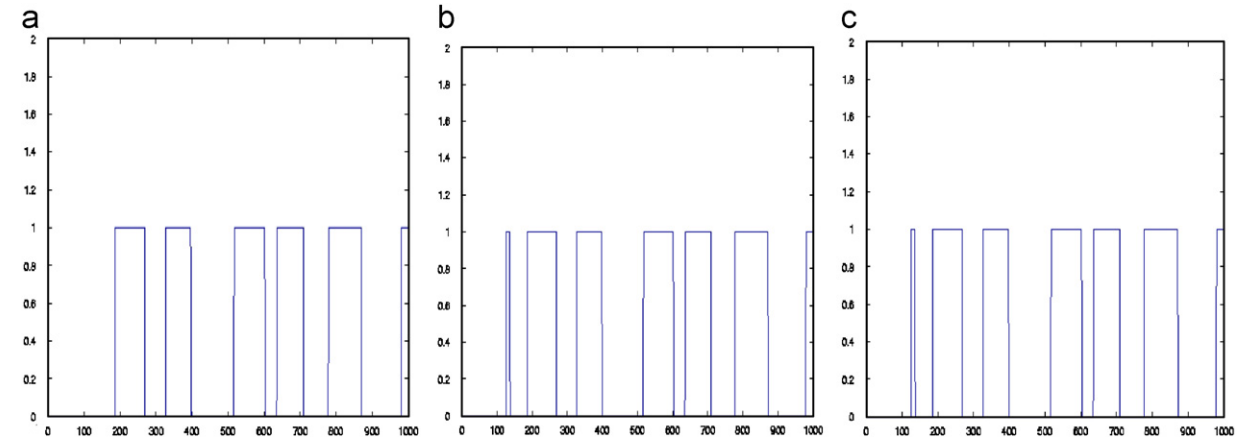

Fig. 3. Segmentation of the observable realization $Y = y$ of an HMC-IN considering three models: (a) HMC-IN with true parameters values—4.1%; (b) HMC-IN with ICE estimates—5.2% and (c) HMC-LMN with ICE estimates—5.2%.

5.2. General long-memory noise model

In this subsection, we propose three series of experiments. For all of them, the HMC X is stationary with $p(x_1=\omega_1,x_2=\omega_1)=p(x_1=\omega_2,x_2=\omega_2)=0.495$, and $p(x_1=\omega_1,x_2=\omega_2)=p(x_1=\omega_2,x_2=\omega_1)=0.005$. We consider samples of size 1000, whose example is presented in Fig. 2(a).

In the first experiment we consider a classical HMC-IN, with means equal, respectively, to 1 and 2, and variances equal to 1. The obtained realisation of $Y=y$, presented in Fig. 2(b), is then segmented by three methods. The first one is the MPM method based on true parameters; thus the result, presented in Fig. 3(a), is the reference one. The second method is the MPM unsupervised method based on the classical HMC-IN and ICE, while the third method is the MPM unsupervised method based on the new HMC-LMN model and the new related ICE. The aims of this experiment are, on the one hand, to show the robustness of the HMC-LMN model and, on the other hand, to see how the new model manages the independent noise.

Table 1
Parameters of the HMC-IN and HMC-LMN models estimated from the observable realization $Y=y$ of an HMC-IN (for numerical reasons α bigger than 100 is not considered). True parameters values are $m_1=1$, $m_2=2$, $\sigma_2^1=\sigma_2^2=1$

IN	HMC-IN		HMC-LMN	
	ω_1	ω_2	ω_1	ω_2
m	0.92	1.99	0.89	1.96
σ^2	1	1	0.98	1.05
α	–	–	>100	>100

According to the results presented in Fig. 3(b) and (c), we see that the new model gives comparable results, which shows its good robustness. This is due to the good behaviour of the parameter estimation method; in fact, according to the results presented in Table 1, means and variances are well estimated, and the estimated α is superior to 100, which means that the covariance decreases very quickly. These results—and other similar results we obtained—seem to us very important; in fact, although the classical HMC-IN is not a particular case of the new HMC-LMN model, the latter can be very close to the former.

The second example is complementary to the first one: both means are equal to 0, and both variances are equal to 1, while α is equal to 0.1 and 1, respectively. Therefore there is no classical HMC-IN which could provide such data—one example of such a sample is presented in Fig. 2(c)—and it is interesting to see the robustness of the classical HMC-IN with respect to data produced by MNC-LMN. According to Fig. 4 we see that the unsupervised segmentation result provided by HMC-LMN is very good, while HMC-IN gives very poor results. Concerning the parameter estimation results presented in Table 2, we see, as expected, that means and variances are poorly estimated while considering the classical HMC-IN. However, we also see that the results obtained with the new ICE used in the new HMC-LMN

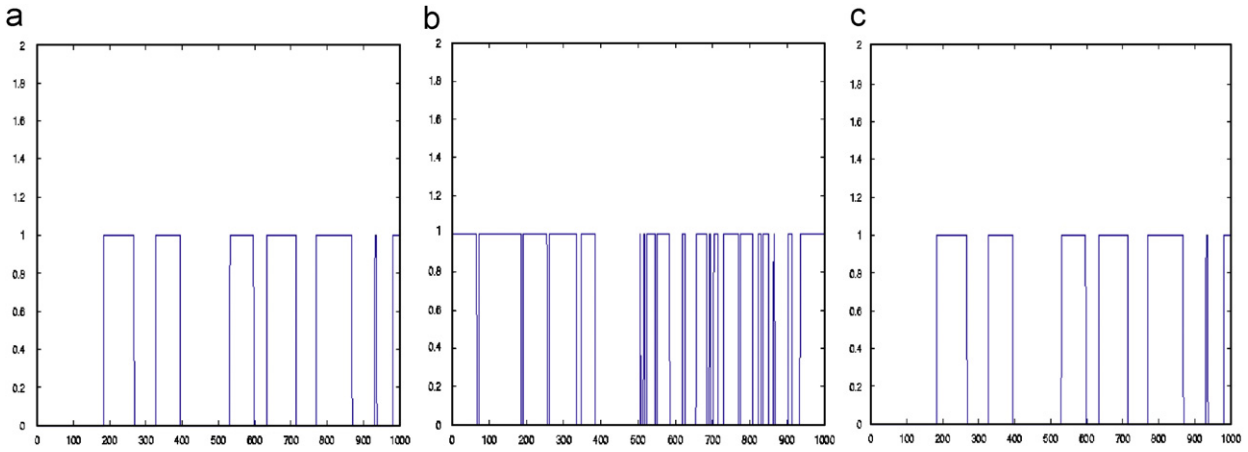

Fig. 4. Segmentation of the observable realization $Y=y$ of an HMC-LMN with same means and same variances considering three models: (a) HMC-LMN with true parameters values—2.1%; (b) HMC-IN with ICE estimates—48% and (c) HMC-LMN with ICE estimates—1.9%.

context are not as good as in the previous case. This could be possibly due to the fact that correlated noise carries less information than the independent one.

Finally, in the third more general example, the means are different (equal to 1 and 2, respectively), and the parameters α are also different (equal to 0.1 and 0.9, respectively). According to the means and variances estimation results presented in Table 3, we see that the new ICE used in the HMC-LMN context gives better results than the classical ICE used in the HMC-IN context; however, the difference is not so striking. Now, HMC-IN cannot take the correlations into account, when HMC-LMN can, and this is probably the reason for the excellent unsupervised segmentation obtained with HMC-LMN, while the results obtained with HMC-IN are quite poor (see Fig. 5). This means that when one class produces a long-memory noise the unsupervised segmentation based on the classical HMC-IN model, which cannot take this fact into account, encounters problems and is not robust.

Finally, the general conclusion involved by these experiments—and other similar experiments we have performed—is that using the new model and related processing leads to superior, or equal, performances than using the classical model and related processing. Moreover, the advantage can turn out to be quite significant.

Table 2
Parameters of the HMC-IN and HMC-LMN models estimated from the observable realization $Y = y$ of an HMC-LMN with same means and same variances (for numerical reasons α bigger than 100 is not considered). True parameters values are $m_1 = m_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, $\alpha_1 = 0.1$, $\alpha_2 = 1$

IN	HMC-IN		HMC-LMN	
	ω_1	ω_2	ω_1	ω_2
m	0.61	−0.41	0.24	0.22
σ^2	0.35	0.26	0.37	0.81
α	–	–	0.28	1.1

Table 3
Parameters of the HMC-IN and HMC-LMN models estimated from the observable realization $Y = y$ of an HMC-LMN with different means and different α (for numerical reasons α bigger than 100 is not considered). True parameters values are $m_1 = 1$, $m_2 = 2$, $\sigma_1^2 = \sigma_2^2 = 1$, $\alpha_1 = 0.1$, $\alpha_2 = 0.9$

IN	HMC-IN		HMC-LMN	
	ω_1	ω_2	ω_1	ω_2
m	1.06	2.25	1.33	1.73
σ^2	0.42	0.67	0.53	1.17
α	–	–	0.22	0.66

Table 4
Initialization of ICE obtained with K-means method

		1st model		2nd model		3rd model	
		ω_1	ω_2	ω_1	ω_2	ω_1	ω_2
True values	m	1	2	0	0	1	2
	σ^2	1	1	1	1	1	1
K-means	m	0.52	2.24	−0.61	0.67	1.6	2.99
Estimates	σ^2	0.39	0.42	0.26	0.23	0.26	0.27

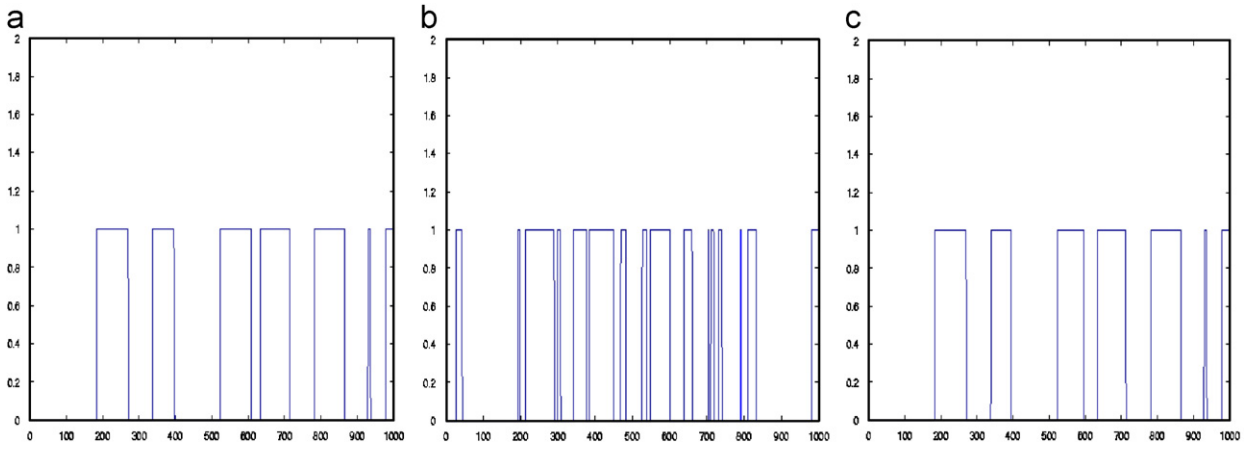

Fig. 5. Segmentation of the observable realization $Y = y$ of an HMC-LMN with different means and different α considering three models: (a) HMC-LMN with true parameters values—4.9%; (b) HMC-IN with ICE estimates—32.4% and (c) HMC-LMN with ICE estimates—4.1%.

Concerning the initialization of ICE, the *K*-means method used provides means and variances given in Table 4, while α is set equal to 0.5.

5.3. Application to a real SAR image

Let us consider a satellite SAR image, presented in Fig. 5(a), which is a photograph of Giens' bay close to Antibes, France. It has been coloured artificially and the colours represent received radar intensity. We transform the colour image into a grey level one by representing the 256 colours by 256 numbers uniformly separated between 0 and 10 ("black = 0" and "white = 10"). Finally, the two-dimensional set of pixels is transformed into a one-dimensional set via the Hilbert–Peano scan.

The image is segmented by two unsupervised methods based on HMC-IN and HMC-LMN, respectively. To limit algorithmic complexity in HMC-LMN, we assume that the distribution of each y_n only depends on the previous values y_k such that $x_n = x_{n-1} = \cdots = x_k$. Thus the density of the joint process (X, Y) can still be written $p(x,y) = p(x_1)\prod_{n=1}^{N-1} p(x_{n+1}|x_n)p(y_{n+1}|x_n, x_{n+1}, y_1, \ldots, y_n)$, with $p(y_{n+1}|x_n, x_{n+1}, y_1, \ldots, y_n) = p(y_{n+1}|x_{n+1})$ if $x_{n+1} \neq x_n$.

Table 5
Initialization of ICE using *K*-means

Classes	m	σ^2
1	1.91	0.35
2	3.36	0.49
3	7.03	1.36

As above, means and variances are initialized using *K*-means, whereas for the long-dependence parameter α has been taken equal to one (Table 5).

The picture is segmented into three classes: "sea" (class 1), "coast or islands" (class 2), and "land" (class 3), corresponding to the colours "blue", "green" and "brown", and the results obtained are presented in Fig. 6. As we do not dispose of the exact image, it is difficult to draw definitive conclusions; however, on the whole, HMC-LMN seems to give better results. In particular, it better recognizes details like islands. Otherwise, HMC-IN accounts more difficulties in distinguishing the land and the sea. Otherwise, according to the estimated values of the parameters given in Table 6, we see that the three noises related to the three classes are all "long-memory" noises. The class 1 (sea) presents the "longest" memory, and the class 3 (land) presents the "shortest" one.

Finally, it can be interesting to test if "sea area", in blue in Fig. 6(c), is really of long-range dependence. For that, we estimate the correlation

Table 6
Estimation of the parameters of a real image according to the models HMC-IN and HMC-LMN

Parameters	HMC-IN			HMC-LMN		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
m	1.69	2.75	5.93	2.29	2.36	5.67
σ^2	0.34	0.11	3.07	0.37	1.12	4.07
α	–	–	–	0.02	0.08	0.31
Computer time	10 s			20 min		

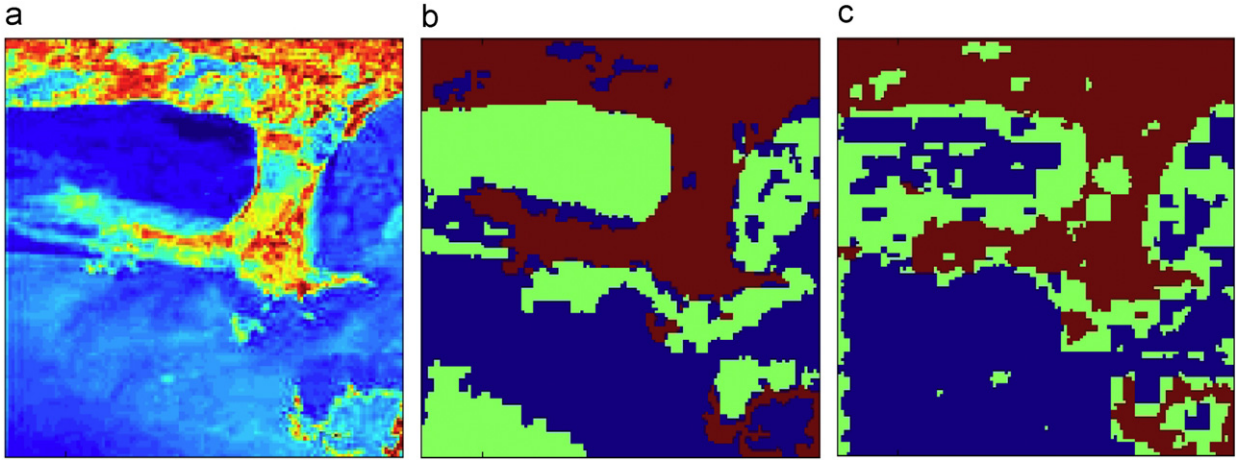

Fig. 6. Unsupervised segmentation of an SAR image using HMC-IN and HMC-LMN: (a) SAR image; (b) HMC-IN and (c) HMC-LMN.

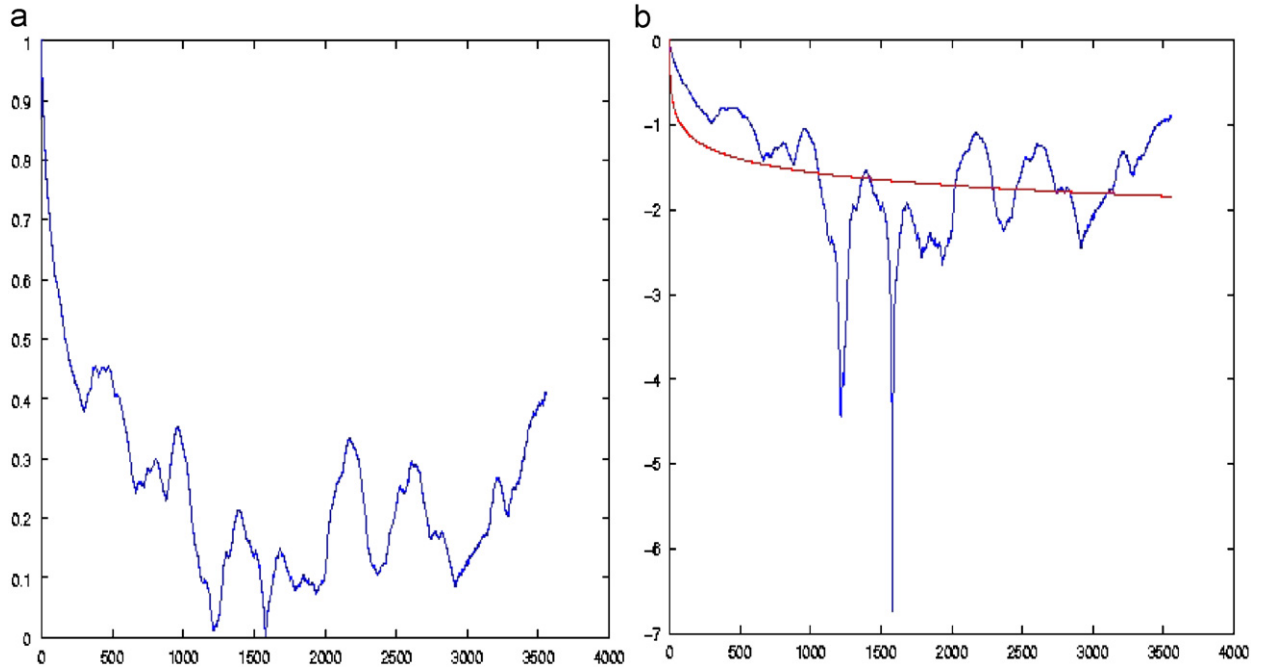

Fig. 7. (a) Correlation function and (b) logarithm of the correlation function and the regression's curve $\hat{\delta}_n = -\hat{\alpha}\log(n+1)$ got for $\hat{\alpha} = 0.22$.

$\gamma = (\gamma(n))$ of the training data corresponding to this area and writing $\delta_n = \log(\gamma(n)) = -\alpha\log(n+1) + \varepsilon_n$, we estimate α by the mean square method. We find $\hat{\alpha} = 0.22$ and we verify from the graphics presented in Fig. 7 that the curve $\hat{\delta}_n = -\hat{\alpha}\log(n+1)$ is very close to $\log(\gamma(n))$ which is characteristic of long-memory dependence. We notice that the estimate $\hat{\alpha} = 0.22$ found here is quite different from the estimate $\hat{\alpha} = 0.02$ found by ICE; however, this is not contradictory because in the Hilbert–Peano scan two points close to each other in the two-dimensional set of pixels can be far from each other obtained in the one-dimensional chain.

6. Conclusion

In this paper, we dealt with the problem of segmentation of a discrete signal hidden by a long-memory noise. Using the general notion of triplet Markov chains (TMCs), we proposed a new general model in which the distribution of the hidden signal is the marginal distribution of a Markov chain, and in which the observations are Gaussian. We showed that the posterior marginal distributions of the hidden signal are computable and thus the classical Bayesian "Maximum of Posterior Marginals" (MPM) segmentation method is computable. Moreover, we proposed an original parameter estimation method inspired by the general "Iterative Conditional Estimation" (ICE) principle, and showed its good appropriateness, when unsupervised MPM segmentation is concerned, to the new model. We provided some experiments which show the interest of the new model and related processing with respect to the classical hidden Markov chain with independent noise (HMC-IN). In fact, when data suit the classical HMC-IN the new model gives comparable results; but when data suit the new model, the results provided by HMC-IN can be much worse than those provided by the new model.

Therefore, the use of new model and the associated parameter estimation method can be considered in every area mentioned in the Introduction, in which the classical HMC-IN already finds interesting applications. Otherwise, the proposed model can also be seen as an approximation of non-stationary fractional Gaussian noise, widely used in different "long-memory" phenomena [29]. Hence the unsupervised segmentation method proposed can be used immediately for unsupervised search of hidden switching states.

As perspective for further work, we can mention the possibility of using marginal distributions of the noise different from Gaussian ones, as suggested in

[25]. This leads to the introduction of Gaussian copulas in the model considered in the paper, which have already provided interesting results in the classical hidden Markov chains model [14].

References

[1] O. Cappé, E. Moulines, T. Ryden, Inference in Hidden Markov Models, Springer Series in Statistics, Berlin, 2005.

[2] Y. Ephraim, Hidden Markov processes, IEEE Trans. Inform. Theory 48 (6) (2002) 1518–1569.

[3] T. Koski, Hidden Markov Models for Bioinformatics, Kluwer Academic Publishers, Netherlands, 2001.

[4] C. Raphael, Automatic segmentation of acoustic musical signals using hidden Markov models, IEEE Trans. Pattern Anal. Mach. 21 (4) (1999) 360–370.

[5] P. Nicolas, L. Bize, F. Muri-Majoube, M. Hoebeke, F. Rodolphe, S. Dusko Ehrlich, B. Prum, P. Bessières, Mining *Bacillus subtilis* chromosome heterogeneities using hidden Markov models, Nucleic Acid Res. 30 (2002) 1418–1426.

[6] E. Bellone, J.P. Hughes, P. Guttorp, A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts, Climate Res. 15 (1) (2000) 1–12.

[7] V. Fortin, L. Perreault, J.D. Salas, Retrospective analysis and forecasting of streamflows using a shifting level model, J. Hydrol. 296 (1–4) (2004) 135–163.

[8] F. Le Ber, M. Benoît, C. Schott, J.-F. Mari, C. Mignolet, Studying crop sequences with CarrotAge, a HMM-based data mining software, Ecol. Model. 191 (1) (2006) 170–185.

[9] D. Hernandez-Hernandez, S.I. Marcus, P.J. Fard, Analysis of a risk-sensitive control problem for hidden Markov chains, IEEE Trans. Autom. Control 44 (5) (1999) 1093.

[10] E. Costamagna, L. Favalli, P. Gamba, Multipath channel modeling with chaotic attractors, Proc. IEEE 90 (5) (2002) 842–859.

[11] H. Kong, E. Shwedyk, Sequence detection and channel state emission over finite state Markov channels, IEEE Trans. Veh. Technol. 48 (3) (1999) 833.

[12] S. Gregoir, F. Lenglart, Measuring the probability of a business cycle turning point by using a multivariate qualitative hidden Markov model, J. Forecast. 19 (2) (2000) 81.

[13] L.C. Thomas, D.E. Allen, N. Morkel-Kingsbury, A hidden Markov chain model for the term structure of bond credit risk spreads, Int. Rev. Finan. Anal. 11 (3) (2002) 311–329.

[14] N. Brunel, W. Pieczynski, Unsupervised signal restoration using hidden Markov chains with copulas, Signal Processing 85 (12) (2005) 2304–2315.

[15] C. Carincotte, S. Derrode, S. Bourennane, Unsupervised change detection on SAR images using fuzzy hidden Markov chains, IEEE Trans. Geosci. Remote Sens. 44 (2) (2006) 432–441.

[16] R. Fjørtoft, Y. Delignon, W. Pieczynski, M. Sigelle, F. Tupin, Unsupervised segmentation of radar images using hidden Markov chains and hidden Markov random fields, IEEE Trans. Geosci. Remote Sens. 41 (3) (2003) 675–686.

[17] N. Giordana, W. Pieczynski, Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation, IEEE Trans. Pattern Anal. Mach. 19 (5) (1997) 465–475.

[18] G. Mercier, S. Derrode, M. Lennon, Hyperspectral image segmentation with Markov chain model, in: Proceedings of the IGARSS 2003, vol. 6, 2003, pp. 3766–3768.

[19] S. Derrode, W. Pieczynski, Signal and image segmentation using pairwise Markov chains, IEEE Trans. Signal Process. 52 (9) (2004) 2477–2489.

[20] W. Pieczynski, C. Hulard, T. Veit, Triplet Markov chains in hidden signal restoration, SPIE's International Symposium on Remote Sensing, Crete, Greece, September 22–27, 2002.

[21] B. Ait-el-Fquih, F. Desbouvries, Kalman filtering in triplet Markov chains, IEEE Trans. Signal Process. 54 (8) (2006) 2957–2963.

[22] S. Bricq, C. Collet, J.-P. Armspach, Triplet Markov chain for 3D MRI brain segmentation using a probabilistic atlas, in: Proceedings of the IEEE International Symposium on Biomedical Imaging, ISBI'06, Arlington, VA, USA, April 6–9, 2006.

[23] W. Pieczynski, F. Desbouvries, On triplet Markov chains, in: Proceedings of the International Symposium on Applied Stochastic Models and Data Analysis, ASMDA 2005, Brest, France, May 2005.

[24] W. Pieczynski, Triplet partially Markov chains and trees, in: Proceedings of the Second International Symposium on Image/Video Communications Over Fixed and Mobile Networks (ISIVC'04), Brest, France, July 2004.

[25] W. Pieczynski, Copules gaussiennes dans les chaînes triplet partiellement de Markov, C. R. Acad. Sci.-Math. 341 (3) (2005) 189–194.

[26] J. Beran, M.S. Taqqu, Statistics for Long-Memory processes Monographs on Statistics and Applied Probability, Chapman & Hall, New York, 1994.

[27] O. Cappé, E. Moulines, J.-C. Pesquet, A. Petropulu, X. Yang, Longe-range dependence and heavy-tail modeling for teletraffic data, IEEE Signal Process. Mag. 19 (3) (2002) 14–27.

[28] F. Chapeau-Blondeau, M. Guglielmi, Modèles de signaux à longue dépendance statistique, in: M. Guglielmi (Ed.), Signaux Aléatoires: Modélisation, Estimation, Détection, Traité IC2, Hermes, Paris, 2004 (Chapter 4).

[29] P. Doukhan, G. Oppenheim, M.S. Taqqu, Theory and Applications of Long-Range Dependence, Birkhauser, Basel, 2003.

[30] W. Pieczynski, P. Lanchantin, Restoring hidden nonstationary process using triplet partially Markov chain with long memory noise, in: Statistical Signal Processing (SSP2005), Bordeaux, France, July 2005.

[31] D. Benboudjema, W. Pieczynski, Unsupervised image segmentation using triplet Markov fields, Comput. Vision Image Understand. 99 (3) (2005) 476–498.

[32] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, W. Pieczynski, Multisensor images segmentation using Dempster–Shafer fusion in Markov fields context, IEEE Trans. Geosci. Remote Sens. 39 (8) (2001) 1789–1798.

[33] F. Salzenstein, W. Pieczynski, Parameter estimation in hidden fuzzy Markov random fields and image segmentation, Graph. Models Image Process. 59 (4) (1997) 205–220.

[34] M. Chen, A. Kundu, J. Zhou, Off-Line handwritten word recognition using a hidden Markov model type stochastic network, IEEE Trans. Pattern Anal. Mach. Intell. 16 (5) (1994).

[35] P. Lanchantin, W. Pieczynski, Unsupervised nonstationary image segmentation using triplet Markov chains, in:

Advanced Concepts for Intelligent Vision Systems (ACVIS 04), Brussels, Belgium, August 31–September 3, 2004.

[36] J. Lapuyade-Lahorgue, W. Pieczynski, Unsupervised segmentation of hidden semi-Markov nonstationary chains, in: Proceedings of the 26th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, MaxEnt 2006, Paris, France, July 8–13, 2006.

[37] J. Lapuyade-Lahorgue, W. Pieczynski, Partially Markov models and unsupervised segmentation of semi-Markov chains hidden with long dependence noise, in: Proceedings of the International Symposium on Applied Stochastic Models and Data Analysis, ASMDA 2007, Chania, Greece, May 2007.

[38] W. Pieczynski, Pairwise Markov chains, IEEE Trans. Pattern Anal. Mach. 25 (5) (2003) 634–639.

[39] W. Pieczynski, Multisensor triplet Markov chains and theory of evidence, Int. J. Approx. Reason. 45 (1) (2007) 1–16.

[40] C.R. Rao, Linear statistical inference and its applications, Wiley Series in Probability and Mathematical Statistics, Wiley, New York, 1973.

[41] P. Lanchantin, W. Pieczynski, Unsupervised restoration of hidden non stationary Markov chain using evidential priors, IEEE Trans. Signal Process. 53 (8) (2005) 3091–3098.

[42] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extension, Wiley, New York, 1997.

[43] W. Pieczynski, Statistical image segmentation, Mach. Graph. Vision 1 (1/2) (1992) 261–268.

[44] J.-P. Delmas, An equivalence of the EM and ICE algorithm for exponential family, IEEE Trans. Signal Process. 45 (10) (1997) 2613–2615.

[45] W. Pieczynski, Convergence of the iterative conditional estimation and application to mixture proportion identification, in: IEEE Statistical Signal Processing Workshop, SSP 2007, Madison, WI, USA, August 26–29, 2007.