Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Signal Processing 92 (2012) 29–42

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Unsupervised segmentation of hidden semi-Markov non-stationary chains

Jérôme Lapuyade-Lahorgue, Wojciech Pieczynski *

Institut Telecom, Telecom Sudparis, Département CITI, CNRS UMR 5157, 9 rue Charles Fourier, 91000 Evry, France

ARTICLE INFO

Article history:
Received 9 September 2010
Received in revised form
30 May 2011
Accepted 5 June 2011
Available online 12 June 2011

Keywords:
Non-stationary hidden semi-Markov chain
Unsupervised segmentation
Iterative conditional estimation
Triplet Markov chains

ABSTRACT

The Bayesian segmentation using Hidden Markov Chains (HMC) is widely used in various domains such as speech recognition, acoustics, biosciences, climatology, text recognition, automatic translation and image processing. On the one hand, hidden semi-Markov chains (HSMC), which extend HMC, have turned out to be of interest in many situations and have improved HMC-based results. On the other hand, the case of non-stationary data can pose an important problem in real-life situations, especially when the model parameters have to be estimated. The aim of this paper is to consider these two extensions simultaneously: we propose using a particular triplet Markov chain (TMC) to deal with non-stationary hidden semi-Markov chains. In addition, we consider a recent particular HSMC having the same computation complexity as the classical HMC. We propose a related parameter estimation method and the resulting unsupervised Bayesian segmentation is validated through experiments; in particular, a real radar image segmentations are provided.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

In the classical "hidden Markov chain" (HMC) model there is a hidden chain $X=(X_1,\ldots,X_N)$ (which will be written X or X_1^N, if there is a risk of confusion) and an observed chain $Y=(Y_1,\ldots,Y_N)$. The hidden chain X is a Markov one, the random variables Y_n are independent conditionally on X, and they verify $p(y_n|x)=p(y_n|x_n)$ for each $n=1,\ldots,N$. HMC is widely used and is quite efficient in numerous situations. In particular, it has been applied in biosciences [27,37], climatology [3], ecology [21,33], control [24], communications [12,26], econometrics and finance [23,49], handwriting and text recognition [11], image processing and computer vision [20,22,35], processing musical signals [46], speech recognition [19,34], or general signal processing [9]. We only cite one or two recent publications in each area, each of them containing a rich bibliography. Moreover, a rich bibliography on classical HMC can also be found in Refs. [9,17,27].

In the case of the discrete hidden chain which we deal with in this paper, these classical models admit two following extensions.

The first one is the hidden semi-Markov chains, in which the hidden process is a semi-Markov chain. The advantage of semi-Markov chains over Markov chains is that in the latter the sojourn time in a given state is necessarily of exponential form, while it is of any form in the former. Such extensions are of interest in different problems, such as the segmentation of medical images [18], speech reconstruction [38], mobility tracking [51], or anomaly detection [48]. Also see recent books or papers [1,2,8,25,39,53,54] containing rich bibliographies on the subject.

The second one consists of taking into account the possibility of the presence of a finite number of different stationarities [29,30,47]. This means that there is an underlying "switching" process, which governs the random changes of the model parameters.

The aim of this paper is to propose a model which is able to take into account these two generalizations simultaneously. We exploit the fact that each of these generalizations can be modeled by particular "triplet Markov

* Corresponding author.
E-mail address: Wojciech.Pieczynski@it-sudparis.eu (W. Pieczynski).

0165-1684/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2011.06.001

Author's personal copy

chains" (TMC) [43,44]. In TMC one adds to the chains X and Y a third chain U, and one assumes that the triplet $T=(X, U, Y)$ is a Markov chain. Thus there is a TMC $T^1=(X, U^1, Y)$ such that the distribution of (X,Y), which is the marginal distribution of T^1, is a hidden semi-Markov chain distribution. Also, there is a TMC $T^2=(X, U^2, Y)$ such that U^2 models the non-stationarity of (X, Y). More precisely, U^2 takes its values in a finite set of states, and each state models a given set of parameters defining a given distribution of (X,Y) [28,29]. In order to exploit these generalizations simultaneously, we will consider a TMC $T=(X, U^1, U^2, Y)=(X, U, Y)$, with $U=(U^1, U^2)$, where U^1 models the semi-Markovianity of X, and U^2 models the different stationarities of (X,Y) [30]. Therefore we have a stationary TMC $T=(X,U,Y)$ which models a non-stationary HSMC (HSMC-NS). We propose to use such a TMC in unsupervised hidden discrete signal segmentation. Estimating the parameters is not a particular problem; in fact, as HSMC-NS is seen as a particular "Pairwise Markov chain" (PMC), the "Iterative Conditional Estimation" (ICE) method already successfully used in Ref. [14] can be adapted to the proposed model. This leads to unsupervised Bayesian segmentation by the classical Maximum Posterior Mode (MPM) method. The interest of the new modeling and related processing is validated by some experiments.

The organization of the paper is the following. The PMC and TMC are recalled in the next section. The third section is devoted to the new model we propose, and the related parameter estimation algorithm is introduced in section four. Different experiments are provided in section five, and the last section contains conclusions and perspectives.

2. Pairwise and triplet Markov chains

Let $X=(X_1,\dots,X_N)$ be a stochastic chain, each $X_1,\dots,X_N$ taking its values in $\Omega=\{\omega_1,...,\omega_K\}$, and let $Y=(Y_1,\dots, Y_N)$ be a real valued process. The aim of Bayesian segmentation is to estimate the hidden realization x of X from an observed realization y of Y. One possible way of such an estimation is to set, for $n=1,\dots,N$, $x_n^*=\mathrm{argmax}_{x_n} p(x_n|y)$. This estimation, which will be used in the whole paper, is called "Marginal Posterior Mode" (MPM) estimation. The estimate $x^*=(x_1^*, ..., x_N^*)$ is computable, even for very large N, when the distribution $p(x,y)$ is not too complex. The most classical model allowing the calculation of x^* is the "hidden Markov chain" model (HMC) in which $p(x,y)$ is given by

$$p(x,y)=p(x_1)\prod_{n=1}^{N-1} p(x_{n+1}|x_n)\prod_{n=1}^{N} p(y_n|x_n). \tag{2.1}$$

The hidden chain X is then a Markov chain, the random variables Y_n are independent conditionally on X, and they verify $p(y_n|x)=p(y_n|x_n)$ for each $n=1,\dots,N$. HMC is widely used because of its efficiency in general; however, the simplicity of $p(y|x)$ is open to criticism in some contexts. In fact, the Markovianity of X assumed in HMC involves that $p(y_n|x)=p(y_n|x_n)$ for stationary invertible chains [29,43], which makes it impossible to take into account a possibly different distribution of the noise on frontiers. A more complete model is the so-called "pairwise Markov chain" (PMC [14,42]), whose distribution is given by

$$p(x,y)=p(x_1,y_1)\prod_{n=1}^{N-1} p(x_{n+1},y_{n+1}|x_n,y_n). \tag{2.2}$$

Then $Z=(X,Y)$ is a Markov chain, and the same is true for $p(y|x)$ and $p(x|y)$. The former property allows one to better model the "noise distribution" $p(y|x)$, while the latter still allows one to calculate the MPM solution x^*. Moreover, setting $p(x_{n+1},y_{n+1}|x_n,y_n)=p(x_{n+1}|x_n,y_n)\ p(y_{n+1}|x_n,y_n,x_{n+1})$, we see that a PMC is an HMC if $p(x_{n+1}|x_n,y_n)=p(x_{n+1}|x_n)$ and $p(y_{n+1}|x_n,y_n,x_{n+1})=p(y_{n+1}|x_{n+1})$. This shows, at a "local" level, how much larger the modeling possibilities of PMC are. Moreover, if $p(x_{n+1}|x_n,y_n)=p(x_{n+1}|x_n)$ holds, we can show that X is a Markov chain [42] (remember that X is not necessarily Markovian in PMC) and we obtain an HMC in which the random variables Y_n are no longer necessarily independent conditionally on X. Let us remember that different experimental studies presented in Ref. [14] show that this larger generality of PMC over HMC can result in greater efficiency when unsupervised data segmentation is concerned: the error ratio can be divided by two.

Considering a triplet Markov chain (TMC) consists of introducing a third stochastic process $U=(U_1,\dots,U_N)$, with each U_n taking its values in a finite set $\Lambda=\{\lambda_1,\dots,\lambda_M\}$, such that $T=(X,U,Y)=((X_n,U_n,Y_n))_{1\le n\le N}$ is a Markov chain. To simplify, let us introduce $V=(V_n)_{1\le n\le N}=(X_n,U_n)_{1\le n\le N}$. Therefore each V_n takes its values in $\Omega\times\Lambda$, and (V,Y) is a PMC. Let us underline the fact that only $T=(X,U,Y)$ is assumed to be Markov and thus no one of the six chains X, U, Y, (X,U), (U,Y), (X,Y) is necessarily Markovian [29,43].

Let us specify how the posterior marginal distributions $p(x_n|y)$ can be computed in the TMC model above. The chain (V,Y) being a PMC, we can introduce the "Forward" probabilities $F_n(v_n)=p(v_n,y_1,\dots,y_n)$, and the "Backward" probabilities $B_n(v_n)=p(y_{n+1},\dots,y_N|v_n,y_n)$, which both extend the classical probabilities used in the HMC model, and which can be calculated by the following forward and backward recursions (see [14,42])

$$F_1(v_1)=p(v_1,y_1);\ F_{n+1}(v_{n+1})=\sum_{v_n\in\Omega\times\Lambda} F_n(v_n)p(v_{n+1},y_{n+1}|v_n,y_n); \tag{2.3}$$

$$B_N(v_N)=1;\ B_n(v_n)=\sum_{v_{n+1}\in\Omega\times\Lambda} B_{n+1}(v_{n+1})p(v_{n+1},y_{n+1}|v_n,y_n). \tag{2.4}$$

Then we have

$$p(v_n,y)=F_n(v_n)B_n(v_n); \tag{2.5}$$

$$p(v_n,v_{n+1},y)=F_n(v_n)p(v_{n+1},y_{n+1}|v_n,y_n)B_{n+1}(v_{n+1}), \tag{2.6}$$

which gives $p(v_n|y)$ and $p(v_{n+1}|v_n,y)$. Recalling that $p(v_n|y)=p(x_n,u_n|y)$, the posterior marginal distribution $p(x_n|y)=\sum_{u_n} p(x_n,u_n|y)$ can then be used in MPM segmentation, and $p(u_n|y)=\sum_{x_n} p(x_n,u_n|y)$ can be used in searching $U=u$ which can be, in some situations, of interest. For example, when using "triplet Markov fields" in image segmentation, the field U can be used to model different stationarities – or different "textures" – and thus

Author's personal copy

searching $U=u$ is equivalent to searching textures, which is an important problem in image processing [4].

Remark 2.1. Let us underline the fact that, in the general case, U models the non-stationarity of the couple (X,Y). However, two following particular cases remain possible. In the first one, the only X is concerned, which means that $p(y|x,u)=p(y|x)$. In the second one, the only Y is concerned, which means that $p(y|x,u)=p(y|x)$. In all papers published until now [4,28–30], and in the present paper, one considers the first case; however, the other two cases could be dealt with as well.

3. Hidden non-stationary semi-Markov chains as a particular TMC

3.1. Classical hidden semi-Markov chains (HSMC) as TMC

Let $X=(X_1,\ldots,X_N)$ be a stochastic chain taking its values in Ω^N, with $\Omega=\{\omega_1,\ldots,\omega_K\}$, and let $Y=(Y_1,\ldots,Y_N)$ be a real valued process as above. X is said to be a semi-Markov chain if its distribution is defined in the following way. Let $p(x_1)$ be a distribution on Ω, and let $p^*(x_{n+1}|x_n)$ be transitions, that we will assume independent from n, such that $x_{n+1}=x_n$ implies $p^*(x_{n+1}|x_n)=0$. Otherwise, for each $\omega_k\in\Omega$ let p_{ω_k} be the distribution of the remaining "duration" in the state ω_k. Thus knowing that $X_{n-1}\neq\omega_k$ and $X_n=\omega_k$, $p_{\omega_k}(m)$ is the probability that $X_{n+i}=\omega_k$ for each $i=0,\ldots,i=m$, and $X_{n+m+1}\neq\omega_k$. In a Markov chain the duration distribution is exponential; in fact, we have $p_{\omega_k}(m)=(\lambda_{kk})^m(1-\lambda_{kk})$, with $\lambda_{kk}=p(x_{n+1}=\omega_k|x_n=\omega_k)$. In semi-Markov chains, which are also known as "variable duration HMMs", it is of any form. Then the transitions $p^*(x_{n+1}|x_n)$ and the distributions p_{ω_k} define the distribution of $X=(X_1,\ldots,X_N)$. For example, for $K=2$, $N=6$, and $x=(\omega_1,\ \omega_1,\ \omega_2,\ \omega_2,\ \omega_2,\ \omega_2,\ \omega_1)$, we have $p(x)=p(\omega_1)\,p_{\omega_1}(1)p^*(\omega_2|\omega_1)p_{\omega_2}(3)p^*(\omega_1|\omega_2)$. In fact, the probability of $X_1=\omega_1$ is $p(\omega_1)$. Then we see, according to $x=(\omega_1,\ \omega_1,\ \omega_2,\ \omega_2,\ \omega_2,\ \omega_2,\ \omega_1)$, that the duration of X_1 in the state ω_1 is 1; the probability that this occurs is thus $p_{\omega_1}(1)$. Once this duration has passed, X_3 must be different from ω_1; its realization follows $p^*(.|\omega_1)$. Thus the probability for $X_1=\omega_1$, $X_2=\omega_1$, and $X_3=\omega_2$ is $p(\omega_1)p_{\omega_1}(1)p^*(\omega_2|\omega_1)$. Continuing like this we find $p(x)=p(\omega_1)p_{\omega_1}(1)p^*(\omega_2|\omega_1)p_{\omega_2}(3)\,p^*(\omega_1|\omega_2)$.

Having $p(x)$, the distribution of (X,Y) is then defined by

$$p(y|x)=\prod_{n=1}^{N}p(y_n|x_n).$$

Such a HSMC can then be seen as a TMC $T=(X,U,Y)$, where $U_n=u_n$ is the duration in the state $X_n=x_n$. The distribution of the Markov chain (X,U) is given by $p(x_1,u_1)$ and the transitions $p(x_{n+1},\ u_{n+1}|x_n,\ u_n)=p(x_{n+1}|x_n,\ u_n)\,p(u_{n+1}|x_{n+1},x_n,u_n)$. Introducing the Kronecker function δ, where $\delta_a(b)=1$ for $a=b$ and $\delta_a(b)=0$ for $a\neq b$, we can write (also see Ref. [52])

$$\begin{aligned}&p(x_{n+1}|x_n,u_n)=\delta_{x_n}(x_{n+1})\text{ if }u_n>0,\\&\text{and }p(x_{n+1}|x_n,u_n)=p^*(x_{n+1}|x_n)\text{ if }u_n=0\end{aligned}\tag{3.1}$$

$$\begin{aligned}&p(u_{n+1}|x_{n+1},x_n,u_n)=\delta_{u_n-1}(u_{n+1})\text{ if }u_n>0,\\&\text{and }p(u_{n+1}|x_{n+1},x_n,u_n)=p_{x_{n+1}}(u_{n+1})\text{ if }u_n=0.\end{aligned}\tag{3.2}$$

Let us notice that $p(u_{n+1}|x_{n+1},x_n,u_n)=\delta_{u_n-1}(u_{n+1})$ if $u_n>0$ simply means that if at n the remaining time of staying in x_n is $u_n>0$, then at $n+1$ this remaining time is $u_{n+1}=u_n-1$. This could also be written $p(u_{n+1}=u_n-1\,|x_{n+1},x_n,u_n>0)=1$.

In the classical HSMC considered here, these transitions are completed by

$$p(y_{n+1}|x_{n+1},u_{n+1},x_n,u_n,y_n)=p(y_{n+1}|x_{n+1}).\tag{3.3}$$

Finally, once we have assumed that $p(x_{n+1}|x_n,u_n,y_n)=p(x_{n+1}|x_n,u_n)$ and $p(u_{n+1}|x_{n+1},x_n,u_n,y_n)=p(u_{n+1}|x_{n+1},x_n,u_n)$ (see Remark 3.1 below for the general form) to obtain the classical HSMC, Eqs. (3.1)–(3.3) define the distribution of $T=(X,U,Y)$.

Besides, setting $V=(X,U)$, we can also say that $T=(V,Y)$ is the very classical HMC with the particular property $p(y_n\,|v_n)=p(y_n|x_n,u_n)=p(y_n|x_n)$. Finally, this particular HMC also is a PMC and thus (2.3)–(2.5), in which

$$p(v_{n+1},y_{n+1}\,|v_n,y_n)=p(v_{n+1}|v_n)p(y_{n+1}|x_{n+1}),\tag{3.4}$$

can be used to compute $p(v_n|y)$, $p(v_{n+1}|v_n,y)$, $p(x_n|y)$, and $p(u_n|y)$.

Let us notice that the formulation (3.1)–(3.3) of the HSMC is not new, as a very close formulation is presented in Ref. [50]. However, using Eqs. (2.3)–(2.5) with Eq. (3.4), which finally means using a particular very classical HMC in the classical way originally proposed in Ref. [19], is, to our knowledge, original and leads to a simpler computation of $p(x_n|y)$ than in Ref. [50].

Remark 3.1. Let us consider a general TMC $T=(X,U,Y)=(T_1,\ldots,T_N)$, with $T_n=(X_n,U_n,Y_n)$ for each $n=1,\ldots,N$. Its transitions in their most general form can be written $p(t_{n+1}|t_n)=p(x_{n+1}|x_n,\ u_n,\ y_n)p(u_{n+1}|x_n,\ u_n,y_n,\ x_{n+1})p(y_{n+1}\,|x_n,\ u_n,\ y_n,\ x_{n+1},\ u_{n+1})$. In this general case, the chain $T=(V,Y)$, with $V=(X,U)$, is a PMC and thus (2.3)–(2.5) can be applied. In the classical HSMC above we have $p(x_{n+1}\,|x_n,u_n,y_n)=p(x_{n+1}|x_n,u_n)$, $p(u_{n+1}|x_n,u_n,y_n,x_{n+1})=p(u_{n+1}\,|x_n,\ u_n)$, and $p(y_{n+1}|x_n,\ u_n,\ y_n,\ x_{n+1},\ u_{n+1})=p(y_{n+1}|x_{n+1})$, but of course, these simplifications are not essential. Eq. (3.1) can be easily extended to a more general distribution in which $p^*(x_{n+1}|x_n,y_n)$ depends on y_n. In a similar way, in Eq. (3.2) $p(u_{n+1}|x_n,\ u_n,\ y_n,\ x_{n+1})$ can depend on (x_n,u_n,y_n). Finally, $p(y_{n+1}|x_{n+1})$ in Eq. (3.3) can be extended to the general $p(y_{n+1}|x_{n+1},\ u_{n+1},\ x_n,\ u_n,\ y_n)$. We see that while using the general "TMC" approach to model the classical HSMC, it is easy to propose a "generalized hidden semi-Markov chain", in which neither is the chain $V=(X,\ U)$ Markovian, nor is the chain X semi-Markovian. However, we will keep the classical HSMC (3.1)–(3.3) in this paper because our aim is to study how to extend it to the non-stationary case, not to study how to extend it to a more general "HSMC".

3.2. Non-stationary classical HSMC (NS-HSMC)

Until now semi-Markovianity has been achieved, now this will be combined with non-stationarity. To introduce the latter, let us first consider the classical HSMC, considered as a TMC (X,U^1,Y), introduced above. Setting $V=(X,U^1)$, we can consider the HMC (V,Y), where Y is

Author's personal copy

observed and V is hidden. Then it is possible to consider different stationarities of (V,Y), which is modeled by introducing a third random chain $U^2=(U_1^2,\ldots,U_N^2)$, each U_n^2 taking its values in a finite set $\Lambda^2=\{1,\ldots,L\}$, as proposed in Refs. [28,29]. More precisely, one considers a stationary TMC (V,U^2,Y), where $p(v,y|u^2)$ is a non-stationary Markovian distribution. Such modeling gave good results in the classical HMC [28,29], and thus we extend them here to HSMC $(X,U^1,Y)=(V,Y)$. Doing so, we arrive at $(V,U^2,Y)=(X,U^1,U^2,Y)$, which is a stationary TMC $T=(X,U,Y)$, with $U=(U^1,U^2)$. The transitions $p(t_{n+1}|t_n)$ in such model can be defined in many different ways. In order to extend Eqs. (3.1)–(3.3) in a consistent manner, we propose to consider the following factorization of the transition $p(t_{n+1}|t_n)$ (remember that $t_n=(x_n,u_n^1,u_n^2,y_n)$, and similarly for t_{n+1})

$$\begin{aligned}p(t_{n+1}|t_n)&=p(x_{n+1},u_{n+1}^1,u_{n+1}^2,y_{n+1}|t_n)\\&=p(u_{n+1}^2|t_n)p(x_{n+1}|t_n,u_{n+1}^2)p(u_{n+1}^1|t_n,u_{n+1}^2,x_{n+1})\\&\quad\times p(y_{n+1}|t_n,u_{n+1}^2,x_{n+1},u_{n+1}^1)\end{aligned}\tag{3.5}$$

Then we propose the following extension of the classical HSMC defined with Eqs. (3.1)–(3.4). We keep $p(u_{n+1}^2|t_n)=p(u_{n+1}^2|x_n,u_n^1,u_n^2,y_n)$ in Eq. (3.5) in its general form, which is simply recalled by Eq. (3.6). Then $p(x_{n+1}|t_n,u_{n+1}^2)$ and $p(u_{n+1}^1|t_n,u_{n+1}^2,x_{n+1})$ in Eq. (3.5) become Eqs. (3.7) and (3.8). Thus Eq. (3.7) is simultaneously a simplification $p(x_{n+1}|t_n,u_{n+1}^2)=p(x_{n+1}|x_n,u_n^1,u_{n+1}^2)$ with respect to Eq. (3.5) and an extension of Eq. (3.1). Similarly, Eq. (3.8) is simultaneously a simplification $p(u_{n+1}^1|t_n,u_{n+1}^2,x_{n+1})=p(u_{n+1}^1|u_n^1,x_{n+1},u_{n+1}^2)$ with respect to Eq. (3.5), and an extension of Eq. (3.2). Finally, $p(y_{n+1}|t_n,u_{n+1}^2,x_{n+1},u_{n+1}^1)$ could be simplified to $p(y_{n+1}|x_{n+1})$ as it is done in the classical model; however, there is no reason for this and it can be kept in its general form, which is simply recalled in Eq. (3.9)

$$p(u_{n+1}^2|t_n)=p(u_{n+1}^2|x_n,u_n^1,u_n^2,y_n)\tag{3.6}$$

$$p(x_{n+1}|t_n,u_{n+1}^2)=p(x_{n+1}|x_n,u_n^1,u_{n+1}^2)=\delta_{x_n}(x_{n+1})\text{ if }u_n^1>0,\quad\text{and}\quad p^*(x_{n+1}|x_n,u_{n+1}^2)\text{ if }u_n^1=0;\tag{3.7}$$

$$\begin{aligned}p(u_{n+1}^1|t_n,u_{n+1}^2,x_{n+1})&=p(u_{n+1}^1|u_n^1,x_{n+1},u_{n+1}^2)\\&=\delta_{u_n^1-1}(u_{n+1}^1)\text{ if }u_n^1>0\\\text{and}\quad p(u_{n+1}^1|t_n,u_{n+1}^2,x_{n+1})&=p(u_{n+1}^1|x_{n+1},u_{n+1}^2)\\&=p_{x_{n+1},u_{n+1}^2}(u_{n+1}^1)\text{ if }u_n^1=0;\end{aligned}\tag{3.8}$$

$$\begin{aligned}&p(y_{n+1}|t_n,u_{n+1}^2,x_{n+1},u_{n+1}^1)\\&\quad=p(y_{n+1}|x_n,u_n^1,u_n^2,y_n,u_{n+1}^2,x_{n+1},u_{n+1}^1).\end{aligned}\tag{3.9}$$

Eqs. (3.5)–(3.9) thus define an extension of the classical HSMC to the non-stationary case. In particular, $p^*(x_{n+1}|x_n,u_{n+1}^2)$ in Eq. (3.7) verifies $p^*(x_{n+1}|x_n,u_{n+1}^2)=0$ for $x_{n+1}=x_n$.

Remark 3.2. Eqs. (3.5)–(3.9) extend the model presented in Ref. [30]; let us notice that similar ideas have been independently proposed in Refs. [16,41], where a switching hidden semi-Markov model (S-HSMM) has been proposed and studied. However, the two models are different in that the switches are Markovian in S-HSMM, while they are semi-Markovian in the model proposed in Ref. [30]. Then we can observe that the model (3.5)–(3.9) extends both of them. In fact, taking in Eq. (3.6) $p(u_{n+1}^2|t_n)=p(u_{n+1}^2|u_n^2)$ we obtain Markovian switches.

3.3. TMC-based definition of a recent HSMC (RHSMC)

In the HSMC defined in sub-section 3.2 above the sojourn time is not bounded, and thus, we are not in the exact case of a TMC where U is finite. However, as the number N of observations used is finite, we can consider that each U_n takes its values in a set with N elements. This gives the computation complexity as polynomial in time.

Let us consider the following recent model, which is a particular HSMC and which allows the computation complexity to be linear in time [31,32]. The classical model above is defined by the transitions $p^*(x_{n+1}|x_n)$ such that $x_{n+1}=x_n$ implies $p^*(x_{n+1}|x_n)=0$, and by the distributions $p_{\omega_k}(u_{n+1})=p(u_{n+1}|x_{n+1}=\omega_k)$ on the set of natural numbers N. Let us modify two things. First, for $x_{n+1}=x_n$ the transition $p^*(x_{n+1}|x_n)$ is no longer necessarily null and becomes a transition of any form. Such new transitions will be denoted with $q(x_{n+1}|x_n)$. Second, the distributions $p_{\omega_k}(u_{n+1})$ are defined on a finite set $\Lambda=\{0,\ldots,P-1\}$, with P independent from N. Such new distributions will be denoted with $q_{\omega_k}(u_{n+1})=q(u_{n+1}|x_{n+1}=\omega_k)$. Such a model, proposed in Ref. [31], will be called a "recent" SMC (RSMC). Thus the couple (X,U) is Markovian finite, each (X_n,U_n) taking its values from $\{\omega_1,\ldots,\omega_K\}\times\{0,\ldots,P-1\}$. Its difference from the classical SMC is that in SMC $U_n=u_n$ is the exact sojourn time in $X_n=x_n$, while in RSMC it is the minimal sojourn time in $X_n=x_n$. The transitions $p(x_{n+1},u_{n+1}|x_n,u_n)$ of an RSMC are defined by Eqs. (3.1), (3.2) with the difference that for $x_{n+1}=x_n$ the transition $p^*(x_{n+1}|x_n)$ in Eq. (3.1), which is here $q(x_{n+1}|x_n)$, is not necessarily null, and $p_{x_{n+1}}(u_{n+1})$ in Eq. (3.2) is replaced with $q_{x_{n+1}}(u_{n+1})$.

Concerning the position of the RSMC with regard to the classical SMC let us notice two points: (i) the classical SMC is a Markov chain when the sojourn time is exponential, while the RSMC is a Markov chain when the minimal sojourn time is null with probability one: $p(u_n=0|x_n)=1$ for each $n=1,\ldots,N,\ldots$ and each $x_n\in\{\omega_1,\ldots,\omega_K\}$. In this case the transition matrices $q(x_{n+1}|x_n)$ related to the RSMC are the transition matrices related to the MC; (ii) as shown in Ref. [32], the distribution of an RSMC can be seen as a distribution of an SMC where the sojourn time distributions are not known explicitly. More precisely, let us consider a RSMC defined by the distributions $q(u_{n+1}|x_{n+1})$ on $\Lambda=\{0,\ldots,P-1\}$, and the transitions $q(x_{n+1}|x_n)$. Let $p^*(x_{n+1}|x_n)$ be transitions obtained from $q(x_{n+1}|x_n)$ by $p^*(x_{n+1}|x_n)=0$ for $x_n=x_{n+1}$, and

$$p^*(x_{n+1}|x_n)=\frac{q(x_{n+1}|x_n)}{1-q(x_{n+1}=x_n|x_n)},$$

and let $p^*(u_{n+1}|x_{n+1})$ be the distributions on N defined by

$$p^*(u_{n+1}=d|x_{n+1}=\omega_j)=q(x_{n+1+d}\neq\omega_j|x_n=\omega_j)$$

Author's personal copy

$$\times \sum_{k=1}^{d}\Big[\sum\nolimits_{s_1+\ldots+s_j=d}[q(u_{n+1}=s_1|x_{n+1}=\omega_j)$$
$$\times q(u_{n+1+s_1}=s_2|x_{n+1+s_1}=\omega_j)$$
$$\times q(u_{n+1+s_1+s_2}=s_3|x_{n+1+s_1+s_2}=\omega_j)\ldots$$
$$\times q(u_{n+1+s_1+\ldots+s_{k-1}}=s_k|x_{n+1+s_1+\ldots+s_{k-1}}=\omega_j)$$
$$\times q(x_{n+1+s_1}=\omega_j|x_{n+1}=\omega_j)$$
$$\times q(x_{n+1+s_1+s_2}=\omega_j|x_{n+1}=\omega_j)\ldots$$
$$\times q(x_{n+1+s_1+\ldots+s_{k-1}}=\omega_j|x_{n+1}=\omega_j)]\Big] \quad (3.10)$$

Finally, RHSMCs are particular HSMCs with the particularity of offering the possibility of Bayesian classification with complexity linear in time.

3.4. Non-stationary recent HSMC (NS-RHSMC)

Here we propose two extensions, of growing generality, of the RHSMCs to the non-stationary model. The first one is similar to the model given by Eqs. (3.6)–(3.9): these equations are kept with the difference that $p(x_{n+1}|u^2_{n+1}, x_n)$ is not necessarily null for $x_{n+1}=x_n$ (i.e., $p^*(x_{n+1}|x_n, u^2_{n+1})$ in Eq. (3.7) is replaced with $q(x_{n+1}|x_n, u^2_{n+1})$), and that U^1_i takes its values from a finite set $\Lambda^1=\{0,1,\ldots,P-1\}$ (*i.e.*, $p(u^1_{n+1}|x_{n+1}, u^2_{n+1})$ in Eq. (3.8) is replaced with $p(u^1_{n+1}|x_{n+1}, u^2_{n+1}) = q_{x_{n+1}, u^2_{n+1}}(u^1_{n+1})$). The second one is more general: Eqs. (3.6) and (3.9) are kept, and Eqs. (3.7), (3.8) are extended to Eqs. (3.11), (3.12) below. Of course, setting $V=(X, U^1, U^2)$, the NS-RHSMC $T=(V,Y)$ defined by Eqs. (3.6), (3.11), (3.12), (3.9) is a pairwise Markov chain with $p(v_{n+1},y_{n+1}|v_n,y_n)=p(v_{n+1}|v_n,y_n)p(y_{n+1}|v_n,y_n,v_{n+1})$, where $p(v_{n+1}|v_n,y_n)$ is defined with Eqs. (3.6), (3.11), (3.12), and $p(y_{n+1}|v_n,y_n,v_{n+1})$ is defined with Eq. (3.9). Therefore we have a classical PMC $T=(V,Y)$, with V discrete and finite, and thus Eqs. (2.3)–(2.6) can be applied. For K classes and L stationarities, such a general NS-RHSMC is thus defined by KL probabilities on Ω conditional on $y_n\in R$ – see Eq. (3.11) – , and K^2L probabilities on Λ^1 conditional on $y_n\in R$—see Eq. (3.12). The motivation of such an extension is the following. Adding the conditioning upon y_n in Eqs. (3.11) and (3.12) means that in the PMC $T=(V,Y)$ we do not necessarily have $p(v_{n+1}|v_n,y_n)=p(v_{n+1}|v_n)$. This means that V is not necessarily a Markov chain, and thus $T=(V,Y)$ is not necessarily a "hidden Markov chain". Now, it has been shown in Ref. [14] that extending the classical HMC to PMC can be of great interest in unsupervised data segmentation; thus it seems to be quite justified to conjecture that it will be of interest in the semi-Markov non-stationary context considered.

$$p(x_{n+1}|x_n,u^1_n,y_n,u^2_{n+1})=\delta_{x_n}(x_{n+1}) \text{ if } u^1_n>0,$$
$$\text{and} \quad q(x_{n+1}|x_n,u^2_{n+1},y_n) \text{ if } u^1_n=0; \quad (3.11)$$

$$p(u^1_{n+1}|x_n,u^1_n,y_n,x_{n+1},u^2_{n+1})=\delta_{u^1_n-1}(u^1_{n+1}) \text{ if } u^1_n>0,$$
$$\text{and} \quad p(u^1_{n+1}|x_n,u^1_n,y_n,x_{n+1},u^2_{n+1})$$
$$=q(u^1_{n+1}|x_n,y_n,x_{n+1},u^2_{n+1}) \text{ if } u^1_n=0; \quad (3.12)$$

We will specify some particular simple NS-RHSMC models dealing with some experiments in the next section.

4. Parameter estimation

The model parameter estimation from the only observed data $Y=y$ is the core point when wishing to perform unsupervised segmentation. The aim of this section is to describe the use of the general "Iterative Conditional Estimation" (ICE) method, which gave conclusive results in similar applications [5,20,22,30] and, in particular, which has already been successfully applied in unsupervised segmentation of PMCs in Ref. [14]. As NS-RHSMCs are particular PMCs, here we adapt the ICE used in Ref. [14] to the considered context. However, as specified in Remark 4.2 below, in the Gaussian case considered in the next section the well-known "Expectation-Maximization" (EM) algorithm could be used as well.

4.1. Iterative conditional estimation (ICE)

Let us consider two random processes (V,Y) whose distribution depends on a vector of parameters $\theta=(\theta_1,\ldots,\theta_m)$. The problem is to estimate θ from Y. The ICE method we will use is based on the following principle. Let $\hat{\theta}(v, y)$ be an estimator of θ from complete data $(V,Y)=(v,y)$ and let us assume that we can sample realizations of V according to $p(v|y)$. The ICE sequence is obtained as follows :

(i) Initialize θ^0;
(ii) compute $\theta^{q+1}_i=E[\hat{\theta}_i(V, Y)|Y=y, \theta^q)$ for the components θ_i for which this computation can be carried on explicitly;
(iii) if there are components θ_i for which the computation above is not feasible, simulate $v^q_1,\ldots,v^q_l$ according to $p(v|y, \theta^q)$ and set for each such component θ_i

$$\theta^{q+1}_i=\frac{[\hat{\theta}_i(v^q_1,y)+\ldots+\hat{\theta}_i(v^q_l,y)]}{l}.$$

We see that the ICE is applicable under two very mild hypotheses:

(H1) existence of an estimator $\hat{\theta}(v, y)$ from the complete data, and

(H2) the ability of simulating V according to $p(v|y)$.

The first hypothesis is not really a constraint because if we are not able to estimate θ from the complete data (v, y), there is no point in searching for an estimator from incomplete ones y. When (V, Y) is a Markov chain, the posterior distribution $p(v|y)$ also provides a Markov chain: it allows us easily to derive V according to the second hypothesis.

Remark 4.1. Concerning the parameter estimation in the case of incomplete data, the most known and used method is the so-called "Expectation-Maximization" (EM) method, whose aim is to iteratively maximize the likelihood $p(y|\theta)$ according to the principle $\theta^{q+1}=\underset{\theta}{\text{argmax}}\, E_{\theta^q}[Log(p_\theta(V, Y))|Y=y]$. In the context of Gaussian NS-RHSMC considered in this paper, which resembles, as far as the parameter estimation problem is concerned, the very classical HMC, EM could be applied in an analogous

Author's personal copy

way as ICE is. However, as we use ICE, let us recall some general properties of both methods:

(i) ICE is more general than EM, and can be easier to perform in complex situations. It is more general because the estimator $\hat{\theta}(v, y)$ can be of any form; in particular, it can be the "maximum likelihood" (ML) estimator or not. Thus, when the ML estimator does not exist, ICE can possibly still be used, when EM cannot. Moreover, it can be easier to perform because the maximization step in the equality defining θ^{q+1} above does not exist in ICE;

(ii) as stated in Ref. [13], in the case of exponential models – and under the condition that $\hat{\theta}(v, y)$ used in ICE is the ML estimator – EM and ICE can produce the same sequence (θ^q);

(iii) many comparisons between EM and ICE have been performed in classical contexts with Gaussian noise, like adaptive estimation [40], hidden Markov chains [5], or hidden Markov trees [36]. In all these situations, the EM formulae are computable and it turns out that both EM and ICE methods are of quite comparable efficiency;

(iv) the use of EM is justified by the theoretical results concerning the optimal behavior of the ML estimator, and by the fact that EM produces a sequence (θ^q) such that the sequence $p(y|\theta^q)$, being increasing, often converges to a local maximum. We have to notice that this does not imply the convergence of (θ^q) to the real parameter θ; however, if the initial value θ^0 is close enough to the real value θ, the convergence can be shown under some mild hypotheses. The idea behind ICE is different and is based on the following. Assuming that $\hat{\theta}(v, y)$ has an interesting quadratic error – or is even optimal, being, for example, an ML estimator in an exponential model – one wishes to approximate it by a function of the only observed variables y. The "best" – with regard to the same "quadratic error" criterion – approximation is the conditional expectation. As this expectation depends on the parameter, we arrive at the point (ii) in the definition of ICE above. Concerning the convergence of ICE, let us mention a recent theoretical result obtained in the case of independent data [45]. As in the case of EM, convergence can be obtained under some reasonable hypotheses if the initial value θ^0 is close enough to the real value θ;

(v) EM encounters more difficulties in hidden Markov field models, where the maximization step cannot be calculated and one is obliged to simplify the model, for example by introducing the "mean field" as indicated in Ref. [10]. ICE can be used without model modification, even in more complex situations, as in the context of recent triplet Markov fields [4].

4.2. ICE in NS-RHSMC

Let us consider NS-RHSMC $T=(X, U^1, U^2, Y)$ specified in sub-section 3.4: the variables X_n, U_n^1, and U_n^2 take their values from $\Omega=\{\omega_1,\ldots,\omega_K\}$, $\Lambda^1=\{0,\ldots,P-1\}$, and $\Lambda^2=\{1,\ldots,L\}$, respectively. Thus we have a Markov chain $T=(V,Y)$, where $V=(X, U^1, U^2)$, and where each V_n takes its values in $\Omega\times\Lambda^1\times\Lambda^2$, and each Y_n takes its values in R. To estimate the parameters, we will assume that the distribution $p(t_n, t_{n+1})$ does not depend on n, which means that the distribution of $T=(V,Y)$ is defined by the distribution $p(t_1, t_2)$.

Let us consider the following particular distribution $p(t_1,t_2)=p(v_1,v_2)p(y_1|v_1)p(y_2|v_2)=p(x_1,u_1^1,u_1^2,x_2,u_2^1,u_2^2)p(y_1|x_1)p(y_2|x_2)$, which means that the PMC $T=(V,Y)$ is in fact a classical HMC with independent noise. This particular case was chosen on purpose, to facilitate comparisons with other classical models in the experiments below. Let us see what the parameters defining the model are exactly. The vector $(V_1,V_2)=(X_1,U_1^1,U_1^2,X_2,U_2^1,U_2^2)$ takes its values in the finite set $(\Omega\times\Lambda^1\times\Lambda^2)^2$, which contains $(KPL)^2$ elements. However, due to the very definition of the model there are some elements with zero probability. Let $v_1=(x_1=\omega_i,u_1^1=j,u_1^2=k)$. For $u_1^1\neq 0$ we necessarily have $v_2=(x_2=\omega_i,u_2^1=j-1,u_1^2=k)$, and for $u_1^1=0$ the vector $v_2=(x_2,u_2^1,u_1^2)$ is of any possible value in $\Omega\times\Lambda^1\times\Lambda^2$. Therefore $p(v_1, v_2)$ is defined on a subset $A=\{a_1,\ldots,a_A\}$ of $(\Omega\times\Lambda^1\times\Lambda^2)^2$ containing $A=K(P-1)L+K^2L^2P=KL(KLP+P-1)$ elements, and thus $p(v_1, v_2)$ is defined by A parameters $\alpha_1=p(a_1),\ldots,\alpha_A=p(a_A)$. Concerning the K distributions $p(y_1|x_1=\omega_1),\ldots,p(y_1|x_1=\omega_K)$ - remember that $p(y_n|x_n)$ does not depend on n - we assume that they depend on vector parameters $\beta_1,\ldots,\beta_K$, respectively. For example, for Gaussian distributions, $\beta_1,\ldots,\beta_K$ are means and variances $\beta_1=(\mu_1,\sigma_1^2),\ldots,\beta_K=(\mu_K,\sigma_K^2)$. Finally, the parameters to be estimated are $\theta=(\alpha,\ \beta)=(\alpha_1,\ldots,\alpha_A,\ \beta_1,\ldots,\beta_K)$.

Then the ICE method can be used once there is, for each $i=1,\ldots,K$, an estimator $\hat{\beta}_i(y_1^i,\ldots,y_n^i)$ of β_i from a sample $y_1^i,\ldots,y_n^i$ produced according to $p(y_1|x_1=\omega_i)$. In fact, (H1) is verified upon using the following estimator $\hat{\theta}(v, y)$. Considering an even $N=2N'$, the parameters α_i are estimated by the classical empirical estimators

$$\hat{\alpha}_i(v_1,\ldots,v_N)=\frac{1}{N'}\sum_{j=1}^{N'}1_{[(v_{2j-1},v_{2j})=a_i]}, \tag{4.1}$$

and the vectors β_i are estimated by $\hat{\beta}_i(y^i)$, where y^i is the sub-sample of $y=(y_1,\ldots,y_N)$ such that y_j is in y^i if $x_j=\omega_i$. In other words, to estimate $\beta=(\beta_1,\ldots,\beta_K)$ we use (iii) of the ICE description in the previous sub-section, with $l=1$.

As (H2) is verified from the fact that $p(v|y)$ is a Markov chain distribution, the ICE method can be used. In particular, to re-estimate $\alpha=(\alpha_1,\ldots,\alpha_A)$ we see that when using $\hat{\alpha}_i(v)$ defined by Eq. (4.1), $\alpha_i^{q+1}=E[\hat{\alpha}_i(V)|Y=y,\alpha^q)$ is computed by

$$\alpha_i^{q+1}=\frac{1}{N'}\sum_{j=1}^{N'}p((v_{2j-1},v_{2j})=a_i|y,\theta^q), \tag{4.2}$$

where $p((v_{2j-1},v_{2j})=a_i|y,\theta^q)$ are computable from Eq. (2.6). To re-estimate $\beta=(\beta_1,\ldots,\beta_K)$, we use (iii) of the ICE description in the previous sub-section, with $l=1$.

Remark 4.2. We said in Remark 4.1 that the well-known EM method could also be used; let us briefly specify how.

Author's personal copy

The re-estimation of $\alpha=(\alpha_1,\ldots,\alpha_A)$ with Eq. (4.2) would be identical in both methods, and the difference is at the $\beta=(\beta_1,\ldots,\beta_K)$ re-estimation. Let us assume that the noise is Gaussian, and thus we have $\beta_i=(m_i,\sigma_i^2)$ with m_i the mean and σ_i^2 the variance of the Gaussian distribution corresponding to the class ω_i. We have seen that in ICE $\beta_i=(m_i,\sigma_i^2)$ is estimated from the sub-sample y^i of $y=(y_1,\ldots,y_N)$ defined by "y_j is in y^i if $x_j=\omega_i$". In EM each $\beta_i=(m_i,\sigma_i^2)$ would be re-estimated from the whole sample $y=(y_1,\ldots,y_N)$ in the following way. After having computed $p(x_j=\omega_i|y,\theta^q)=\sum_{(u_i^1,u_i^2)}p(x_j=\omega_i,u_i^1,u_i^2|y,\theta^q)]$, we would set

$$m_i^{q+1}=\frac{\sum_{j=1}^N y_j[p(x_j=\omega_i|y,\theta^q)}{\sum_{j=1}^N[p(x_j=\omega_i|y,\theta^q)]},$$
$$(\sigma_i^2)^{q+1}=\frac{\sum_{j=1}^N (y_j-m_i^{q+1})^2[p(x_j=\omega_i|y,\theta^q)]}{\sum_{j=1}^N[p(x_j=\omega_i|y,\theta^q)]}. \quad (4.3)$$

We obtain formulas which extend the classical EM formulas which are well known in the classical hidden Markov chains.

5. Experiments

We present three series of experiments. Their aim is to show that the new model is of interest, when coupled with the related ICE-based estimation method, in the unsupervised segmentation context. Let us recall that the unsupervised segmentation context can be quite different from the supervised, which means that the model parameters are known. In fact, when all parameters are known the new model must be of interest, simply because of its larger generality. However, in unsupervised context things are more complicated. In fact, let us assume that a given set of data perfectly corresponds to the classical HMC. Of course HMC is a RHSMC and thus when the parameters are known both models are strictly the same and thus they will provide the same results. Things are different in the unsupervised context: using HMC directly needs fewer parameters to be estimated than using RHSMC. Therefore, when the parameters have to be estimated, using HMC instead of RHSMC could possibly give significantly better results. To study whether this occurs or not is the very aim of the first series of two experiments presented in sub-section 5.1.

In sub-section 5.2 we study the converse problem. When the data does follow an RHSMC which is not an HMC, can the use of a HMC provide comparable results? In other words, are there situations in which RHSMC presents an interest in the unsupervised segmentation context? As above, the problem is somewhat different in supervised and unsupervised cases. In the former case the RHSMC based results must be better, according to the very Bayesian theory. In unsupervised case this is not necessarily true, especially when the data are very noisy. In fact, if the RHSMC parameters estimation turned out to be less efficient than the HMC parameters estimation, the HMC based segmentation could provide better results that the RHSMC based one. The stronger the noise the higher the risk that such cases occur is; thus we directly study a rather noisy case.

In the third series the hidden data suit neither HMC nor RHSMC and thus it is of interest to test which one between the two models will be more efficient in unsupervised context. As above, the parameter estimation is of importance and thus the simpler HMC model, having fewer parameters to be estimated, could possibly be more efficient than the more complex RHSMC.

Finally, we present some segmentation of a real radar image in sub-section 5.4.

Let us notice that the calculations Eqs. (2.3)–(2.5) can be used when N is not too large; however, when it is, both $F_N(v_n)$ and $B_1(v_1)$ tend to 0, when N tends to infinity, at exponential rate. To remedy this problem we use, as in Refs. [14,15], the following "normalized" forward and backward probabilities

$$\alpha_n(v_n)=p(v_n|y_1,\ldots,y_n), \quad (5.1)$$

$$\beta_n(v_n)=\frac{p(y_{n+1},\ldots,y_N|v_n,y_n)}{p(y_{n+1},\ldots,y_N|y_1,\ldots,y_n)}, \quad (5.2)$$

which can be calculated by the following recursions:

$$\alpha_1(v_1)=p(v_1|y_1);$$
$$\alpha_{n+1}(v_{n+1})=\frac{\sum_{v_n\in\Omega\times\Lambda}\alpha_n(v_n)p(v_{n+1},y_{n+1}|v_n,y_n)}{\sum_{(v_n^*,v_{n+1}^*)\in(\Omega\times\Lambda)^2}\alpha_n(v_n^*)p(v_{n+1}^*,y_{n+1}|v_n^*,y_n)} \quad (5.3)$$

$$\beta_N(v_N)=1;\ \beta_n(v_n)=\frac{\sum_{v_{n+1}\in\Omega\times\Lambda}\beta_{n+1}(v_{n+1})p(v_{n+1},y_{n+1}|v_n,y_n)}{\sum_{(v_n^*,v_{n+1}^*)\in(\Omega\times\Lambda)^2}\alpha_n(v_n^*)p(v_{n+1}^*,y_{n+1}|v_n^*,y_n)}. \quad (5.4)$$

Having calculated $\alpha_n(v_n)$ and $\beta_n(v_n)$, we compute $p(v_n,v_{n+1}|y)$ and $p(v_n|y)$ with

$$p(v_n,v_{n+1}|y)=\frac{\alpha_n(v_n)p(v_{n+1},y_{n+1}|v_n,y_n)\beta_{n+1}(v_{n+1})}{\sum_{(v_n^*,v_{n+1}^*)\in(\Omega\times\Lambda)^2}\alpha_n(v_n^*)p(v_{n+1}^*,y_{n+1}|v_n^*,y_n)\beta_{n+1}(v_{n+1}^*)}, \quad (5.5)$$

$$p(v_n|y)=\alpha_n(v_n)\beta_n(v_n). \quad (5.6)$$

In particular, $p(v_n|y)$ gives $p(x_n|y)=\sum_{u_n\in\Lambda}p(v_n=(x_n,u_n)|y)$ and $p(u_n|y)=\sum_{x_n\in\Omega}p(v_n=(x_n,u_n)|y)$.

Finally, dividing Eq. (5.5) by Eq. (5.6) we obtain

$$p(v_{n+1}|v_n,y)=\frac{p(v_{n+1},y_{n+1}|v_n,y_n)\beta_{n+1}(v_{n+1})}{\beta_n(v_n)\sum_{(v_n^*,v_{n+1}^*)\in(\Omega\times\Lambda)^2}\alpha_n(v_n^*)p(v_{n+1}^*,y_{n+1}|v_n^*,y_n)\beta_{n+1}(v_{n+1}^*)} \quad (5.7)$$

Both $p(v_n,v_{n+1}|y)$ and $p(v_{n+1}|v_n,y)$ will be used in the parameter estimation method discussed in sub-sections 5.1–5.3 below.

5.1. Unsupervised segmentation of simulated hidden Markov and semi-Markov chains

The series of tests considered consists in simulating an HMC and segmenting it in unsupervised manner, by using the HMC or the RHSMC model. As an HMC is a special case of an RHSMC, it is interesting to see whether the unsupervised segmentations based on the RHSMC model are

Author's personal copy

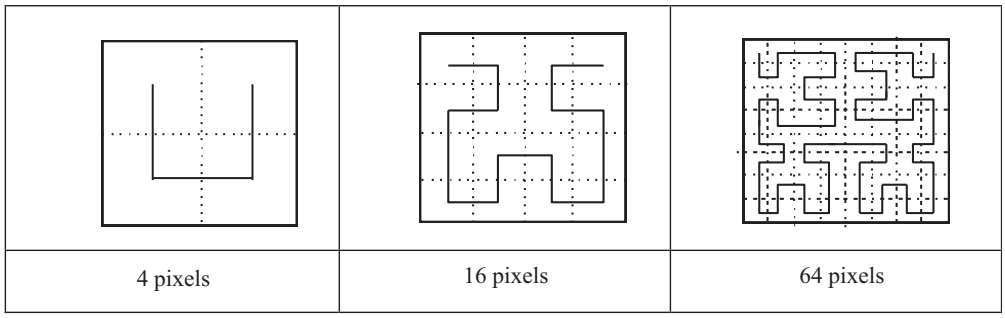

Fig. 1. Construction of a Hilbert–Peano scan used to represent realizations of mono-dimensional sequences as bi-dimensional images.

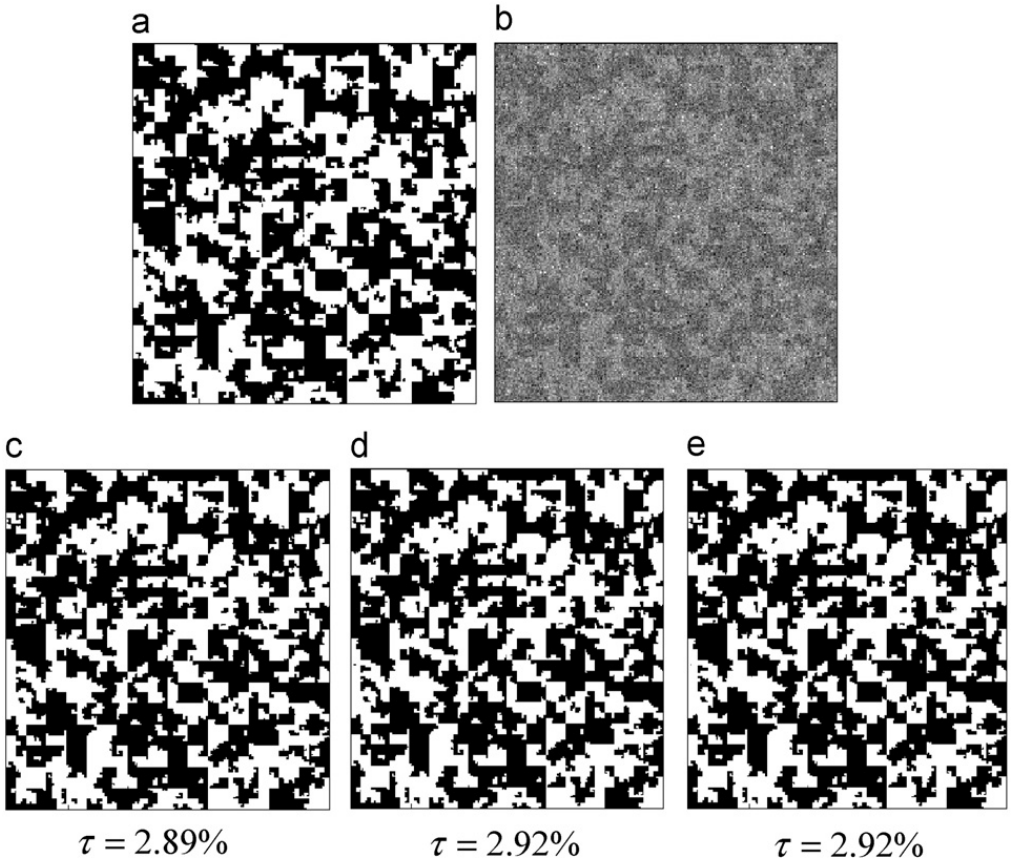

Fig. 2. Unsupervised segmentation of an HMC. From left to right: (a) simulation x of X, (b) simulation y of Y, (c) segmentation of y based on the HMC model and the real parameters used in (a) and (b), (d) unsupervised segmentation of y based on the HMC model and the ICE method, and (e) unsupervised segmentation of y based on the RHSMC models and the ICE method. τ is the proportion of misclassified points.

close or not to those based on the HMC model. We present two series of results, with increasing noise level.

In order to visualize the results, we will use images of size $N=256\times 256$. Such a bi-dimensional set of pixels is transformed into a mono-dimensional set using a Hilbert–Peano scan, as presented in Fig. 1. Such a representation is quite pleasant because it allows one to appreciate visually the degree of the noise, and also the quality difference between two segmentation results. However, let us insist on the fact that in this sub-section and the following one, this is only a representation, and the problem we deal with is the problem of mono-dimensional chains.

More precisely, we simulate a two-class stationary Markov chain X whose distribution is defined by $p(x_1=\omega_1, x_2=\omega_1)=p(x_1=\omega_2, x_2=\omega_2)=0.495$, $p(x_1=\omega_1, x_2=\omega_2)=p(x_1=\omega_2, x_2=\omega_1)=0.005$, and whose realization is presented in Fig. 2(a). The noise distribution $p(y|x)$ is defined by two Gaussian distributions $p(y_i|x_i=\omega_1)\sim N(1,1)$, $p(y_i|x_i=\omega_2)\sim N(2,1)$ and one realization y of Y is presented in Fig. 2(b). In RHSMC used we take $P=10$. Although the noise is rather strong, we see that both HMC and RHSMC based unsupervised methods give identical results, which are very close to the segmentation based on real parameters. This shows that when RHSMC turns out to be a HMC, the ICE method is capable of finding that out, and thus, roughly

Author's personal copy

speaking, when the data suits HMC, there is no risk of using the broader RHSMC model.

The estimates of the noise parameters are given in Table 1, and the estimated $p(x_1,x_2)$ by the ICE method in both the HMC and the RHSMC models are $\hat{p}(x_1=\omega_1, x_2=\omega_1)=p(x_1=\omega_2,x_2=\omega_2)=0.496$, $\hat{p}(x_1=\omega_1,x_2=\omega_2)=p(x_1=\omega_2,x_2=\omega_1)=0.004$, which is very close to the real values.

In order to further challenge the proposed method, we considered an experiment with a lot more noise. Namely, let us consider the same $p(x_1, x_2)$ as above, and $p(y_i|x_i=\omega_1)\sim N(1,10)$, $p(y_i|x_i=\omega_2)\sim N(2,10)$. The results obtained are presented in Fig. 3.

The estimates of the noise parameters are given in Table 2. For the HMC, the estimated $p(x_1, x_2)$ are $\hat{p}(x_1=\omega_2, x_2=\omega_2)=0.495$ and $\hat{p}(x_1=\omega_1, x_2=\omega_2)=\hat{p}(x_1=\omega_2, x_2=\omega_1)=0.005$. The estimates related to RHSMC are

$$\hat{p}(x_2=\omega_j|x_1=\omega_i,u_n^1=0)=\begin{pmatrix}0.84 & 0.16\\ 0.12 & 0.88\end{pmatrix},$$

$$\hat{p}(u_{n+1}^1|x_n=\omega_1,u_n^1=0)=(0.15,0.15,0.13,0.11,0.09,0.08,0.07,0.07,0.07,0.08)$$

and

$$\hat{p}(u_{n+1}^1|x_n=\omega_2,u_n^1=0)=(0.14,0.14,0.12,0.11,0.09,0.08,0.08,0.08,0.08,0.08)$$

According to Fig. 3, we see that the noise is really significant. In fact, the misclassification error is $\tau=17.42\%$ when using the real parameters. The good behavior of the ICE method in the HMC case is confirmed, as the misclassification error $\tau=18.38\%$ is quite close to the error in the supervised case. We also see that using the RHSMC models

Table 1
Noise parameters estimated by the ICE method for the HMC and the RHSMC models in the case in Fig. 2. True parameters are $m_1=1$, $m_2=2$, and $\sigma_1^2=\sigma_2^2=1$.

	HMC		RHSMC	
	ω_1	ω_2	ω_1	ω_2
m	0.99	2.01	0.99	2.01
σ^2	0.99	1.01	1.01	1.01

Table 2
Noise parameters estimated by the ICE method for the HMC and the RHSMC models in the case in Fig. 3. True parameters are $m_1=1$, $m_2=2$, and $\sigma_1^2=\sigma_2^2=10$.

	HMC		RHSMC	
	ω_1	ω_2	ω_1	ω_2
m	1.02	2.01	0.76	2.07
σ^2	10.31	9.67	9.99	9.71

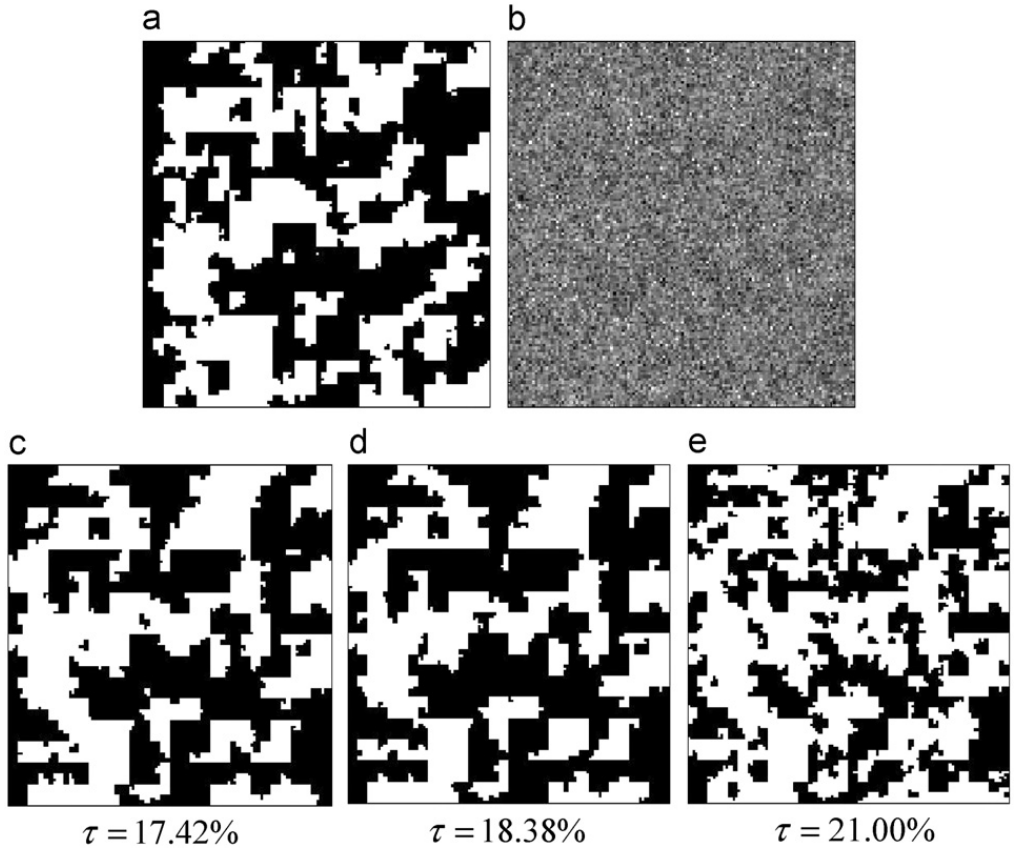

Fig. 3. Unsupervised segmentation of an HMC. From left to right: (a) simulation x of X, (b) simulation y of Y, (c) segmentation of y based on the HMC model and the real parameters used in (a) and (b), (d) unsupervised segmentation of y based on the HMC model and the ICE method, and (e) unsupervised segmentation of y based on the RHSMC models and the ICE method.

Author's personal copy

instead of the HMC model degrades the results, which seems to be mainly due to the estimation of the means. However, we also notice that in order to obtain such degradation, we had to consider a very high noise level, which probably goes beyond cases met in real applications.

5.2. Estimation of a simulated hidden semi-Markov chain

Here, we deal with the converse problem: the data is sampled according to an RHSMC and the observations are segmented, in an unsupervised manner, by both the RHSMC and the HMC-based methods. Of course, the very Bayesian theory implies that when the RHSMC considered is not an HMC, the RHSMC with results based on real parameters must be better than the HMC based on any other parameters; however, it is interesting to see whether the latter are close to the former, especially in an unsupervised segmentation context. In other words, it is interesting to look at what the RHSMC model brings in addition to the HMC model.

Let us consider a two-class RHSMC (X, U^1, Y), where U^1 takes its values in $\Lambda^1=\{0,...,9\}$ and the distribution $p(u^1_{n+1}|x_{n+1},u^1_n=0)$ is uniform for any value x_{n+1} of X_{n+1}. The transitions $p(x_{n+1}|x_n,u^1_n=0)$ are given by $p(x_{n+1}=x_n\ |u^1_n=0)=0.99$, and $p(x_{n+1}\neq x_n|u^1_n=0)=0.01$, and the noise distributions are $p(y_i|x_i=\omega_1)\sim N(1,20)$, $p(y_i|x_i=\omega_2)\sim N(2,20)$. Therefore, to better appreciate the differences among the two models, we directly consider a very noisy case.

The estimated parameters of the RHSMC model are

$$\hat{p}(x_2=\omega_j|x_1=\omega_i,u^1_n=0)=\begin{pmatrix}0.89 & 0.11\\ 0.11 & 0.89\end{pmatrix},$$

$$\hat{p}(u^1_{n+1}|x_n=\omega_1,u^1_n=0)$$
$$=(0.11,0.11,0.11,0.11,0.10,0.10,0.09,0.09,0.09,0.09)$$

and

$$\hat{p}(u^1_{n+1}|x_n=\omega_2,u^1_n=0)$$
$$=(0.11,0.11,0.11,0.11,0.10,0.10,0.09,0.09,0.09,0.09).$$

The corresponding results are presented in Fig. 4.

The estimates of the noise parameters are presented in Table 3. Overall, using the ICE in the HMC or the RHSMC context gives similar results. Also, given the very high level of the noise we note the good behavior of the ICE method.

Table 3
Noise parameters estimated by the ICE method for the HMC and the RHSMC models.

	HMC		RHSMC	
	ω_1	ω_2	ω_1	ω_2
m	0.85	2.48	0.60	2.33
σ^2	19.78	19.01	19.81	20.05

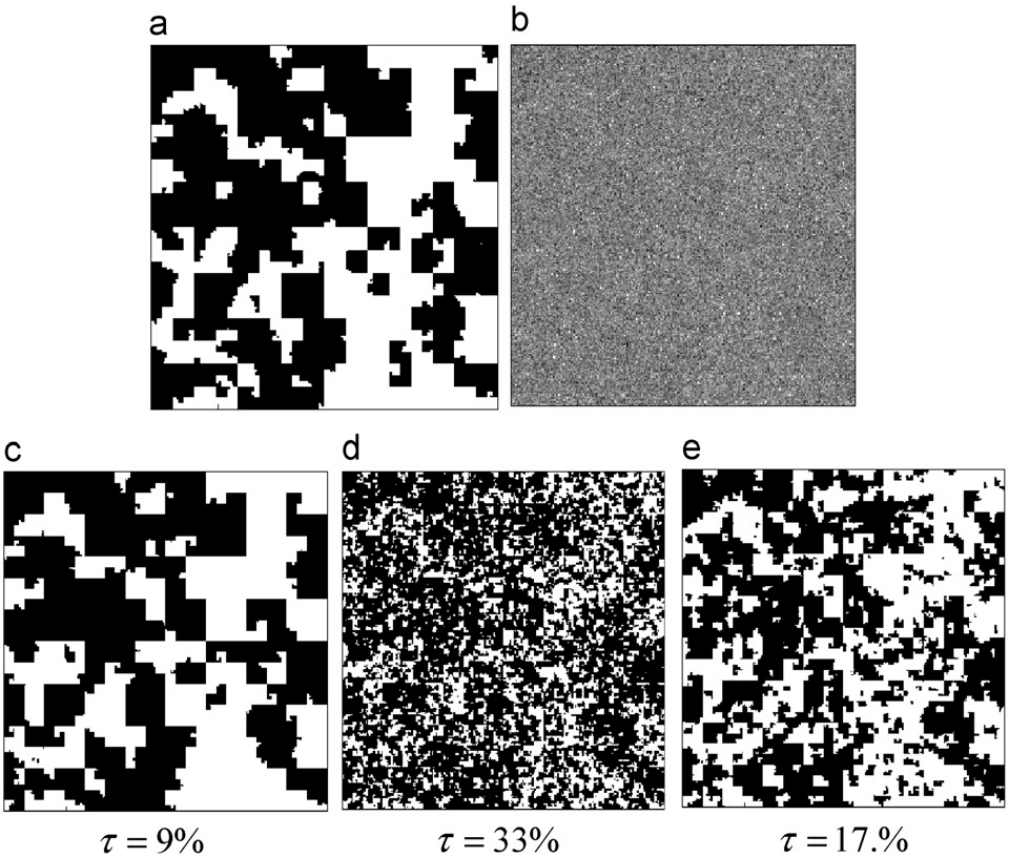

Fig. 4. Unsupervised segmentation of an RHSMC. From left to right: (a) simulation x of X, (b) simulation y of Y, (c) segmentation of y based on the RHSMC model and the real parameters used in (a) and (b), (d) unsupervised segmentation of y based on the HMC model and the ICE method, and (e) unsupervised segmentation of y based on the RHSMC models and the ICE method. τ is the proportion of misclassified points.

Author's personal copy

5.3. Segmentation of a "hand-made" non-stationary picture

In this sub-section we present some results concerning the new NS-RHSMC model. The presented results are of importance in real applications because the data considered are non-stationary and they are obtained neither with an NS-HMC model nor with an NS-RHSMC one.

The simplified NS-HSMC used here and in sub-section 5.4 is the model considered in sub-section 4.2:

$$p(x_{n+1}|x_n,u_n^1,y_n,u_{n+1}^2)=\delta_{x_n}(x_{n+1}) \text{ if } u_n^1>0,$$
$$\text{and } p^*(x_{n+1}|x_n,u_{n+1}^2) \text{ if } u_n^1=0; \tag{5.8}$$

$$p(u_{n+1}^1|x_n,u_n^1,y_n,x_{n+1},u_{n+1}^2)=\delta_{u_n^1-1}(u_{n+1}^1) \text{ if } u_n^1>0,$$
$$\text{and}$$
$$p(u_{n+1}^1|x_n,u_n^1,y_n,x_{n+1},u_{n+1}^2)=p(u_{n+1}^1|x_{n+1},u_{n+1}^2) \text{ if } u_n^1=0 \tag{5.9}$$

$$p(u_{n+1}^2|t_n)=p(u_{n+1}^2|u_n^2). \tag{5.10}$$

$$p(y_{n+1}|t_n,u_{n+1}^2,x_{n+1},u_{n+1}^1)=p(y_{n+1}|x_{n+1}). \tag{5.11}$$

The realization $X=x$ is obtained by considering a collage of two different textures and the result is presented in Fig. 5(a). The observed data are obtained by using two Gaussian distributions $p(y_i|x_i=\omega_1)\sim N(1,2)$, $p(y_i|x_i=\omega_2)\sim N(2,2)$ and we can see, according to Fig. 5, that such a noise is rather strong. Then $Y=y$ is segmented in three unsupervised ways, based on three models with increasing generality: HMC, NS-HMC, and NS-RHSMC. Thus the important point here is to study whether NS-RHSMC can improve results obtained with NS-HMC, recalling that the results obtained with both NS-HMC and NS-RHSMC are of two kinds: recover the classes $X=x$, and recover the "textures" $U^2=u^2$. Therefore, we consider an HMC (X,Y), an NS-HMC (X,U^2,Y), and an NS-RHSMC (X, U^1, U^2, Y). In both NS-HMC (X,U^2,Y) and NS-RHSMC

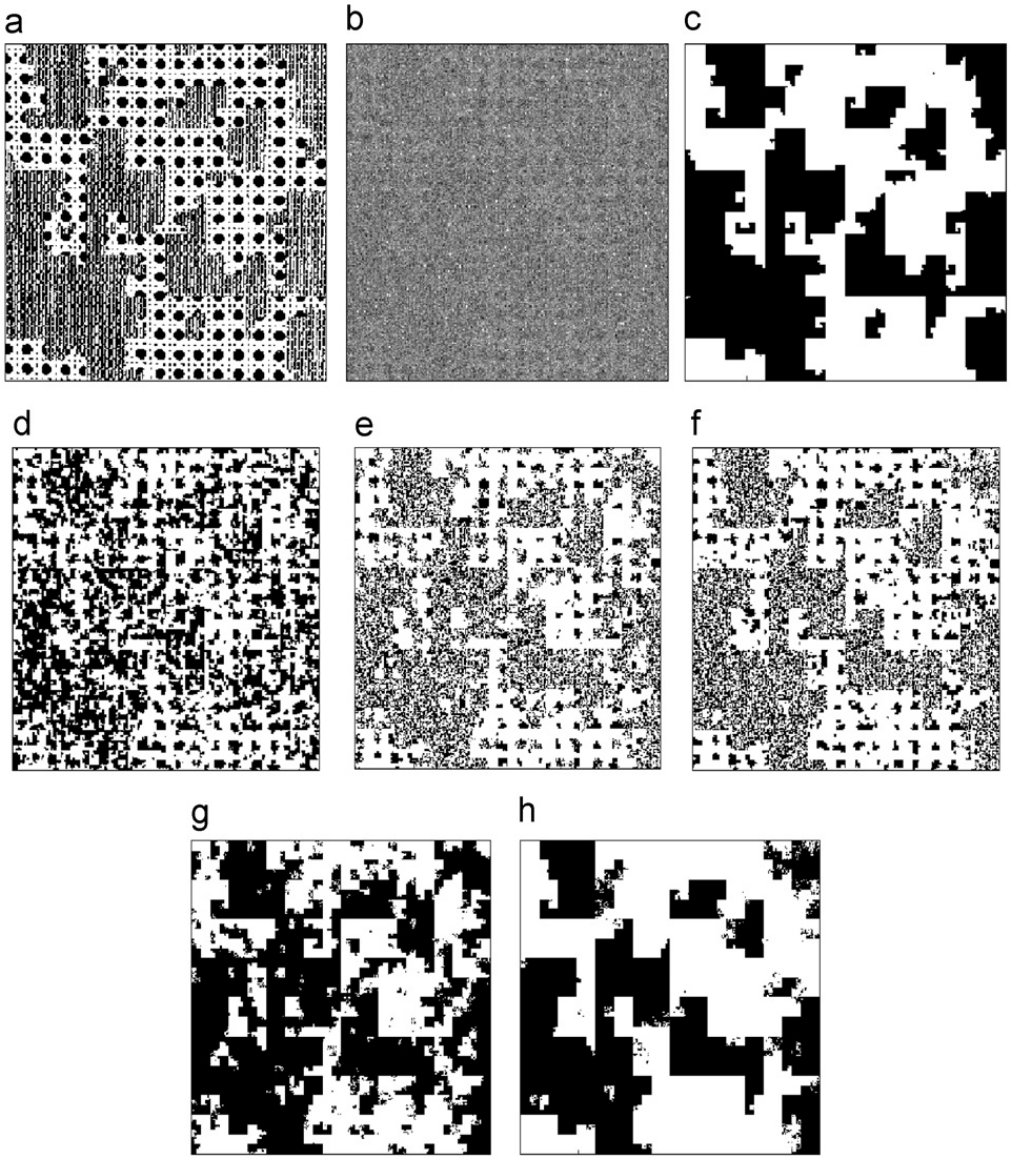

Fig. 5. Estimation of X by using the HMC, NS-HMC, and NS-RHSMC models. Estimation of $U^2=u^2$ by using NS-HMC and NS-RHSMC. Parameters estimated with the ICE method, τ is the misclassification error: (a) hand-made $X=x$, (b) observed $Y=y$, (c) texture chain $U^2=u^2$, (d) HMC segmentation τ=32%, (e) NS-HMC segmentation τ=30%, (f) NS-RHSMC segmentation τ=25%, (g) NS-HMC estimation of u : τ=18%, (h) NS-RHSMC estimation of u : τ=9%.

Author's personal copy

(X, U^1, U^2, Y) the two different textures are considered as being modeled by a two-value random chain U^2, and the semi-Markovianity in (X, U^1, U^2, Y) is modeled by U^1, which will take its values in $\Lambda^1=\{0,\ldots,4\}$. In all cases considered, the parameters are estimated by ICE described in sub-section 4.2.

Table 4
Estimated parameters in the case of the HMC, NS-HMC and NS-RHSMC models. The real parameters are $m_1=1$, $m_2=2$, and $\sigma^2=2$ for both classes ω_1, ω_2.

	HMC		NS-HMC		NS-RHSMC	
	ω_1	ω_2	ω_1	ω_1	ω_1	ω_2
m	1.18	1.89	0.89	1.93	0.96	1.89
σ^2	2.07	2.19	1.92	2.05	2.01	2.08

According to Fig. 5, we can see that NS-RHSMC-based results are significantly better than the NS-HMC-based ones. Concerning the misclassification error, they are $\tau=25\%$ and $\tau=30\%$, respectively. What is more, the difference between NS-RHSMC and NS-HMC is larger than the difference between NS-HMC and HMC. The difference in texture classification efficiency between NS-RHSMC and NS-HMC is still more significant, the error ratio being $\tau=9\%$ and $\tau=18\%$, respectively. Also, we show in Table 4 the noise estimates by the three ICE algorithms used and it turns out that the results obtained in both NS-RHSMC and NS-HMC cases are good and similar. This means, on the one hand, that the efficiency difference between NS-RHSMC and

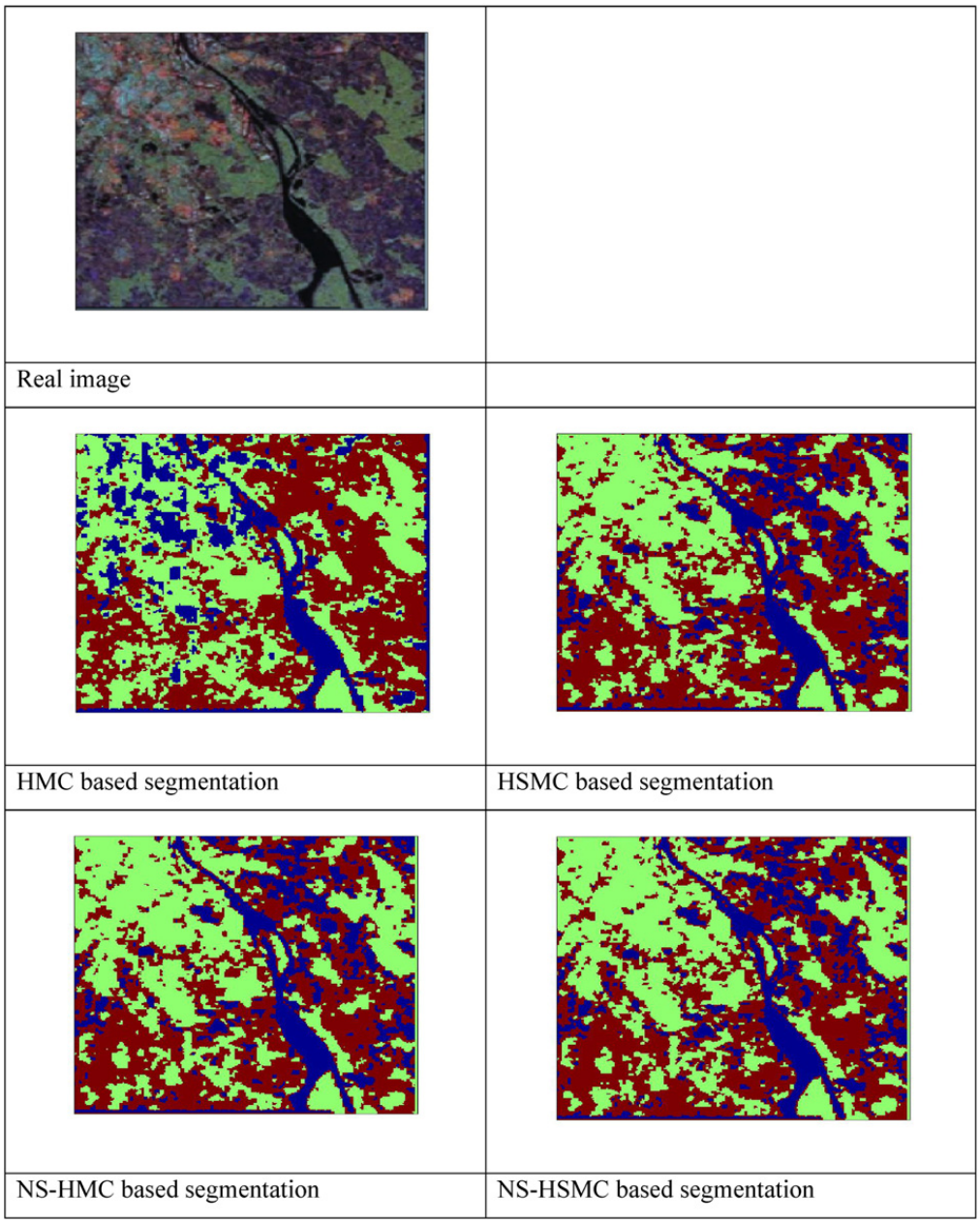

Fig. 6. Three unsupervised segmentations of a real SAR airborne image of the «Rhone Valley», France. ω_1 (blue) is the river, ω_2 (green) is the vegetation, and ω_3 (brown) is what remains.

Author's personal copy

NS-HMC is due to the superiority of the former over the latter, and not to some difference in parameter estimates. On the other hand, this also means that the ICE works well in both NS-RHSMC and NSHMC cases.

Remark 5.1. We showed above that the proposed new model can be applied in image segmentation. Let us notice that where statistical image segmentation is concerned, the Markov model which is mostly used is the hidden Markov field (HMF) model. However, as shown in the classical hidden Markov chains and hidden Markov fields models, the former can be useful with respect to the latter. In some situations they may be of comparable efficiency. In some others, they can be used to initialize the HMF-based methods [20].

5.4. Real image segmentation

Finally, let us consider a real radar SAR airborne image. This image, seen as a realization $Y=y$, is presented in Fig. 6. The size here is 256×256 pixels, and the set of pixels is converted into mono-dimensional sequence using the Hilbert–Peano scan. The observed sequence is then segmented by four methods based on HMC, NS-HMC, RHSMC and NS-RHSMC, respectively. We consider three classes "water", "vegetation", and "others". For NS-HMC and NS-RHSMC, we considered two different stationarities and $\Lambda^1=\{0,...,9\}$. The results obtained are presented in Fig. 6.

As we have no ground truth it is difficult to draw general conclusions. However, both HSMC and NS-HMC based segmentations seem clearly of better quality than the HMC based one, mainly because of the false presence of "water" in the upper left corner of the HMC-based segmentation. Thus each of these two models is of interest when replacing the classical HMC. However, it does not clearly appears, in the example considered, that these advantages have cumulative interest. In fact, the difference between the HSMC based segmentation and the NS-HSMC based seems negligible, and it is still true when comparing the NS-HMC based segmentation and the NS-HSMC based one.

6. Conclusion

We proposed in this paper a new non-stationary hidden semi-Markov chain model and a related parameter estimation method. Both of them have been applied to unsupervised Bayesian signal segmentation and the interest of the whole method has been validated through different experiments. As the new model can be applied in any area where the classical hidden Markov chains are of interest, its applicative possibilities are extremely wide and, in particular, contain different domains mentioned in the introduction.

As perspective, let us mention the possibility of extending the Gaussian noise model considered here to a noise with marginal distributions of any form. In fact, the general method based on Copulas presented in Ref. [7] – and extended to the non-stationary case in Ref. [29] – can be adapted to the proposed model and the good behavior of the parameter estimation method let us envisage its possible extension to more complex situations. The use of a 3-dimensional Hilbert–Peano scan, as introduced in Ref. [6], opens the way to different investigations related to 3-dimensional – or spatio-temporal – data segmentation.

References

[1] M. Azimi, P. Nasiopoulos, R.K. Ward, Offline and online identification of hidden semi-Markov models, IEEE Transactions on Signal Processing 53 (8) (2005) 2658–2663.

[2] V.S. Barbu, N. Limnios, Semi-Markov chains and hidden semi-Markov models toward applications, Lectures Notes in Statistics, Springer, 2008.

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

[4] D. Benboudjema, W. Pieczynski, Unsupervised statistical segmentation of non stationary images using triplet Markov fields, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (8) (2007) 1367–1378.

[5] B. Benmiloud, W. Pieczynski, Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images, Traitement du Signal 12 (5) (1995) 433–454.

[6] 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, Virginia, USA, April 6–9, 2006.

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

[8] Y. Guedon, Estimating hidden semi-Markov chains from discrete sequences, Journal of Computational and Graphical Statistics 12 (3) (2003) 195–224.

[9] O. Cappé, E. Moulines, T. Ryden, Inference in hidden Markov models, Springer, Series in Statistics, Springer, 2005.

[10] G. Celeux, F. Forbes, N. Peyrard, EM procedures using mean field-like approximations for Markov model-based segmentation, Pattern Recognition 36 (2006) 131–144.

[11] M. Chen, A. Kundu, J. Zhou, Off-Line handwritten work recognition using a hidden Markov model type stochastic network, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (5) (1994).

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

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

[14] S. Derrode, W. Pieczynski, Signal and image segmentation using pairwise Markov chains, IEEE Transactions on Signal Processing 52 (9) (2004) 2477–2489.

[15] P.A. Devijver, Baum's forward-backward algorithm revisited, Pattern Recognition Letters 3 (1985) 369–373.

[16] T.V. Duong, H.H. Bui, D.Q. Phung, S. Venkatesh, Activity recognition and abnormality detection with the switching hidden semi-Markov model, in: Proceedings of the Conference on Computer Vision and Pattern Recognition, (CVPR 2005), vol. 1, pp. 838–845, 20–25 June 2005.

[17] Y. Ephraim, Hidden Markov processes, IEEE Transactions on Information Theory 48 (6) (2002) 1518–1569.

[18] S. Faisan, L. Thoraval, J.-P. Armspach, M.-N. Metz-Lutz, F. Heitz, Unsupervised learning and mapping of active brain functional MRI signals based on hidden semi-Markov event sequence models, IEEE Transactions on Medical Imaging 24 (2) (2005) 263–276.

[19] J.D. Ferguson, Variable duration models for speech, in: Proceedings of the Symposium on Application of Hidden Markov Models to Text and Speech, Institute for Defense Analysis, Princeton, NJ, October 1980, pp. 143–179.

[20] 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 Transactions on Geoscience and Remote Sensing 41 (3) (2003) 675–686.

[21] V. Fortin, L. Perreault, J.D. Salas, Retrospective analysis and forecasting of streamflows using a shifting level model, Journal of Hydrology 296 (1–4) (2004) 135–163 20.

Author's personal copy

[22] N. Giordana, W. Pieczynski, Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation, IEEE Transactions on Pattern Analysis and Machine 19 (5) (1997) 465–475.
[23] S. Gregoir, F. Lenglart, Measuring the probability of a business cycle turning point by using a multivariate qualitative hidden Markov model, Journal of Forecasting 19 (2) (2000) 81.
[24] D. Hernandez-Hernandez, S.I. Marcus, P.J. Fard, Analysis of a risk-sensitive control problem for hidden Markov chains, IEEE Transactions on Automatic Control 44 (5) (1999) 1093.
[25] J. Janssen, R. Manca, Applied semi-Markov processes, Springer, New York, 2006.
[26] H. Kong, E. Shwedyk, Sequence detection and channel state emission over finite state Markov channels, IEEE Transactions on Vehicular Technology 48 (3) (1999) 833.
[27] T. Koski, Hidden Markov models for bioinformatics, Kluwer Academic Publishers, Netherlands, 2001.
[28] P. Lanchantin, W. Pieczynski, Unsupervised non stationary image segmentation using triplet Markov chains, in: Proceedings of the Advanced Concepts for Intelligent Vision Systems (ACVIS 04), Aug. 31–Sept. 3, Brussels, Belgium, 2004.
[29] P. Lanchantin, J. Lapuyade-Lahorgue, W. Pieczynski, Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise, Signal Processing 91 (2) (February 2011) 163–175.
[30] J. Lapuyade-Lahorgue, W. Pieczynski, Unsupervised segmentation of hidden semi-Markov non stationary 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.
[31] J. Lapuyade-Lahorgue, W. Pieczynski, Unsupervised segmentation of new semi-Markov chains hidden with long dependence noise, Signal Processing 90 (11) (November 2010) 2899–2910.
[32] J. Lapuyade-Lahorgue, Sur Diverses Extensions des Chaînes de Markov Cachées Avec Application au Traitement des Signaux Radar, Ph.D., Université Pierre et Marie Curie, Paris, France, 2008.
[33] F. Le Ber, M. Benoît, C. Schott, J.-F. Mari, C. Mignolet, Studying crop sequences with CarrotAge, a HMM-based data mining software, Ecological Modelling 191 (1) (2006) 170–185.
[34] S.E. Levinson, Continuously variable duration hidden Markov models for automatic speech recognition, Computer Speech and language 1 (1986) 29–45.
[35] G. Mercier, S. Derrode, M. Lennon, Hyperspectral image segmentation with Markov chain model, in: Proceedings of IGARSS 2003, vol. 6, 2003, pp. 3766–3768.
[36] E. Monfrini, Identifiabilité et Méthode des Moments Dans les Mélanges Généralisés de Distributions du Système de Pearson, Thèse de l'Université Paris VI, soutenue le 4 janvier, 2002.
[37] 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 Research 30 (2002) 1418–1426.
[38] M.D. Moore, M.I. Savic, Speech reconstruction using a generalized HSMM (GHSMM), Digital Signal Processing 14 (1) (2004) 37–53.
[39] M. Ostendorf, V. Dikalakis, O. Kimball, From HMMs to segment models, IEEE Transactions on Speech and Audio Processing 4 (5) (1996) 360–378.
[40] A. Peng, W. Pieczynski, Adaptive mixture estimation and unsupervised local Bayesian image segmentation, Graphical Models and Image Processing 57 (5) (1995) 389–399.
[41] D.Q. Phung, T.V. Duong, S. Venkatesh, H.H. Bui, Topic transition detection using hierarchical hidden Markov and semi-Markov models, in: Proceedings of the 13th Annual ACM International Conference, 2005, pp. 11–20.
[42] W. Pieczynski, Pairwise Markov chains, IEEE Transactions on Pattern Analysis and Machine 25 (5) (2003) 634–639.
[43] W. Pieczynski, Multisensor triplet Markov chains and theory of evidence, International Journal of Approximate Reasoning 45 (1) (2007) 1–16.
[44] 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.
[45] W. Pieczynski, Convergence of the iterative conditional estimation and application to mixture proportion identification, in: Proceedings of the IEEE Statistical Signal Processing Workshop, SSP 2007, Madison, Wisconsin, USA, 26–29 August, 2007.
[46] C. Raphael, Automatic segmentation of acoustic musical signals using hidden Markov models, IEEE Transactions on Pattern Analysis and Machine 21 (4) (1999) 360–370.
[47] F. Salzenstein, C. Collet, S. Le Cam, M. Hatt, Non stationary fuzzy Markov chains, Pattern Recognition Letters 28 (16) (2007) 2201–2208.
[48] X. Tan, H. Xi, Hidden semi-Markov model for anomaly detection, Applied Mathematics and Computation 205 (2) (2008) 562–567.
[49] L.C. Thomas, D.E. Allen, N. Morkel-Kingsbury, A hidden Markov chain model for the term structure of bond credit risk spreads, International Review of Financial Analysis 11 (3) (2002) 311–329.
[50] Y. Xie, S.-Z. Yu, A large-scale hidden semi-Markov model for anomaly detection on user browsing behaviors, IEEE/ACM Transactions on Networking 17 (1) (2009) 54–65.
[51] S.-Z. Yu, H. Kobayashi, A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking, Signal Processing 83 (2) (2003) 235–250.
[52] S.-Z. Yu, H. Kobayashi, An efficient forward-backward algorithm for an explicit duration hidden Markov model, IEEE Signal Processing Letters 10 (1) (2003) 11–14.
[53] S.-Z. Yu, H. Kobayashi, Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden Markov model, IEEE Transactions on Signal Processing 54 (5) (2006) 1947–1951.
[54] S.-Z. Yu, Hidden semi-Markov models, Artificial Intelligence 174 (2) (2010) 215–243.