Signal Processing 91 (2011) 163–175

Contents lists available at ScienceDirect

Signal Processing

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

Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise

Pierre Lanchantin, Jérôme Lapuyade-Lahorgue, Wojciech Pieczynski *

Institut Telecom, Telecom SudParis, CITI Department, CNRS UMR 5157, 9 rue Charles Fourier, 91000 Evry, France

ARTICLE INFO

Article history:
Received 12 May 2009
Received in revised form
9 March 2010
Accepted 28 May 2010
Available online 8 June 2010

Keywords:
Hidden Markov chains
Triplet Markov chains
Unsupervised signal segmentation
Non-stationary data segmentation
Image segmentation
Texture classification
Copulas
Non-Gaussian noise
Correlated noise
Stochastic EM
Iterative conditional estimation

ABSTRACT

Hidden Markov chains (HMC) are a very powerful tool in hidden data restoration and are currently used to solve a wide range of problems. However, when these data are not stationary, estimating the parameters, which are required for unsupervised processing, poses a problem. Moreover, taking into account correlated non-Gaussian noise is difficult without model approximations. The aim of this paper is to propose a simultaneous solution to both of these problems using triplet Markov chains (TMC) and copulas. The interest of the proposed models and related processing is validated by different experiments some of which are related to semi-supervised and unsupervised image segmentation.

1. Introduction

The aim of this paper is to propose different workable extensions of the hidden Markov chain (HMC) model, which is well known and widely used. Hundreds of papers dealing with different subjects using HMC are published each year; let us mention general books [9,28] and papers [21,47] containing rich bibliographies. The HMC is a pairwise stochastic process $Z=(X,Y)=(X_n,Y_n)_{n=1}^N$ in which X is hidden and Y is observable. In this paper, each X_n will take its values from a finite set of classes $\Omega=\{\omega_1,\ldots,\omega_K\}$ and each Y_n will take its values from R. In such a model X is assumed to be a Markov chain and the distribution of Z is given by $p(z)=p(z_1)\prod_{n=2}^N p(z_n|z_{n-1})$, with $p(z_n|z_{n-1})=p(x_n|x_{n-1})p(y_n|x_n)$. It allows one to estimate a realization $X=x$ from the observed data $Y=y$ in different Bayesian ways. When the parameters of the model are unknown, they have to be estimated from $Y=y$ by methods like, for example, "Expectation-Maximization" (EM [9,35,47]). To achieve it, Z is assumed to be stationary: $p(z_{n-1},z_n)$ do not depend on $2\le n\le N$. However, when $p(z_{n-1},z_n)$ do depend on $2\le n\le N$, the estimation algorithm gives a stationary model which can lead, when used in signal segmentation, to very poor results. The main aim of this paper, which extends the ideas proposed in [29,32], is to model such a lack of stationarity by introducing, in a quite general context, an auxiliary hidden process.

The general idea comes from the fact that the very classical stationary HMC $Z=(X, Y)$ can be interpreted as a way of modelling a switching chain Y, which is performed by adding an auxiliary chain X whose stochastic "switches" model different stationarities of Y. Thus we can deal with switching $Z=(X, Y)$ using exactly the same idea; it can be modelled by adding an auxiliary chain U, whose stochastic "switches" model different stationarities

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

0165-1684/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2010.05.033

of $Z=(X, Y)$. Then we arrive at a chain $T=(X, U, Y)$ which, while assumed Markovian, becomes a "triplet Markov chain" (TMC) introduced in [40]. TMCs form a very rich family of models and can be used in different problems [1,45]. In particular, an original link between TMCs and the "theory of evidence" [53,55], established in [30,43], provides a way of performing the "Dempster–Shafer fusion" in a Markovian context. More recently, a particular TMC introduced in [44] allows one to exactly calculate the conditional expectation in the presence of "Markov switches", while the classical models need approximate calculation methods.

Let us notice that in our model there is precise information about the non-stationarity: there are M different stationarities. This is different from the non-stationarity dealt with in [30], where such information is not given, and this is different from the situation described in [54], where the non-stationarity is studied through the dependence of the transitions of the hidden Markov chain on the duration process.

The basic idea of modelling the non-stationarity of $Z=(X, Y)$ by a TMC $T=(X, U, Y)$ proposed in [29] is extended here and benefits from two novelties.

First, we propose a general result (Proposition in Section 2.1) specifying the conditions under which different marginal chains X, U, Y, (X, Y), (X, U), (U, Y) of a TMC $T=(X, U, Y)$ remain Markovian. This result has different practical implications. It allows us to better understand what the large generality of TMC consists of. Subsequently, it allows us to propose different particular and original TMC, which can be not much more complex, when the number of parameters defining the model is concerned, than the classical HMC. Another practical implication is the fact that it remains valid in the "mixed-states" case, in which the distribution of each X_n can simultaneously contain Dirac masses and a part which is continuous with respect to the Lebesgue measure. Such Markov models were first introduced in the context of hidden Markov fields to model "fuzzy" segmentation in [50], and showed how useful it was in different situations [48,49]. Afterwards, an extended version of such mixed-states models was proposed in [4], with application to motion texture modelling. Finally, mixed-state hidden Markov chains, first proposed in [10], found similar applications [11,13,24,33,51,52].

Second, we propose to use copulas to model possibly non-Gaussian and correlated noises. The interest of non-Gaussian marginal distributions of the noise $p(y|x)$ is well known; in particular, it is vital in modelling and processing of radar data [15,16,23,37,38]. It can also be useful for color object tracking [27]. Here we add a new possibility to model such situations. In fact, two marginal distributions of $p(y|x)$ of a given shape can be correlated in different ways. For example, two Gaussian margins $p(y_i|x)$, $p(y_j|x)$ can be correlated in a "Gaussian way" – then the couple (Y_i, Y_j) is Gaussian conditionally on $X=x$ – but they can also be correlated in a different way - then we have Gaussian margins with non-Gaussian couple. Such situations are modelled with "copulas" [39]. The introduction of copulas in hidden Markov chains is relatively recent and has provided promising results [7]. This allows us to deal with non-Gaussian correlated noise and here we extend this idea to TMC in the context of switching chains $Z=(X, Y)$.

In addition, we propose a parameter estimation method, based on the general "iterative conditional estimation" (ICE [46]) principle, which has already been successfully used in similar contexts [2,19,22,23] and which makes feasible different unsupervised Bayesian segmentation methods of non-stationary data. Let us remark that some relationships between ICE and EM have been studied in [17].

Finally, we can sum up the proposed novelties, which are of interest in any problem in which the classical HMCs are of interest, in four points:

i. specification of general conditions under which different marginal chains X, U, Y, (X, Y), (X, U), (U, Y) of a TMC $T=(X, U, Y)$ remain Markovian;
ii. general modelling of switching chain $Z=(X, Y)$, which is Markov or not, by a triplet Markov chain $T=(X, U, Y)$;
iii. introduction of copulas in TMC;
iv. parameter estimation and unsupervised segmentation of switching data hidden with non-Gaussian correlated noise.

The interest of these different novelties is attested by some experiments. In particular, the unsupervised image segmentation is addressed. To achieve it, the bi-dimensional set of pixels is transformed into a mono-dimensional sequence via a Hilbert–Peano scan, as used in [22,23], and then the different Markov chain based methods can be applied in the image processing context. Such methods have been successfully used in different problems, like multi-sensor image segmentation [23], hyperspectral image segmentation [36], multi-scale image segmentation [18], fuzzy image segmentation [24,51], or even three-dimensional image segmentation, where a three-dimensional Hilbert–Peano scan is used [5]. Although the hidden Markov field based methods are intuitively better adapted to image processing, the hidden Markov chains based ones, which are much less time-consuming, can be competitive [22]. However, let us underline the fact that Markov chain based methods are more a complement than a rival to Markov field based ones and both of them present advantages and drawbacks. Let us also mention that some results concerning unsupervised segmentation of non-stationary images using triplet Markov fields can be seen in [2,42].

Finally, let us underline that the image segmentation domain is just an example of possible application and any area in which the classical HMC is used can be concerned by the proposed models and related processing.

2. Modelling switching pairwise chains with triplet Markov chains

2.1. Triplet Markov chains

Let $Z=(X,Y)=(X_n,Y_n)_{n=1}^N$ be a pairwise random chain in which X is searched for and Y is observed. In a triplet

Markov chain (TMC) one introduces a third stochastic process $U=(U_n)_{n=1}^N$, and assumes that $T=(X, U, Y)$ is a Markov chain. TMCs are a very rich model and their generality goes far beyond the classical HMC. This greater generality is specified in the Proposition below. Although its proof is strongly inspired by the proofs in [41,43], its generality is new. There are two aspects of this greater generality. The first one is that the chain G introduced in the proposition can be equal to X, U, Y, (X, U), (X, Y), or (U, Y), and the same for H which completes G to give $T=(X, U, Y)$. The second one is the fact that the σ-additive measures γ and η introduced in the proposition can be of different kind, including the «mixing» kind, which allows one to deal with the fuzzy models.

Thus let $W=(G,H)=(G_n,H_n)_{n=1}^N$ be a random chain, where each $W_n=(G_n, H_n)$ takes its values from $\Gamma \times \mathrm{H}$. Let γ be a σ-additive measure on Γ, and let η be a σ-additive measure on H. Let us assume that W is a Markov chain and let us denote by the same letter p different densities with respect to different measures linked with γ and η. In practice, each γ and η measure is either the classical counting measure or the Lebesgue measure, but other measures, such as mixed-states measures mentioned in the introduction, can also be considered. We will say that W is stationary and reversible if $p(w_n, w_{n+1})$ does not depend on $1 \le n \le N-1$ and $p(w_n=a, w_{n+1}=b)=p(w_n=b, w_{n+1}=a)$ for each $1 \le n \le N-1$, a, and b.

We can state the following:

Proposition. Let $W=(G,H)=(G_n,H_n)_{n=1}^N$ be a stationary and reversible Markov chain, whose realizations are denoted by $w=(g,h)=(g_n,h_n)_{n=1}^N$. The three following conditions:

(i) G is a Markov chain (i.e., $W=(G, H)$ is a hidden Markov chain);
(ii) for each $2 \le n \le N$, $p(h_n|g_n,g_{n-1})=p(h_n|g_n)$; and
(iii) for each $1 \le n \le N$, $p(h_n|g)=p(h_n|g_n)$,

are equivalent.

The proof is in the Appendix.

The condition (ii) is very practical when evaluating the generality degree of a given model. For example, in this paper we are interested in TMC $T=(X, U, Y)$ such that each X_n takes its values from $\Omega=\{\omega_1,...,\omega_K\}$, each Y_n takes its values from the set of real numbers R, and each U_n takes its values from a finite set $\Lambda=\{\lambda_1,...,\lambda_M\}$. When T is stationary, its distribution is given by $p(t_1,t_2)$. We can wonder whether (X, U) is a Markov chain or not; to see this, we can apply the Proposition to $G=(X, U)$, with γ counting measure on $\Gamma=\Omega\times\Lambda$, and $H=Y$, with η Lebesgue measure on R. Assuming the Proposition hypotheses verified and writing $p(t_1,t_2)$ as $p(t_1,t_2)=p(x_1,u_1,x_2,u_2)p(y_1,y_2|x_1,u_1,x_2,u_2)$ we can say that (X, U) is a Markov chain if and only if $p(y_1|x_1,u_1,x_2,u_2)=p(y_1|x_1,u_1)$. We can also wonder whether X is a Markov chain or not; to see this, we can apply the Proposition to $G=X$, with γ counting measure on $\Gamma=\Omega$, and $H=(U, Y)$, with η the product of counting measure on Λ and by Lebesgue measure on R. Writing $p(t_1,t_2)$ as $p(t_1,t_2)=p(x_1,x_2)p(u_1,y_1,u_2,y_2|x_1,x_2)$ we can say that X is a Markov chain if and only if $p(u_1,y_1|x_1,x_2)=p(u_1,y_1|x_1)$. Of course, there are numerous other possibilities. Roughly speaking, as each variable among X_n, U_n, and Y_n can be either discrete or continuous, there are eight different basic cases. We say "basic" because each of these variables can be a vector with some components continuous, and some others discrete. Moreover, counting measure and Lebesgue one can be replaced, as stated in the introduction, by a mixed-state measure, which adds numerous new possibilities. Let us also notice that condition (iii) shows that complex noises cannot be taken into account in the classical hidden Markov chains, in which the hidden chain is a Markov one.

2.2. Bayesian segmentation using TMC

Let $T=(X, U, Y)$ be a TMC with the distribution written $p(t)=p(t_1)\prod_{n=2}^N p(t_n|t_{n-1})$. To simplify the notation, let $V=(V_n)_{n=1}^N=(X_n,U_n)_{n=1}^N$. Therefore each V_n takes its values from $\Omega\times\Lambda$, and $T=(V, Y)$ is a Markov chain. Thus (V, Y) is a "pairwise Markov chain" (PMC) and we can introduce the "Forward" quantities $\alpha_n(v_n)=p(v_n, y_1, ..., y_n)$, and the "Backward" quantities $\beta_n(v_n)=p(y_{n+1}, ..., y_N| v_n, y_n)$, which are both calculated by the following forward and backward recursions ([19,41]):

$$\alpha_1(v_1)=p(v_1, y_1);$$
$$\alpha_{n+1}(v_{n+1})=\sum_{v_n\in\Omega\times\Lambda}\alpha_n(v_n)p(v_{n+1},y_{n+1}|v_n,y_n); \tag{2.1}$$

$$\beta_N(v_N)=1;$$
$$\beta_n(v_n)=\sum_{v_{n+1}\in\Omega\times\Lambda}\beta_{n+1}(v_{n+1})p(v_{n+1},y_{n+1}|v_n,y_n). \tag{2.2}$$

Let us remark that the "Forward" probability above is the same that the one in the classical HMC, and the "Backward" probability above is an extension of the "Backward" probability in the classical HMC, the latter being defined by $\beta_n(v_n)=p(y_{n+1},...,y_N|v_n)$. In addition the related recursions (2.1), (2.2) extend to PMC the well known "Forward" and "Backward" recursions of classical HMCs.

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

$$p(v_n|y)=\frac{\alpha_n(v_n)\beta_n(v_n)}{\sum_{v_n\in\Omega\times\Lambda}\alpha_n(v_n)\beta_n(v_n)}, \tag{2.3}$$

and the transitions of the posterior Markov distribution $p(v|y)$ are calculated by

$$p(v_{n+1}|v_n,y_1, ..., y_N)=\frac{p(v_{n+1},y_{n+1}|v_n,y_n)\beta_{n+1}(v_{n+1})}{\sum_{v_{n+1}\in\Omega\times\Lambda}p(v_{n+1},y_{n+1}|v_n,y_n)\beta_{n+1}(v_{n+1})}. \tag{2.4}$$

Having calculated $p(v_n|y)$, one can then compute $p(x_n|y)$ and $p(u_n|y)$ with

$$p(x_n|y)=\sum_{u_n\in\Lambda}p(v_n|y), \quad p(u_n|y)=\sum_{x_n\in\Omega}p(v_n|y). \tag{2.5}$$

These three marginal posterior probabilities being computable, it is then possible to use different Bayesian methods to estimate $v=(x, u)$. For example, knowing the posterior marginal distributions $p(x_n|y)$, one can use the Bayesian Maximum Posterior Mode

(MPM) segmentation method $\hat{s}_{MPM}(y_1, ..., y_N) = (\hat{x}_1, ..., \hat{x}_N)$, with $\hat{x}_n = \arg\max_{x_n \in \Omega} p(x_n|y)$. The realization $u=(u_1, ..., u_N)$ of $U=(U_1, ..., U_N)$ can be estimated from $p(u_n|y)$ in a similar way. These estimates will be used in the experiments below.

2.3. Modelling switching pairwise chain by a triplet Markov one

Let $Z=(X, Y)$ be a pairwise Markov chain, where Y is observed and X is hidden. Let us assume that for each $n=1, \ldots, N-1$ the transition distribution $p(x_{n+1}, y_{n+1}|x_n, y_n)$ is one distribution among M possible ones, which means that there are M sets of parameters and the chain $Z=(X, Y)$ can switch from one to another at random unknown time. We propose to model this non-stationarity by introducing a third stochastic process $U=(U_n)_{n=1}^N$, with each U_n taking its values from a set of M values $\Lambda=\{\lambda_1, ..., \lambda_M\}$, such that $T=(Z,U)=(Z_n,U_n)_{n=1}^N$ is a homogeneous Markov chain. Thus the distribution of $T=(Z, U)$ is given by $p(x_1,y_1,u_1)$ and the transitions $p(x_{n+1}, y_{n+1}, u_{n+1}|x_n, y_n, u_n)$, which can be written

$$p(x_{n+1},y_{n+1},u_{n+1}|x_n,y_n,u_n) = p(u_{n+1}|x_n,y_n,u_n) \times p(x_{n+1},y_{n+1}|u_{n+1},x_n,y_n,u_n) \tag{2.6}$$

The first transition $p(u_{n+1}|x_n, y_n, u_n)$ models the apparitions of switches, while the second one $p(x_{n+1}, y_{n+1}|u_{n+1}, x_n, y_n, u_n)$, which will be taken equal to $p(x_{n+1}, y_{n+1}|u_{n+1}, x_n, y_n)$ because of there are exactly M possible transitions, given by u_{n+1}. Let us remark that when, in addition, $p(z_n, u_n, z_{n+1}, u_{n+1})$ does not depend on n the homogeneous chain $T=(Z, U)$ is stationary.

Finally, as already specified in Introduction, we model a switching $Z=(X, Y)$ by a stationary $T=(Z, U)$.

2.4. Learning Gaussian TMC with SEM

Let us consider the following particular stationary $T=(X, Y, U)$, which will be used in experiments below. The chain $V=(X, U)$ is a stationary Markov chain, and the distribution $p(y|v)$ of Y conditional on $V=(X, U)$ is Gaussian and verifies

$$p(y|v) = p(y|x) = \prod_{n=1}^{N} p(y_n|x_n) \tag{2.7}$$

Let us notice that in spite of the simplicity of $p(y|v)$, the couple (X, Y) is not necessarily a HMC, because, according to the Proposition in Section 2.2, X is not necessarily a Markov chain. Otherwise, we assume that the Gaussian distributions $p(y_n|x_n)$ are independent from $n=1,...,N$. Thus for K classes $\Omega=\{\omega_1,...,\omega_K\}$, we have to estimate K means $\mu_1,\ldots,\mu_K$, and K variances $\sigma_1^2, ..., \sigma_K^2$ of the K Gaussian densities $p(y_n|x_n=\omega_1),\ldots,p(y_n|x_n=\omega_K)$. Furthermore, each U_n taking its values from $\Lambda=\{\lambda_1,...,\lambda_M\}$, the distribution of the stationary Markov chain $V=(U, X)$ is given by $(KM)^2$ parameters $p_{ij}=p(v_1=i, v_2=j)$ (to simplify, we consider the set $\Omega\times\Lambda$ as being $\{1,2,.....,KM\}$), which is a probability on $[\Omega\times\Lambda]^2$.

In Section 3 we will use the Stochastic "Expectation-Maximization" (SEM [12]) method, which is a stochastic approximation of the "Expectation-Maximization" (EM) method [35,47], and which runs as follows:

(i) consider an initial value $\theta^0 = (p_{ij}^0, \mu_k^0, (\sigma_k^0)^2)$, for $0 \le i, j \le KM$, and $1 \le k \le K$ obtained in some classical way specific to a given application;

(ii) for each $q \in N^*$:

- simulate $V=v^q$ according to the Markov chain $p(v|y)$ based on θ^q;
- calculate $\theta^{q+1} = (p_{ij}^{q+1}, \mu_k^{q+1}, (\sigma_k^{q+1})^2)$ with

$$p_{ij}^{q+1} = \frac{1}{N-1}\sum_{n=1}^{N-1} 1_{[v_n^q = i, v_{n+1}^q = j]}, \tag{2.8}$$

$$\mu_k^{q+1} = \frac{\sum_{n=1}^{N} y_n 1_{[x_n^q = \omega_k]}}{\sum_{n=1}^{N} 1_{[x_n^q = \omega_k]}}, \tag{2.9}$$

$$(\sigma_k^{q+1})^2 = \frac{\sum_{n=1}^{N} (y_n - \mu_k^{q+1})^2 1_{[x_n^q = \omega_k]}}{\sum_{n=1}^{N} 1_{[x_n^q = \omega_k]}}; \tag{2.10}$$

(iii) stop the procedure when the sequence (θ^q) has became "steady" according to some criteria, specific to a given application.

3. Semi-supervised segmentation of switching chains and "non-stationary" images hidden with Gaussian noise.

Let (X, Y) be the classical Gaussian hidden Markov chain with independent noise (HMC), with the distribution of the noise given by $p(y|x) = \prod_{n=1}^{N} p(y_n|x_n)$. We will compare the HMC and the proposed TMC model in the context of semi-supervised MPM segmentation, where the model parameters are estimated with SEM and the number of classes are set manually.

We present three series of experiments. In the first series, the data (X, Y) are simulated according to a chosen $U=u$. In the second series, the hidden data $X=x$ are given and the Gaussian noise is simulated. In the third series we deal with a collage of Brodatz [6] patterns segmentation. It these experiments, the number of real and auxiliary classes will be supposed to be known or will be set manually such that the segmentations will be semi-supervised. However, it is possible to automatically compute these numbers by using Bayes Information Criterion [3] as it will be shown in Section 4.2.3, which will lead to unsupervised segmentation.

Let us consider two classes $\Omega=\{\omega_1, \omega_2\}$ and three stationarities $\Lambda=\{\lambda_1,\lambda_2,\lambda_3\}$. The length of the chain before conversion using a Hilbert–Peano path is $N=128\times128$. The realization $U=u$ is fixed arbitrarily: the first half of $u=(u_1,...,u_N)$ is λ_1 (grey), the third quarter is λ_2 (black), and the last quarter is λ_3 (grey-dark). Using the Hilbert–Peano scan this gives an image $U=u$ as shown in Fig. 1. The sequence $X=x$ is simulated in the following way. $X_1=x_1$ is sampled from $\Omega=\{\omega_1, \omega_2\}$ (with ω_1 "white" and ω_2 black) according to the distribution (0.5, 0.5), and $X_2=x_2,\ldots, X_{N/2}=x_{N/2}$ are sampled using the transition matrix M_1.

Then $X_{(N/2)+1}=x_{(N/2)+1},\ldots,X_{3N/4}=x_{3N/4}$ are sampled using the transition matrix M_2, and, finally, $X_{3N/4}=x_{3N/4},\ldots,X_N=x_N$ are sampled using the transition matrix M_3. In the experiment, the matrices considered are

$$M_1=\begin{bmatrix}0.98 & 0.02\\ 0.02 & 0.98\end{bmatrix},\quad M_2=\begin{bmatrix}0.5 & 0.5\\ 0.5 & 0.5\end{bmatrix},\quad \text{and}$$

$$M_3=\begin{bmatrix}0.02 & 0.98\\ 0.98 & 0.02\end{bmatrix}.$$

The observed $Y=y$ is sampled according to $p(y_1|x_1),\ldots,p(y_N|x_N)$, where $p(y_n|x_n=\omega_1)$ is Gaussian with mean 1 and variance 1, and $p(y_n|x_n=\omega_2)$ is Gaussian with mean 3 and variance 1. The realization $X=x$ is then estimated by the Bayesian MPM method from $Y=y$ in two different ways. The first segmentation is obtained using the parameters estimated with the SEM algorithm assuming that (X, Y) is a classical HMC with two real classes, with the hidden chain assumed stationary, which gives $\hat{X}^1=\hat{x}^1$ (HMC), $\tau=26\%$ (see Fig. 1, where τ is the error ratio). The second segmentation is obtained using the parameters estimated with the SEM algorithm assuming that (X, U, Y) is a TMC presented above with two real classes and three auxiliary classes, which gives $\hat{X}^2=\hat{x}^2$ (TMC), $\tau=5\%$. We also show in Fig. 1 the estimates $\hat{U}=\hat{u}$ of the auxiliary process $U=u$. We clearly see how the use of TMCs can improve the results obtained with the classical HMC model by considering three different stationarities. Of course, the results are good because the data are not strongly noisy; however, this example shows that the lack of stationarity of X can have heavy negative consequences on the efficiency of the classical HMC based methods, and this result also shows that these negative consequences can be avoided using a TMC.

In addition, different parameters are well estimated (Tables 1 and 2), which shows the good behavior of the SEM associated with TMCs; while the classical SEM associated with the classical HMCs gives poorer results.

In the second series, let us consider the binary image of a zebra presented in Fig. 2, with $N=256\times256$. The skin of the zebra has got stripes of uneven size and shape. For instance, the stripes on the neck and on the legs of the zebra are thinner than those on the body. The observed $Y=y$ is then sampled in the same way as in the first experiment and the realization $X=x$ is estimated by the Bayesian MPM method from $Y=y$ using the parameters estimated with the SEM algorithm considering the HMC and the TMC models. We can see in Fig. 1 that the stripes are not well restored while using HMC. The result $\hat{X}^1=\hat{x}^1$ involves the error ratio of $\tau=7.2\%$ (see Fig. 2). When using a $T=(X, U, Y)$ with 3 auxiliary classes, we can take three different stationarities into account (Fig. 2) on the estimates $\hat{U}=\hat{u}$ of the auxiliary processes: black labels are assigned to the background, grey labels to the body, which presents large stripes and white labels are assigned to the neck and to the legs of the zebra which present thin stripes. By considering different stationarities, we improve the parameters estimation (Tables 3 and 4) and

Table 1
Estimates of noise parameters.

	μ_1	σ_1^2	μ_2	σ_2^2
True	1.00	1.00	3.00	1.00
HMC	0.98	1.03	2.33	1.91
TMC	1.02	1.01	2.98	0.94

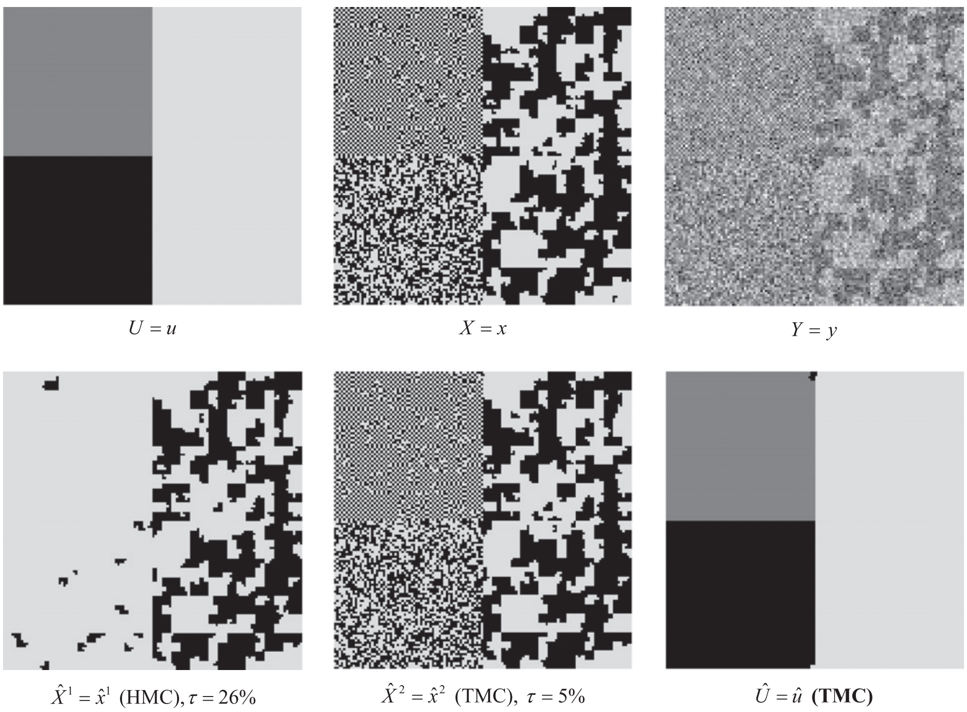

Fig. 1. Different images corresponding to the first experiment and their semi-supervised HMC and TMC based segmentations.

Table 2
Estimates of the transition matrix $p(v_n|v_{n-1})$ (A) and $p(v_1)$ (B).

(A)	$v_1=(\omega_1,\lambda_1)$	$v_2=(\omega_2,\lambda_1)$	$v_3=(\omega_1,\lambda_2)$	$v_4=(\omega_2,\lambda_2)$	$v_5=(\omega_1,\lambda_3)$	$v_6=(\omega_2,\lambda_3)$
(ω_1,λ_1)	0.98	0.02	~0	2e-4	~0	~0
(ω_2,λ_1)	0.02	0.98	~0	~0	0.01	~0
(ω_1,λ_2)	~0	5e-4	0.50	0.49	0.02	~0
(ω_2,λ_2)	~0	~0	0.52	0.47	~0	~0
(ω_1,λ_3)	~0	~0	~0	~0	~0	1
(ω_2,λ_3)	~0	~0	5e-4	~0	0.95	0.04
(B)	0.26	0.24	0.13	0.12	0.09	0.13

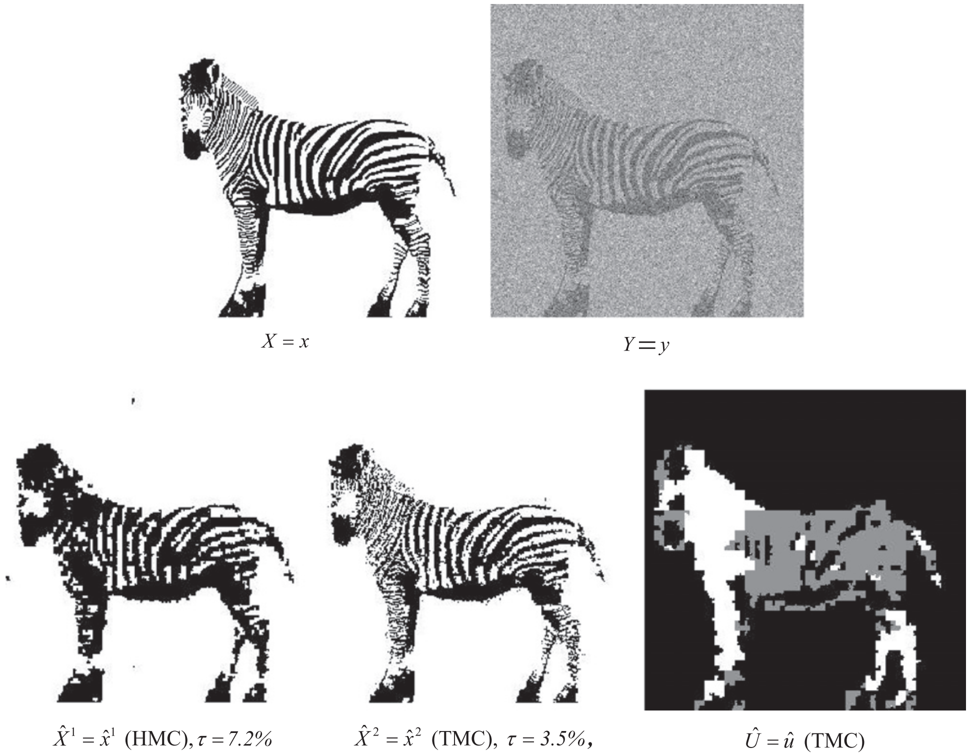

Fig. 2. Different images corresponding to the second experiment and their semi-supervised HMC and TMC based segmentations.

Table 3
Estimates of noise parameters.

	μ_1	σ_1^2	μ_2	σ_2^2
True	1.00	1.00	3.00	1.00
HMC	1.47	1.31	2.99	1.00
TMC	1.02	1.02	2.99	1.00

the segmentation quality result ($\hat{X}^2=\hat{x}^2$ involves the error ratio of τ=3.5%) as well.

The third series of experiments deals with the semi-supervised segmentation of Brodatz patterns [6] which are often used in texture segmentation. We built an image $Y=y$ according to $U=u$ presented in Fig. 3 by considering three different Brodatz textures. We assume that $T=(X, U, Y)$ is a TMC model with 3 auxiliary classes and 4 real classes. The segmentation results are presented in Fig. 3. The interesting point is the fact that $\hat{U}=\hat{u}$ is well estimated with τ=2.5%, which shows that TMC models can be useful in the context of texture segmentation. Let us notice that we do not have the true $X=x$.

4. Semi-supervised segmentation of switching hidden Markov chains with copulas

4.1. Copulas in triplet Markov chains

Copulas are particular cumulative distribution functions (cdf), which allow one to easily propose different correlate variables distributions. Let $C:[0,1]^2 \to [0,1]$ be a continuous cdf on $[0,1]^2$ such that the corresponding marginal distributions are uniform distributions on [0,1]. Such a cdf is called "copula".

Let Y_1, Y_2 be two real random variables, F_1, F_2 the cdfs of their laws, and F the cdf of the law of $Y=(Y_1, Y_2)$. For F_1, F_2 continuous, there is a single copula $C:[0,1]^2 \to [0,1]$, such

Table 4
Estimates of the transition matrix $p(\upsilon_n|\upsilon_{n-1})$ (A) and $p(\upsilon_1)$ (B).

(A)	$v_1=(\omega_1, \lambda_1)$	$v_2=(\omega_2, \lambda_1)$	$v_3=(\omega_1, \lambda_2)$	$v_4=(\omega_2, \lambda_2)$	$v_5=(\omega_1, \lambda_3)$	$v_6=(\omega_2, \lambda_3)$
(ω_1, λ_1)	0.98	$\sim$0	$\sim$0	0.01	$\sim$0	$\sim$0
(ω_2, λ_1)	$\sim$0	0.99	$\sim$0	$\sim$0	$\sim$0	$\sim$0
(ω_1, λ_2)	$\sim$0	$\sim$0	0.85	0.14	$\sim$0	$\sim$0
(ω_2, λ_2)	0.02	$\sim$0	0.10	0.88	$\sim$0	$\sim$0
(ω_1, λ_3)	$\sim$0	$\sim$0	$\sim$0	$\sim$0	0.6	0.4
(ω_2, λ_3)	$\sim$0	$\sim$0	$\sim$0	$\sim$0	0.34	0.65
(B)	0.06	0.68	0.06	0.08	0.05	0.07

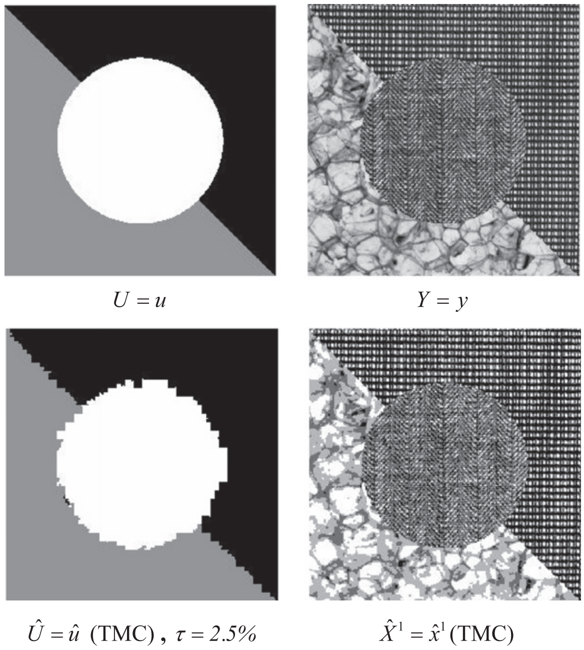

Fig. 3. Image corresponding to the third experiment and its semi-supervised TMC based segmentations.

that [39] (also see [7,8,14,20,34] and the references therein):

$$\forall (y_1,y_2)\in R^2,\ F(y_1,y_2)=C(F_1(y_1),F_2(y_2)) \tag{4.1}$$

Let $Y=(Y_1, Y_2)$ be a random vector with the distribution defined by the cdf $F(y_1,y_2)=C(F_1(y_1),F_2(y_2))$. An important property is that if we consider φ_1, φ_2 two continuous strictly increasing functions from R to R, the distribution of $W=(W_1, W_2)$ defined from $Y=(Y_1, Y_2)$ by $W_1=\varphi_1(Y_1)$, $W_2=\varphi_2(Y_2)$, admits different marginal distributions but is defined by the same copula C. This means that one can use a given distribution F to define any another distribution F' having the same copula than F and having any desired marginal distributions F'_1, F'_2. In fact, it is sufficient to take $W_1=\varphi_1(Y_1)=F'_1(F_1^{-1}(Y_1))$ and $W_2=\varphi_2(Y_2)=F'_2(F_2^{-1}(Y_2))$. This is a simple way of creating correlated variables distributions within the desired margins.

Let us assume that the cdf F of the random variables Y_1, Y_2 is derivable. Derivating (4.1) gives

$$p(y_1,y_2)=p(y_1)p(y_2)c[F_1(y_1),F_2(y_2)] \tag{4.2}$$

Let us consider two examples which will be used in the experiments of the next sub-section.

Example 4.1. Gaussian copula

The Gaussian copula C_G is the copula associated to a Gaussian distribution $p(y_1, y_2)$ with (4.2). As the copula C_G does not depend on marginal distributions $p(y_1)$, $p(y_2)$, let us assume them to be of mean 0 and of variance 1. Let us denote by Φ its cdf. Setting in (4.2) $u_1=\Phi(y_1)$, and $u_2=\Phi(y_2)$, we find

$$c_G(u_1,u_2)=\frac{f_G(\Phi^{-1}(u_1),\Phi^{-1}(u_2))}{p_G(\Phi^{-1}(u_1))p_G(\Phi^{-1}(u_2))}, \tag{4.3}$$

where f_G is the density of the Gaussian distribution $N\left[(0,0),\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right]$ on R^2 and p_G the density of the Gaussian distribution $N(0,1)$ on R. We see that the Gaussian copula on $[0,1]^2$ is defined by one real parameter ρ.

Example 4.2. Student copula

Let W be a strictly non-negative real random variable and let A be a Gaussian random vector with the distribution $N[(0,0), \Sigma]$ on R^2. Let $m\in R^2$, and $Y=m+W^{-(1/2)}A$. Such a model, which can be seen as a Gaussian distribution with stochastically affected covariance matrix, is used in radar signal processing, where W is called "texture" and A the "speckle". The distribution of Y can then be calculated for some particular distributions of W. Denoting by $\gamma(a, b)$ the gamma distribution given by the density $f(y)=\frac{b^a}{\Gamma(a)}e^{-by}y^{a-1}1_{y\geq 0}$, where a, $b>0$ and $\Gamma(a)=\int_0^{+\infty}y^{a-1}e^{-y}\,dy$, we can state that if the distribution of W is $\gamma(\frac{v}{2}, \frac{v}{2})$, then the distribution of Y is the following "Student law with v degrees of freedom"

$$f_S(y_1, y_2)=\frac{\Gamma((v+2)/2)|\Sigma|^{-(1/2)}}{\pi v\Gamma(v/2)}\left(1+\frac{2(y_1-m)'\Sigma^{-1}(y_2-m)}{v}\right)^{-(v+2)/2}, \tag{4.4}$$

and the marginal distribution of (4.4) is (with σ^2 denoting the variance of Gaussian marginal variable of A)

$$p_S(y_1)=\frac{\Gamma(v+1/2)}{\sqrt{\pi v\sigma^2}\,\Gamma(v/2)}\left(1+\frac{2(y_1-m)^2}{v\sigma^2}\right)^{-(v+1)/2} \tag{4.5}$$

The Student copula is then obtained from (4.4) and (4.5) using (4.2), in a similar way as was done for the Gaussian copula. Denoting by Φ_S the cdf of the standardized (mean equal to 0 and variance equal to 1) Student distribution on R and setting $u_1=\Phi_S(y_1)$, $u_2=\Phi_S(y_2)$, we have

$$c_S(u_1,u_2)=\frac{f_S(\Phi_S^{-1}(u_1),\Phi_S^{-1}(u_2))}{p_S(\Phi_S^{-1}(u_1))p_S(\Phi_S^{-1}(u_2))},\tag{4.6}$$

where f_S is given by (4.4) with m=0 and $\Sigma=\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$, and p_S is given by (4.5) with m=0 and σ^2=1. We see that the Student copula on $[0,1]^2$ is defined by two real parameters ρ and ν.

Let T=(X, U, Y) be a stationary reversible TMC. Setting V=(X, U) to simplify, the distributions of T is then defined by $p(t_1, t_2)=p(v_1, v_2)p(y_1, y_2|v_1, v_2)$. The distributions $p(y_1, y_2|v_1, v_2)$ can then be defined using different margins and different copulas. To our knowledge, such models were first proposed in the context of the classical hidden Markov chains in [7], and different experiments have validated their interest. Thus here we extend this kind of model to TMCs, with application to switching hidden chains.

Let us notice the great richness of such a family of models. If there are K possible values for each X_n and M possible values for each U_n, there are KM possible values for each V_n=(X_n, U_n), and thus there are K^2M^2 possible values for (v_1, v_2). Because of the reversibility, there are therefore $(KM(KM-1))/2$ possible distributions $p(y_1, y_2| v_1, v_2)$ on R^2, and each of them is defined by two marginal distributions and a copula. For example, if there are three stationarities and two classes, there are fifteen possibly different marginal distributions on R, and fifteen possibly different copulas on R^2. Let us also notice that considering different margins can be necessary in real situations; as shown in [15], different classes can present different shapes of noise in the same radar image.

4.2. Experiments

The aim of the experiments presented below is to test whether using the true copula is of importance. Let us consider two copulas: the Gaussian copula and the Student one. In all experiments the marginal distributions of the observed chain conditionally on the hidden data will be Gaussian. Thus when the copula used is the Gaussian copula we are faced with the classical Gaussian case. When the copula used is the Student copula, we have an original model with Gaussian margins and a Student copula.

The number of classes is set manually and the parameters are estimated with the "Iterative Conditional Estimation" (ICE [2,19,22,23,46]). Let $\theta=(\theta_1, ..., \theta_m)$ be the set of parameters. To apply ICE, we need an estimator $\hat\theta(v, y)$ from complete data (V, Y)=(v, y). Likely to EM, ICE is an iterative method and runs as follows:

(i) Initialize θ^0;
(ii) compute $\theta_i^{q+1}=E[\hat\theta_i(V, Y)|Y=y, \theta^q]$ for the components θ_i for which this computation is workable;
(iii) for other components θ_i, simulate v_1^q, ..., v_l^q according to $p(v|y,\theta^q)$ and put $\theta_i^{q+1}=(\hat\theta_i(v_1^q, y)+ ... +\hat\theta_i(v_l^q, y))/l$.

The marginal distributions are Gaussian in our model and, as (ii) cannot be used for all components, we sample one realization v^q (we take l=1) of V and we use (iii), with the estimates defined by (2.9) and (2.10), to re-estimate the means and the variances. To re-estimate the copulas, we use the classical estimate of correlations

$$\rho_{km}^{q+1}=\frac{\sum_{n=1}^{N-1}(y_n-\mu_k^{q+1})(y_{n+1}-\mu_m^{q+1})1_{[x_n=\omega_k,x_{n+1}=\omega_m]}}{\sum_{n=1}^{N-1}1_{[x_n=\omega_k,x_{n+1}=\omega_m]}}\tag{4.7}$$

Applying (4.7) to the sampled v^q gives the re-estimated correlation, which gives Gaussian or Student copula, according to which case we are in.

The parameters p_{ij} are estimated from the complete data by (2.8); applying the principle (ii) of ICE gives

$$p_{ij}^{q+1}=\frac{1}{N-1}\sum_{n=1}^{N-1}p(v_n=i,v_{n+1}=j|y),\tag{4.8}$$

knowing that $p(v_n, v_{n+1}|y)=p(v_n|y)p(v_{n+1}|v_n, y)$ is computable with (2.3)–(2.4).

We propose two series of experiments. In the first one, the data are either simulated with a Gaussian copula or simulated with a Student copula. We estimate the hidden realization and the parameters by using both models. The aim of these experiments is to know which of these two models is robust and which of these two models is worth using if the data are not simulated with the appropriate model.

In all experiments, U_n takes its values from $\Lambda=\{\lambda_1, \lambda_2\}$ and its transition matrix is $R=[r_{ij}]_{1\le i,\ j\le 2}$, with $r_{11}=r_{22}$=0.9999 and $r_{12}=r_{12}$=0.0001. The two transition matrices corresponding to the two states of U_n are $M^1=[m_{ij}^1]_{1\le i,\ j\le 2}$, with $m_{11}^1=m_{22}^1=0.99$ and $m_{12}^1=m_{21}^1=0.01$, and $M^2=[m_{ij}^2]_{1\le i,j\le 2}$, with $m_{11}^2=m_{22}^2$=0.90 and $m_{12}^2=m_{21}^2$=0.10. The marginal distributions are Gaussian N(0,1) and N(2,1). The sample size is 256×256.

4.2.1. Data with Student copula

In this first series we will consider the Student copula with the parameters (ρ,ν)=(0,1). We sample the TMC T=(X, U, Y) and the sampled realizations are presented in Fig. 4 as being three images, using the Hilbert–Peano scan. The same Fig. 4 contains the estimates of (X, U)=(x, u) based on the true parameters, which is thus the reference one. Of course, it is just a representation and this experiment is not about "image segmentation" problem. Fig. 5 contains the results of unsupervised segmentation of the data Y=y, once with the true Student copula and once with the wrong Gaussian copula. Tables 5 and 6 contain the parameters estimates.

According to the results we can say that there are situations in which taking into account the true copula is vital, and considering Gaussian copula instead of the true Student one leads to very poor results. Let us recall the distributions of the chain (X, U) are exactly the same in the both models, and, what is more, the marginal distributions $p(y_n|x, u)$ are also exactly the same. We can see how important the nature of copulas alone can be.

4.2.2. Data with a Gaussian copula

Let us consider the Gaussian copula defined by ρ=0.9. As the marginal distributions are Gaussian, we have a

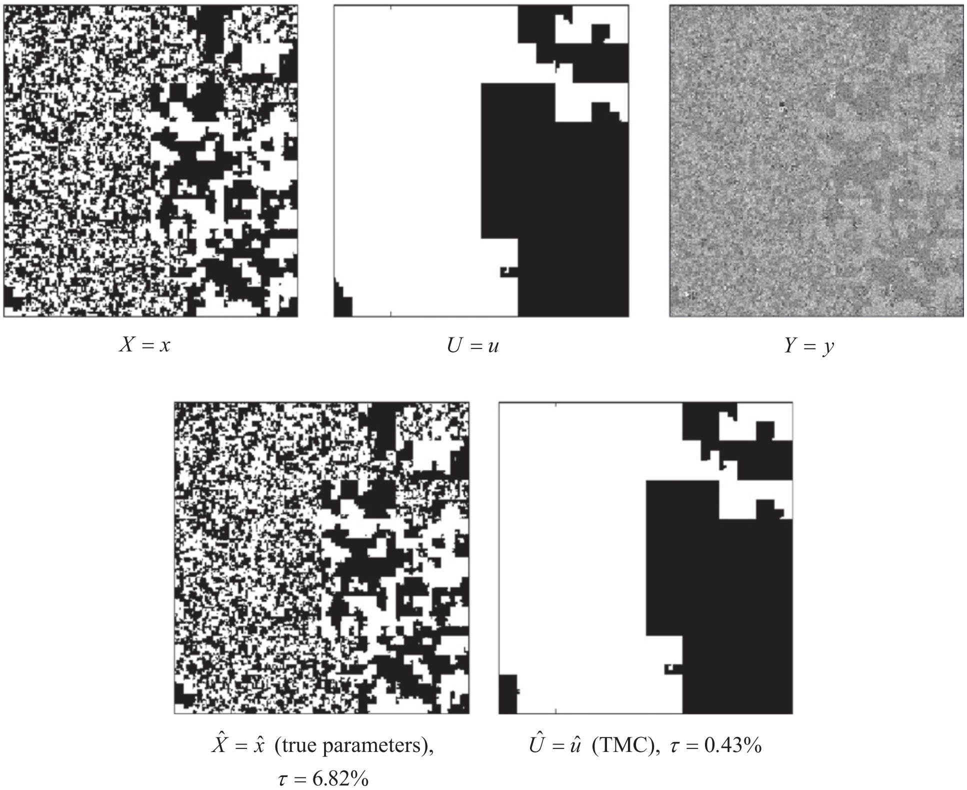

Fig. 4. Simulated data with Student copula and true parameters based estimation of $X=x$ and $U=u$. τ is the error ratio.

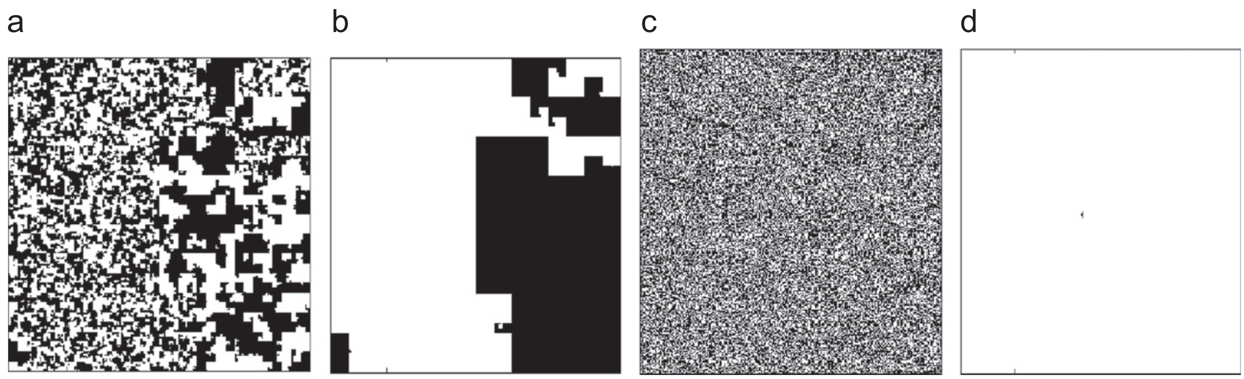

Fig. 5. Semi-supervised segmentation, parameters estimated with ICE given in Tables 5 and 6. (a) and (b) use the correct Student copula. (c) and (d): use of the Gaussian copula. (a) SC $\hat{X}=\hat{x}$, τ=7.32%; (b) SC $\hat{U}=\hat{u}$, τ=0.51%; (c) GC $\hat{X}=\hat{x}$, τ=43.14%; and (d) GC $\hat{U}=\hat{u}$, τ=37.8%.

Table 5
Parameters estimated with ICE. True parameters, with Student copula, are $\mu_1=0$, $\mu_2=2$, $\rho_1=\rho_2=0$, $\sigma_1^2=\sigma_2^2=1$.

	μ_1	σ_1^2	ρ_1	μ_2	σ_2^2	ρ_2
GC	0.10	2.39	0.85	1.93	2.32	0.89
SC	0.10	2.40	0.12	1.90	2.37	0.10

correlated Gaussian noise. Thus the aim here is to test whether some Student copula based model can compete. The sampled realizations of the TMC $T=(X, U, Y)$ and the estimates of $(X, U)=(x, u)$ based on the true parameters are presented in Fig. 6, while Fig. 7 contains the results of unsupervised segmentation of the data $Y=y$, once with the true Gaussian copula and once with the wrong Student copula. Tables 7 and 8 contain the parameter estimates.

As in the previous sub-section, we can say that there are situations in which taking into account the true copula is vital, and considering Student copula instead of the true Gaussian one can lead to very poor results. As above, the distributions of the chain (X, U) and the marginal distributions $p(y_n|x, u)$ are exactly the same in both the models. As above, we see how important the nature of copulas alone can be.

Table 6
Parameters estimated with ICE.

	Transition $p(u_{n+1}\|u_n)$	Transition $p(x_{n+1}\|x_n,u_{n+1})$
GC	$R=\begin{pmatrix}0.9627 & 0.0373\\ 0.0373 & 0.9627\end{pmatrix}$	$M_1=\begin{pmatrix}0.934 & 0.066\\ 0.066 & 0.934\end{pmatrix}$, $M_2=\begin{pmatrix}0.632 & 0.368\\ 0.368 & 0.632\end{pmatrix}$
SC	$R=\begin{pmatrix}0.9999 & 0.0001\\ 0.0001 & 0.9999\end{pmatrix}$	$M_1=\begin{pmatrix}0.993 & 0.007\\ 0.007 & 0.993\end{pmatrix}$, $M_2=\begin{pmatrix}0.925 & 0.075\\ 0.075 & 0.925\end{pmatrix}$
True parameters	$R=\begin{pmatrix}0.9999 & 0.0001\\ 0.0001 & 0.9999\end{pmatrix}$	$M_1=\begin{pmatrix}0.99 & 0.01\\ 0.01 & 0.99\end{pmatrix}$, $M_2=\begin{pmatrix}0.90 & 0.10\\ 0.10 & 0.90\end{pmatrix}$

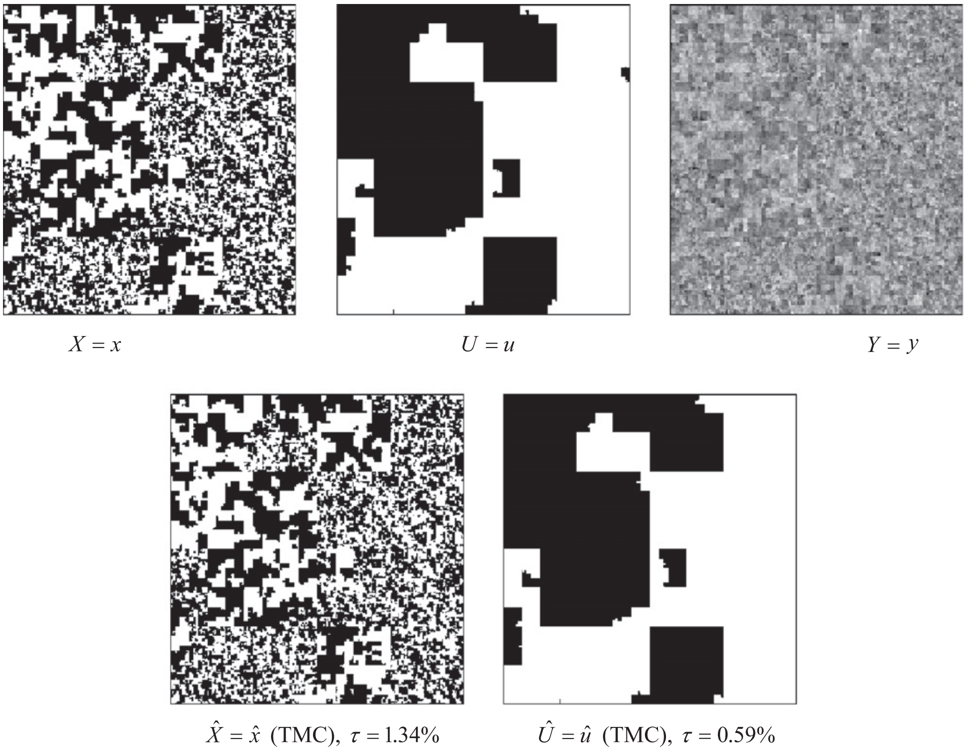

Fig. 6. Simulated data with Gaussian copula and true parameters based estimation of $X=x$ and $U=u$. τ is the error ratio.

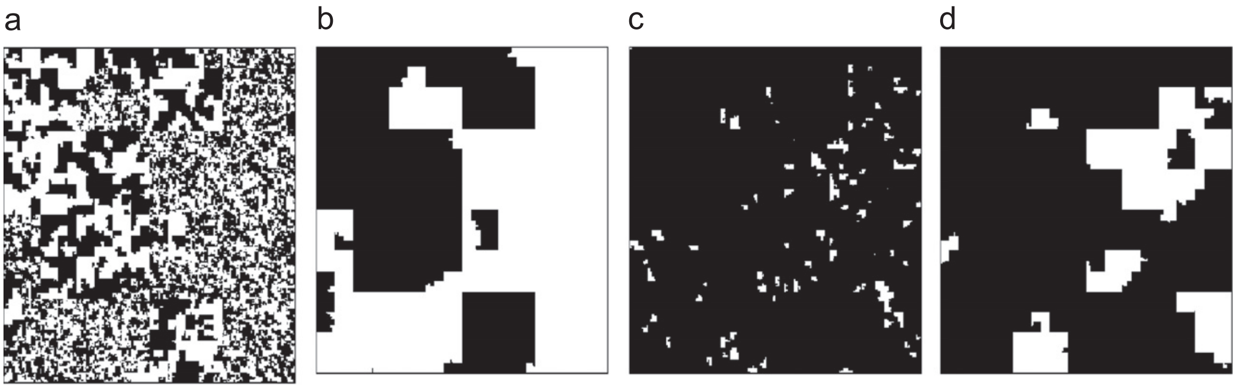

Fig. 7. Semi-supervised segmentation, parameters estimated with ICE given in Tables 7 and 8. (a) and (b) use the correct Gaussian copula, while (c) and (d) use the Student copula. (a) GC $\hat{X}=\hat{x}$, τ=1.36%; (b) GC $\hat{U}=\hat{u}$, τ=0.53%; (c) SC $\hat{X}=\hat{x}$, τ=49.57%; and (d) SC $\hat{U}=\hat{u}$, τ=36.71%.

Table 7
Parameters estimated with ICE. True parameters, with Gaussian copula, are $\mu_1=0$, $\mu_2=2$, $\rho_1=\rho_2=0.9$, $\sigma_1^2=\sigma_2^2=1$.

	μ_1	σ_1^2	ρ_1	μ_2	σ_2^2	ρ_2
GC	−0.01	1.00	0.90	1.98	1.02	0.90
SC	0.92	1.95	0.93	1.74	2.02	0.93

Table 8
Parameters estimated with ICE.

	Transition $p(u_{n+1}\vert u_n)$	Transition $p(x_{n+1}\vert x_n,u_{n+1})$
GC	$R=\begin{pmatrix}0.9999 & 0.0001\\ 0.0001 & 0.9999\end{pmatrix}$	$M_1=\begin{pmatrix}0.991 & 0.009\\ 0.009 & 0.991\end{pmatrix}$, $M_2=\begin{pmatrix}0.900 & 0.100\\ 0.100 & 0.900\end{pmatrix}$
SC	$R=\begin{pmatrix}0.9990 & 0.0010\\ 0.0010 & 0.9990\end{pmatrix}$	$M_1=\begin{pmatrix}0.995 & 0.005\\ 0.005 & 0.995\end{pmatrix}$, $M_2=\begin{pmatrix}0.96 & 0.04\\ 0.04 & 0.96\end{pmatrix}$
True parameters	$R=\begin{pmatrix}0.9999 & 0.0001\\ 0.0001 & 0.9999\end{pmatrix}$	$M_1=\begin{pmatrix}0.99 & 0.01\\ 0.01 & 0.99\end{pmatrix}$, $M_2=\begin{pmatrix}0.90 & 0.10\\ 0.10 & 0.90\end{pmatrix}$

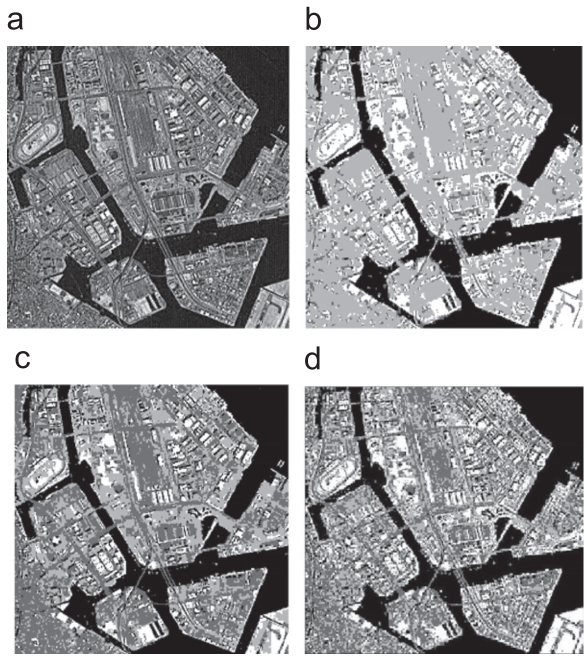

Fig. 8. (a) Real Tokyo image, (b) unsupervised MPM segmentation based on the classical stationary HMC model with 4 real classes, (c) unsupervised MPM segmentation based on the Gaussian TMC (Gaussian margins and Gaussian copula) with two stationarities, and (d) unsupervised MPM segmentation based on the TMC with Gaussian margins and Student copula.

4.2.3. Unsupervised segmentation of real image

We now consider the segmentation of a real 256×256 size satellite image of Tokyo, which is presented in Fig. 8(a). It is segmented by an MPM method based on three models: the classical Gaussian HMC, the TMC with Gaussian margins and a Gaussian copula, and a TMC with Gaussian margins and a Student copula.

To estimate the number of real classes and auxiliary switching classes, we compare the models using the Bayes Information Criterion (BIC) [26]. This criterion is defined as $BIC=-2LL+p\log(N)$ where LL is the log-likelihood of the model, p its number of independent parameters and N the number of data. We do not take into account parameters estimated to be zeros according to the convention established in [3]. The best model is the one with the lowest BIC. We successively segment the image considering the TMC, Gaussian margins, and a Gaussian copula with different numbers or real and auxiliary states and compute the BIC value until it reaches a minimum value. Results are presented on Table 9. Among the different models we tested, the one with 4 classes and 2 auxiliary states is the best according to BIC. It is also interesting to note that the model with 3 real classes and 2 auxiliary classes is better than the one with 4 real classes and 1 auxiliary class which means that in some case, it is better to consider several stationarities than to increase the number of classes.

The segmentation results considering 4 classes are presented in Fig. 8(b–d), respectively. We notice that using TMC with 2 stationarities (c) instead of the classical HMC-IN (b) obviously improves the segmentation. The difference between (c) and (d) is less striking; however, (d) seems to restore fine details better in many spots. Thus using a Student copula instead of the Gaussian one, which here is equivalent to using the classical Gaussian distribution, can be of interest in real image segmentation.

5. Conclusion and perspectives

We proposed a new model based on the triplet Markov chains, which allows one to deal with the hidden switching data noisy with non-Gaussian correlated noise. Different experiments validated the interest of the new model, in the case of simulated as well as real data.

Let us mention some directions as perspectives to further work. Extending Markov chains to Markov trees, or even

Table 9
Comparison between models according to BIC criterion. The case of Gaussian margins and Gaussian copula.

Number of real states	Number of auxiliary states	LL	Number of parameters	BIC
2	1	−338,750	7	677,580
2	2	−337,460	13	675,060
2	3	−337,360	23	**674,980**
3	1	−334,310	14	668,780
3	2	−332,070	27	664,440
3	3	−331,760	55	**664,130**
4	1	−333,020	21	666,280
4	2	−331,210	43	**662,610**

more general Bayesian networks [25], seems to be the natural way for further investigations. Another possibility would be to investigate different uses of the mixed-states measure. This has been done in the context of hidden fuzzy chains in [52], and similar applications to motion texture models should be of interest. Let us also mention that the mono-sensor case dealt within this paper can be extended to the multisensor case, in which copulas are used to model the dependence among sensors [8]. Thus in the multisensor case copulas could be used twice: first to model the "temporal" dependence, and second to model the sensors' dependence. Finally, the use of copulas and non-Gaussian correlated noise in the context of long-memory noise models recently proposed in [31] should also provide interesting perspectives for further works.

Appendix. Proof of the Proposition

(i) implies (ii). W being a Markov chain, we can write

$$p(w)=\frac{p(w_1,w_2)\dots p(w_{N-1},w_N)}{p(w_2)\dots p(w_{N-1})}=\underbrace{\frac{p(g_1,g_2)\dots p(g_{N-1},g_N)}{p(g_2)\dots p(g_{N-1})}}_{a(g)}\underbrace{\frac{p(h_1,h_2|g_1,g_2)\dots p(h_{N-1},h_N|g_{N-1},g_N)}{p(h_2|g_2)\dots p(h_{N-1}|g_{N-1})}}_{b(h)} \quad (1)$$

In particular, (W_1, W_2, W_3) also is a Markov chain and thus (1) is valid for N=3. Therefore, let us consider (1) with N=3. As G is a Markov chain according to (i), $(p(g_1,g_2)\,p(g_2,g_3))/p(g_2)$ in (1) is the distribution of G, which means that the integral of $b(h_1,h_2,h_3)$ in (1) is equal to one: $\int_{H^3} b(h_1,h_2,h_3)\,d\eta(h_1)\,d\eta(h_2)\,d\eta(h_3)=1$. After having integrated $b(h_1,h_2,h_3)$ with respect to h_1 and h_2, we obtain

$$\int_H \frac{p(h_2|g_1,g_2)p(h_2|g_2,g_3)}{p(h_2|g_2)}\,d\eta(h_2)=1 \quad (2)$$

Let us consider the scalar product $\langle f^1,f^2\rangle=\int_H (f^1(h_2)f^2(h_2)/p(h_2|g_2))\,d\eta(h_2)$ and the associated norm $\|f\|=\sqrt{\langle f,f\rangle}$. According to (2) we have $\|p(h_2|g_1,g_2)-p(h_2|g_2,g_3)\|^2=\|p(h_2|g_1,g_2)\|^2+\|p(h_2|g_2,g_3)\|^2-2\int_H (p(h_2|g_1,g_2)p(h_2|g_2,g_3)/p(h_2|g_2))\,d\eta(h_2)=0$, and thus $p(h_2|g_1,g_2)=p(h_2|g_2,g_3)$ for every g_1, g_3. This means that $p(h_2|g_1,g_2)$ is independent from g_1, which gives (ii).

(ii) implies (i) and (iii). As $p(h_n|g_{n-1},g_n)=p(h_n|g_n,g_{n+1})=p(h_n|g_n)$, the integration of $b(h)$ with respect to $h_1,\dots,h_N$ gives $a(g)$, which means that $p(g)=a(g)$ and thus G is a Markov chain. Moreover, the integration of $b(h)$ with respect to $h_1,\dots,h_{n-1},h_{n+1},\dots,h_N$ gives $p(g,h_n)=a(g)p(h_n|g_n)=p(g)p(h_n|g_n)$. Thus $p(h_n|g)=\frac{p(g,h_n)}{p(g)}=\frac{p(g)p(h_n|g_n)}{p(g)}=p(h_n|g_n)$, which gives (iii).

(iii) implies (ii). We have

$$\begin{aligned}p(h_n|g_n,g_{n-1})&=\int_{H^{N-2}} p(g_1,\dots,g_{n-2},g_{n+1},\dots,g_N)\\&\quad\times p(h_n|g_1,\dots,g_N)\,d\gamma(g_1)\dots,d\gamma(g_{n-2})\\&\quad\times d\gamma(g_{n+1})\,d\gamma(g_N)\\&=\int_{H^{N-2}} p(g_1,\dots,g_{n-2},g_{n+1},\dots,g_N)\\&\quad\times p(h_n|g_n)\,d\gamma(g_1)\dots,d\gamma(g_{n-2})\,d\gamma(g_{n+1})\,d\gamma(g_N)\\&=p(h_n|g_n)\int_{H^{N-2}} p(g_1,\dots,g_{n-2},g_{n+1},\dots,g_N)\\&\quad\times d\gamma(g_1)\dots,d\gamma(g_{n-2})\,d\gamma(g_{n+1})\,d\gamma(g_N)=p(h_n|g_n),\end{aligned}$$

which ends the proof.

References

[1] B. Ait-el-Fquih, F. Desbouvries, Kalman filtering in triplet Markov chains, IEEE Transactions on Signal Processing 54 (8) (2006) 2957–2963.
[2] 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.
[3] Y.M. Bishop, S.E. Fienberg, P.W. Holland, in: Discrete Multivariate Analysis, MIT Press, Cambridge, 1975.
[4] P. Bouthemy, C. Hardouin, G. Piriou, J.-F. Yao, Mixed-state auto-models and motion texture modelling, Journal of Mathematical Imaging and Vision 25 (3) (2006) 387–402.
[5] S. Bricq, C. Collet, J.-P. Armspach, Unifying framework for multi-modal brain MRI segmentation based on hidden Markov chains, Medical Image Analysis 12 (6) (2008) 639–652.
[6] P. Brodatz, in: Textures: A Photographic Album for Artists and Designers, Dover Publications, New York, 1966.
[7] N. Brunel, W. Pieczynski, Unsupervised signal restoration using hidden Markov chains with copulas, Signal Processing 85 (12) (2005) 2304–2315.
[8] N. Brunel, J. Lapuyade-Lahorgue, W. Pieczynski, Modeling and unsupervised classification of multivariate hidden Markov chains with copulas, IEEE Transactions on Automatic Control 55 (2) (2010) 338–349.

[9] O. Cappé, E. Moulines, T. Ryden, in: Inference in Hidden Markov Models, Springer, Series in Statistics, 2005.
[10] C. Carincotte, S. Derrode, G. Sicot, J.-M. Boucher, Unsupervised image segmentation based on a new fuzzy HMC model, ICASSP'04, Montreal, Canada, May 17–21, 2004.
[11] C Carincotte, S. Derrode, S. Bourennane, Unsupervised change detection on SAR images using fuzzy hidden Markov chains, IEEE Transactions on Geosciences and Remote Sensing 44 (2) (2006) 432–441.
[12] G. Celeux, D. Chauveau, J. Diebolt, Stochastic versions of the EM algorithm: an experimental study in the mixture case, Journal of Statistical Computation and Simulation 55 (4) (1996) 287–314.
[13] T. Crivelli, B. Cernuschi-Frias, P. Bouthemy, J. F. Yao, Temporal modelling of motion textures with mixed-states Markov chains, ICASSP 2008, March 30–April 04, Las Vegas, USA, 2008, pp. 881–884.
[14] M. Davy, A. Doucet, Copulas: a new insight into positive time-frequency distributions, IEEE Signal Processing Letters 10 (7) (2003) 215–218.
[15] Y. Delignon, W. Pieczynski, Modeling non Rayleigh speckle distribution in SAR images, IEEE Transactions on Geoscience and Remote Sensing 40 (6) (2002) 1430–1435.
[16] O. D'Hondt, C. Lopez-Martinez, L. Ferro-Famil, E. Pottier, Spatially nonstationary anisotropic texture analysis in SAR images, IEEE Transactions on Geoscience and Remote Sensing 45 (12) (2007) 3905–3917.
[17] J.-P. Delmas, An equivalence of the EM and ICE algorithm for exponential family, IEEE Transactions on Signal Processing 45 (10) (1997) 2613–2615.
[18] S. Derrode, G. Mercier, Unsupervised multiscale oil slick segmentation from SAR images using a vector HMC model, Pattern Recognition 40 (3) (2007) 1135–1147.
[19] S. Derrode, W. Pieczynski, Signal and image segmentation using pairwise Markov chains, IEEE Transactions on Signal Processing 52 (9) (2004) 2477–2489.
[20] P. Embrechts, F. Lindskog, A. McNeil, Modeling dependence with copulas, ⟨http://www.math.uiowa.edu/∼lwang/finance/DependenceWithCopulas.pdf⟩.
[21] Y. Ephraim, Hidden Markov processes, IEEE Transactions on Information Theory 48 (6) (2002) 1518–1569.
[22] 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.
[23] 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.
[24] M. Hatt, F. Lamare, N. Boussion, A. Turzo, C. Collet, F. Salzenstein, et al., Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET, Physics in Medicine and Biology 52 (2007) 3467–3491.
[25] F.V. Jensen, in: An Introduction to Bayesian Networks, UCL Press, 2000.
[26] R.W. Katz, On some criteria for estimating the order of a Markov chain, Technometrics 23 (1981) 243–249.
[27] W. Ketchantang, S. Derrode, L. Martin, S. Bourennane, Pearson-based mixture model for color object tracking, Machine Vision and Applications 19 (5–6) (2008).
[28] T. Koski, in: Hidden Markov Models for Bioinformatics, Kluwer Academic Publishers, 2001.
[29] P. Lanchantin, W. Pieczynski, Unsupervised non-stationary image segmentation using triplet Markov chains, In: Proceedings of the Advanced Concepts for Intelligent Vision Systems (ACIVS 04), August 31–September 3, Brussels, Belgium, 2004.
[30] P. Lanchantin, W. Pieczynski, Unsupervised restoration of hidden non stationary Markov chain using evidential priors, IEEE Transactions on Signal Processing 53 (8) (2005) 3091–3098.
[31] P. Lanchantin, J. Lapuyade-Lahorgue, W. Pieczynski, Unsupervised segmentation of triplet Markov chains with long-memory noise, Signal Processing 88 (5) (2008) 1134–1151.
[32] J. Lapuyade-Lahorgue, W. Pieczynski, Unsupervised segmentation of non-stationary hidden Markov chains with copulas, In: Proceedings of the Fourth World Conference of the International Association for Statistical Computing (IASC 2008), December 5–8, Yokohama, Japan, 2008.
[33] S. Le Cam, F. Salzenstein, C. Collet, Fuzzy pairwise Markov chain to segment correlated noisy data, Signal Processing 88 (10) (2008) 2526–2541.
[34] S. Le Cam, A. Belghith, C. Collet, F. Salzenstein, Wheezing sounds detection using multivariate generalized gaussian distributions, In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009), April 19–24, Taipei, Taiwan, pp. 541–544, 2009.
[35] G.J. McLachlan, T. Krishnan, in: EM Algorithm and Extensions, Wiley, Series in Probability and Statistics, 1997.
[36] G. Mercier, S. Derrode, M. Lennon, Hyperspectral image segmentation with Markov chain model, In: Proceedings of IGARSS 2003 6 (2003) 3766–3768.
[37] G. Moser, J. Zerubia, S.B. Serpico, SAR amplitude probability density function estimation based on a generalized Gaussian model, IEEE Transactions on Image Processing 15 (6) (2006) 1429–1442.
[38] S. Nadarajah, S. Kotz, Intensity models for non-Rayleigh speckle distributions, International Journal of Remote Sensing 29 (2) (2008) 529–541.
[39] R.B. Nelsen., in: An introduction to Copulas. Number 139 in Lecture Notes in Statistics, Springer-Verlag, 1998.
[40] W. Pieczynski, C. Hulard, T. Veit, Triplet Markov chains in hidden signal restoration, In: Proceedings of the SPIE's International Symposium on Remote Sensing, September 22–27, Crete, Greece, 2002.
[41] W. Pieczynski, Pairwise Markov chains, IEEE Transactions on Pattern Analysis and Machine 25 (5) (2003) 634–639.
[42] W. Pieczynski, D. Benboudjema, Multisensor triplet Markov fields and theory of evidence, Image and Vision Computing 24 (1) (2006) 61–69.
[43] W. Pieczynski, Multisensor triplet Markov chains and theory of evidence, International Journal of Approximate Reasoning 45 (1) (2007) 1–16.
[44] W. Pieczynski, Exact calculation of optimal filter in semi-Markov switching model, In: Proceedings of the Fourth World Conference of the International Association for Statistical Computing (IASC 2008), December 5–8, Yokohama, Japan, 2008.
[45] 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.
[46] W. Pieczynski, Statistical image segmentation, Machine Graphics and Vision 1 (1–2) (1992) 261–268.
[47] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE 77 (2) (1989) 257–286.
[48] S. Ruan, C. Jaggi, J. Xue, J. Fadili, D. Bloyet, Brain tissue classification of magnetic resonance images using partial volume modeling, IEEE Transactions on Medical Imaging 19 (12) (2000) 1179–1187.
[49] S. Ruan, B. Moretti, J. Fadili, D. Bloyet, Fuzzy Markovian segmentation in application of magnetic resonance images, Computer Vision and Image Understanding 85 (2002) 54–69.
[50] F. Salzenstein, W. Pieczynski, Parameter estimation in hidden fuzzy Markov random fields and image segmentation, Graphical Models and Image Processing 59 (4) (1997) 205–220.
[51] F. Salzenstein, C. Collet, Fuzzy Markov random fields versus chains for multispectral image segmentation, IEEE Transactions on Pattern Analysis and Machine 28 (11) (2006) 1753–1767.
[52] F. Salzenstein, C. Collet, S. Le Cam, M. Hatt, Non stationary fuzzy Markov chains, Pattern Recognition Letters 28 (16) (2007) 2201–2208.
[53] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton NJ, 1976.
[54] B. Sin, J.H. Kim, Nonstationary Markov models, Signal Processing 46 (1) (1995) 31–46.
[55] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, International Journal of Approximate Reasoning 9 (1993) 1–35.