Unsupervised segmentation of non-stationary hidden Markov chains with copulas

Jérôme Lapuyade-Lahorgue[1] Wojciech Pieczynski[1]
[1]Telecom SudParis, 91000 Evry, FRANCE.
E-mail(1): jerome.lapuyade@it-sudparis.eu
E-mail(2): wojciech.pieczynski@it-sudparis.eu

Keywords: Hidden Markov Models, Unsupervised Segmentation, Non Stationarity, Triplet Markov Chains, Copulas.

1 Introduction

The Hidden Markov Chains (HMC), also called Hidden Markov Models, is a well known and widely used model. Its applications covers numerous fields, including Acoustics, Biosciences, Climatology, Ecology, Control, Communications, Econometrics and Finance, Handwriting and Text Recognition, Image Processing and Computer Vision, Signal Processing, or others : hundreds of papers are written on the subject each year. Rich bibliography can be found in recent books (Cappé and al. (2005), Koski (2001)), or tutorial papers (Ephraim and Marhav (2002), Willsky (2002). The HMC is a couple of stochastic processes $Z=(X,Y)=(X_n,Y_n)_{n=1}^N$ in which X is hidden and Y is observable. We will consider in this paper that each X_n takes its values in a finite set of classes $\Omega=\{\omega_1,\dots,\omega_K\}$ and each Y_n takes its values in R. The distribution of Z is given by

$$p(z)=p(z_1)\prod_{n=2}^{N}p(z_n|z_{n-1}) \tag{1}$$

with

$$p(z_n|z_{n-1})=p(x_n|x_{n-1})p(y_n|x_n)\,. \tag{2}$$

It allows one to estimate a realization $X=x$ from the observed data $Y=y$ in different Bayesian ways. Then HMC have been generalized to "Pairwise Markov Chains" (PMC) (Pieczynski (2003)) in which $p(z_n|z_{n-1})$ are of any form, and it has been showed in (Derrode and Pieczynski (2004)) that such an extension can significantly improve the efficiency of Bayesian segmentation. Subsequently, PMC have been extended to "Triplet Markov Chains" (TMC) in which one adds a third chain U and one assumes that the triplet $T=(X,U,Y)$ is a Markov chain. TMC form a very rich family of particular models (Pieczynski and Desbouvries (2005)); in particular, TMC can be used to deal with non stationary HMC (Lanchantin and Pieczynski (2004)). Otherwise, copulas are a simple and efficient tool to model the dependence between two random variables (Nelsen (1998)), and their recent introduction in PMC results in a very rich set of possibilities of modeling the noise (Brunel and Pieczynski (2005)). Such models are of interest in the situations where the noise is correlated and is Gaussian, as, for example, in radar image processing context (Delignon and Pieczynski (2002), Nadarajah and Kotz (2008)).

The aim of this paper is simultaneously to use these two ideas and apply TMC with copulas to deal with non stationary chains hidden by non necessarily Gaussian and correlated noise. We also propose a parameter estimation method based on the general "Iterative Conditional Estimation" (ICE) principle, which has already been successfully used in the context of PMC in (Derrode and Pieczynski (2004)). Some experiments of unsupervised Bayesian segmentation of non stationary hidden random chains with non Gaussian and correlated noise are also presented and discussed.

2 Hidden non stationary chains and triplet ones

Let $Z = (X, Y) = (X_n, Y_n)_{n=1}^N$ be a chain as above, with Y observable and X hidden. Let us assume that Z is not stationary in that the transitions $p(z_n | z_{n-1})$ depend on n. However, we assume that there are only a finite number M of possible distributions for each $p(z_n | z_{n-1})$. Such a situation can be modelled by a stationary "Triplet Markov Chain": one introduces a third random chain $U = (U_n)_{n=1}^N$, with each U_n taking its values in a finite set $\Lambda = \{\lambda_1, ..., \lambda_M\}$, and one assumes that the triplet chain $T = (X, U, Y)$ is a Markov chain. As described in (Lanchantin and Pieczynski (2004)), it is then possible to search simultaneously U and X from the observed Y.

Let us shortly recall how the Bayesian segmentation runs in the context of TMC.

The distribution of $T = (X, U, Y)$ is thus given by $p(t) = p(t_1)p(t_2|t_1)...p(t_N|t_{N-1})$. To simplify notation, we introduce the process $V = (V_n)_{n=1}^N = (X_n, U_n)_{n=1}^N$. Therefore each V_n takes its values in $\Omega \times \Lambda$, and $T = (V, Y)$ is a Markov chain. The chain (V, Y) being a PMC, we can introduce the "Forward" quantities $\alpha(v_n) = p(v_n, y_1, ..., y_n)$, and the "Backward" quantities $\beta(v_n) = p(y_{n+1}, ..., y_N | v_n)$, which are both calculated by the following forward and backward recursions (see (Pieczynski (2003) or (Derrode and Pieczynski (2004)):

$$\alpha(v_1) = p(v_1, y_1);\ \alpha(v_{n+1}) = \sum_{v_n \in \Omega \times \Lambda} \alpha(v_n) p(v_{n+1}, y_{n+1} | v_n, y_n) \tag{3}$$

$$\beta(v_N) = 1;\ \beta(v_n) = \sum_{v_{n+1} \in \Omega \times \Lambda} \beta(v_{n+1}) p(v_{n+1}, y_{n+1} | v_n, y_n) \tag{4}$$

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

$$p(v_n | y) \propto \alpha(v_n)\beta(v_n) \tag{5}$$

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

$$p(v_{n+1} | v_n, y_1, ..., y_N) \propto p(v_{n+1}, y_{n+1} | v_n, y_n)\beta(v_{n+1}) \tag{6}$$

Having calculated $p(v_n | y)$, we can then compute

$$p(x_n | y) = \sum_{u_n \in \Lambda} p(v_n | y), \tag{7}$$

Therefore, one can easily calculate $p(x_n | y)$, which makes the use of Bayesian Maximum Posterior Mode (MPM) segmentation method $\hat{s}_{MPM}(y_1, ..., y_N) = (\hat{x}_1, ..., \hat{x}_N)$ possible, with $\hat{x}_n = \arg\max_{x_n \in \Omega} p(x_n | y)$.

Otherwise, we can also calculate $p(u_n|y) = \sum_{x_n \in \Omega} p(v_n|y)$ which thus makes also possible the use of MPM to estimate the different stationarities.

3 Copulas

Let $h(y_1, y_2)$ be a probability density on R^2, $H(y_1, y_2)$ the associated cumulative distribution function (cdf), $h_1(y_1)$ and $h_2(y_2)$ the marginal densities, and $H_1(y_1)$, $H_2(y_2)$ the cdf associated with them. Then there exists a function C defined on $[0,1]^2$, called "copula" such that

$$H(y_1, y_2) = C(H_1(y_1), H_2(y_2)), \tag{8}$$

see (Nelsen (1998)). Assuming C derivable, which will made in the following, taking the derivative of (8) with respect to y_1, y_2, and introducing $c(u,v) = \frac{\partial\partial C(u,v)}{\partial u \partial v}$, we have

$$h(y_1, y_2) = h_1(y_1) h_2(y_2) c(H_1(y_1), H_2(y_2)) \tag{9}$$

Otherwise, it is then possible to show that a copula is a pdf on $[0,1]^2$ such that the marginal distributions are both uniform distributions on $[0,1]$. Conversely, having H_1, H_2 and C pdf on $[0,1]^2$ with marginal distributions uniform on $[0,1]$, we can use (8) to define a pdf H. Therefore, a given H cdf on R^2 defines a copula C with (8), and this copula can also be used to define any another H' from any another H_1', H_2'.
According to (9), c is defined from h with

$$c(y_1, y_2) = \frac{h(H_1^{-1}(y_1), H_2^{-1}(y_2))}{h_1(H_1^{-1}(y_1)) h_2(H_2^{-1}(y_2))} \tag{10}$$

There are numerous possibilities of defining particular copulas (Nelsen (1998)). Let us specify two of them, which will be used in experiments below.

(i) Gaussian copula
Let $h(y_1, y_2)$ be the density of a Gaussian vector with correlation ρ and marginal distributions having null means and variances equal to one. Applying (10) we find

$$c_G(y_1, y_2) = |\Sigma|^{-\frac{1}{2}} \exp[-\frac{1}{2} \zeta(y_1, y_2)(\Sigma^{-1} - Id)^t \zeta(y_1, y_2)] \tag{12}$$

with $\zeta(y_1, y_2) = (H_1^{-1}(y_1), H_2^{-1}(y_2))$ and $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. So, a bivariate Gaussian copula is defined by just one parameter, which is the correlation ρ.

(ii) Student copula
Let us recall that the Student distribution "with ν degrees of freedom" (also called "T law") on R^2 is

given by the density

$$g(y_1,y_2)=\frac{\Gamma(\frac{\nu+2}{2})|\Sigma|^{-\frac{1}{2}}}{\pi\nu\Gamma(\frac{\nu}{2})}(1+\frac{2q(y_1,y_2)}{\nu})^{\frac{-(\nu+2)}{2}} \tag{13}$$

with, for $y=(y_1,y_2)$, $q(y)=\frac{1}{2}y'\Sigma^{-1}y$, where y' is the transpose of the vector y. Let g_1, g_2 be the marginal distributions of (13) , and G_1, G_2 the associated cdf functions. According to (10), the associated copula, called "Student copula", is given by:

$$c_S(y_1,y_2)=|\Sigma|^{-\frac{1}{2}}\frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})^2}\frac{(1+\frac{1}{\nu}\zeta(y_1,y_2)'\Sigma^{-1}\zeta(y_1,y_2))^{\frac{-(\nu+2)}{2}}}{(1+\frac{G_1^{-1}(y_1)}{\nu})^{\frac{-(\nu+1)}{2}}(1+\frac{G_2^{-1}(y_2)}{\nu})^{\frac{-(\nu+1)}{2}}}, \tag{14}$$

with $\Sigma=\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$ the correlation matrix.

where $\zeta(y_1,y_2)=(G_1^{-1}(y_1),G_2^{-1}(y_2))$.

These two copulas can be used to define different correlated distributions on R^2. For example, we can take two Gaussian marginal distributions g_1, g_2 and the Student copula c_S. Then we have two correlated Gaussian real variables, but the couple is not a Gaussian vector. As we will see in the sixth section, such distribution used in the context of triplet Markov chains can give different results than the Gaussian vector. We can also make a converse choice by taking two marginal Student distributions and the Gaussian copula. Of course, it is also possible to take one marginal Gaussian, the other Student, and the copula Gaussian or Student.

4 Copulas in Triplet Markov Chains

Let $T=(X,U,Y)$ be a TMC as above, and let $T=(V,Y)$ be the associated PMC with $V=(X,U)$. Let us assume that $T=(V,Y)$ is stationary and reversible, which means that $p(t_n,t_{n+1})$ does not depend on n, and $p(t_n=a,t_{n+1}=b)=p(t_n=b,t_{n+1}=a)$ for every triplets a, b. Its distribution is defined by

$$p(t_1,t_2)=p(v_1,y_1,v_2,y_2)=p(v_1,v_2)p(y_1,y_2|v_1,v_2), \tag{15}$$

Thus the distribution of the whole chain $T=(X,U,Y)$ is defined by the discrete distribution $p(v_1,v_2)$ and a finite set of distributions $p(y_1,y_2|v_1,v_2)$ on R^2. According to the previous section, each distribution $p(y_1,y_2|v_1,v_2)$ is defined by its marginal distribution and a copula. To simplify notations, let us set $p_{ij}(y_1,y_2)=p(y_1,y_2|v_1=i,v_2=j)$. Thus we have $(KM)^2$ marginal distributions $p_{ij}(y_1)$

(which are the same that the $(KM)^2$ marginal distributions $p_{ij}(y_2)$ because of the reversibility of T), and $KM(KM-1)/2$ copulas.
Let us consider the following particular model which will be used in experiments presented in the next section. The distribution of is given by

$$p(x,u,y)=p(u_1)p(x_1,u_1)p(y_1,x_1)\times\prod_{n=1}^{N-1}p(u_{n+1}|u_n)p(x_{n+1}|x_n,u_{n+1})p(y_{n+1}|x_n,x_{n+1},y_n) \tag{16}$$

Then it is possible to show that U and (X,U) are Markov chains, but X is not necessarily a Markov chain. Otherwise, the chains U and Y are independent conditionally on X. Recalling that $(v_1,v_2)=(x_1,u_1,x_2,u_2)$, the particular transitions () imply

$$p(y_1,y_2|v_1,v_2)=p(y_1,y_2|x_1,x_2) \tag{17}$$

5 Parameter estimation with Iterative Conditional Estimation (ICE)

The general principle of the "Iterative conditional estimation" (ICE) that we will use in experiments below and which is an alternative method to the well known "expectation-Maximization" (EM) method, is the following. Let us consider two random processes V, Y whose distribution depends on a parameter $\theta=(\theta_1,...,\theta_m)$. The problem is to estimate θ from Y. The aim of the well known EM method is to iteratively maximize the likelihood $p(y|\theta)$. The ICE principle we propose to use is somewhat different from EM and is often easier to perform in complex situations. To simplify, let $\theta\in R$. Let us assume that we have at our disposal some estimator $\hat{\theta}(V,Y)$, which has a good efficiency measured by the mean square error $E_\theta[(\theta-\hat{\theta}(V,Y))^2]$. Let us notice that $\hat{\theta}(V,Y)$ is the Maximum Likelihood estimator or not. As V is not available, the idea of ICE is to approximate $\hat{\theta}(V,Y)$ by some function of Y. The best approximation, in the sense of mean square error, is the conditional expectation $\tilde{\theta}(Y)=E_\theta[\hat{\theta}(V,Y)|Y]$. Of course, $\tilde{\theta}$ is no longer an estimator because it does depend on θ; however, the principle leads to an iterative method whose general run resembles to the EM method.
More precisely, ICE is an iterative method based on the following principle. Let $\hat{\theta}(V,Y)$ be an estimator of θ from complete data and let us assume that we can sample realizations of V according to $p(v|y,\theta)$. ICE runs as follows:
(i) take an initial value θ^0;
(ii) using $Y=y$ and the current value of the parameter θ^q, 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 feasible;
(iii) for other components θ_i, simulate v_1^q, …, v_l^q independent realizations of V according to $p(v|y,\theta^q)$ and put $\theta_i^{q+1}=[\hat{\theta}(v_1^q,y)+...+\hat{\theta}(v_l^q,y)]/l$.
Let us notice that in (iii) one simply approximates, using the large numbers law, the expectation by the empirical mean. In principle, the greater is l the better is the approximation; however, in practice taking small l, or even $l=1$, has little influence on the final estimation results. Otherwise, we will see that in the problem we are concerned with (ii) is computable for the components θ_i defining the distribution $p(v)$, while it is not computable for the components θ_i defining $p(y|v)$.
We see that ICE is applicable under two very slight hypotheses: existence of an estimator $\hat{\theta}(V,Y)$ from the complete data, and the ability of simulating V according to $p(v|y,\theta)$. The first hypothesis is

not really a constraint because if we are not able to estimate θ from complete data (v, y), there is no point in searching an estimator from incomplete ones y. The second hypothesis is always verified for TMC we consider in this paper, as the distribution $p(v|y)$ is a Markov one.
Finally, we can use ICE for estimating TMC with copulas once we have an estimator $\hat{\theta}(V,Y)$ from the complete data. Such an estimator depends on the forms of the different copulas used in the model; we will see one of them in the next section.

6 Experiments

Let us consider $\Omega = \{\omega_1, \omega_2\}$ and $\Lambda = \{\lambda_1, \lambda_2\}$. Thus we have two classes and two different stationarities. All marginal distributions considered in this section will be Gaussian; then we will consider two copulas : Gaussian copula and Student one.
According to the ICE principle, we have to specify estimators from the complete data (x,u,y). We first estimate the Gaussian margins using the very classical formulas $m_{\omega_j}^{q+1} = [\sum_{n=1}^{N} y_n I(x_n = \omega_j)]/[\sum_{n=1}^{N} I(x_n = \omega_j)]$ and $(\sigma_{\omega_j}^{q+1})^2 = [\sum_{n=1}^{N} (y_n - m^{q+1}{}_{\omega_j})^2 I(x_n = \omega_j)]/[\sum_{n=1}^{N} I(x_n = \omega_j)]$.
Let $F_{\omega_j}^{q+1}$ be the cdf of the Gaussian distribution with mean $m_{\omega_j}^{q+1}$ and variance $(\sigma_{\omega_j}^{q+1})^2$, ϕ be the cdf of the Gaussian centered distribution on R^2, and ϕ_ν be the cdf of Student centered distribution on R^2, with the parameter ν. If the copula defining $p(y_n, y_{n+1}|x_n, x_{n+1})$ is Gaussian (Student, respectively), we consider $z_n = \phi^{-1} \circ F_{\omega_j}^{q+1}(y_n)$ ($z_n = \phi_\nu^{-1} \circ F_{\omega_j}^{q+1}(y_n)$, respectively). Then $p(z_n, z_{n+1}|x_n, x_{n+1})$ is a Gaussian (Student, respectively) distribution and it is sufficient to estimate the correlation $\rho_{x_n, x_{n+1}}$, which is made classically from $(z_1, \ldots, z_N)$ and $(x_1, \ldots, x_N)$.
Let us assume $p(x_n = \omega_i, u_n = \lambda_k, x_{n+1} = \omega_j, u_{n+1} = \lambda_l) = p(x_n = \omega_j, u_n = \lambda_l, x_{n+1} = \omega_i, u_{n+1} = \lambda_k)$, and

$$p(y_n, y_{n+1}|x_n = \omega_i, u_n = \lambda_k, x_{n+1} = \omega_j, u_{n+1} = \lambda_l) = p(y_{n+1}, y_n|x_n = \omega_j, u_n = \lambda_l, x_{n+1} = \omega_i, u_{n+1} = \lambda_k).$$

In the two series of experiments below we take the following values:

$$p(u_n = \lambda_1, u_{n+1} = \lambda_1) = p(u_n = \lambda_2, u_{n+1} = \lambda_2) = 0.49995$$
$$p(u_n = \lambda_1, u_{n+1} = \lambda_2) = p(u_n = \lambda_2, u_{n+1} = \lambda_1) = 0.00005,$$
$$p(x_n = \omega_1, x_{n+1} = \omega_1|u_{n+1} = \lambda_1) = p(x_n = \omega_2, x_{n+1} = \omega_2|u_{n+1} = \lambda_1) = 0.495,$$
$$p(x_n = \omega_1, x_{n+1} = \omega_2|u_{n+1} = \lambda_1) = p(x_n = \omega_2, x_{n+1} = \omega_1|u_{n+1} = \lambda_1) = 0.005,$$
$$p(x_n = \omega_1, x_{n+1} = \omega_1|u_{n+1} = \lambda_2) = p(x_n = \omega_2, x_{n+1} = \omega_2|u_{n+1} = \lambda_2) = 0.45,$$
$$p(x_n = \omega_1, x_{n+1} = \omega_2|u_{n+1} = \lambda_2) = p(x_n = \omega_2, x_{n+1} = \omega_1|u_{n+1} = \lambda_2) = 0.05.$$

Otherwise, we have $p(y_n, y_{n+1}|x_n, x_{n+1}) = p(y_n|x_n)p(y_{n+1}|x_{n+1})c_{x_n, x_{n+1}}(F_{x_n}(y_n), F_{x_{n+1}}(y_{n+1}))$, where $F_{x_n}(y_n)$ is cdf associated with $p(y_n|x_n)$ and $c_{x_n, x_{n+1}}$ is the associated copula. Then $p(y_{n+1}|x_n, x_{n+1}, y_n) = p(y_{n+1}|x_{n+1})c_{x_n, x_{n+1}}(F_{x_n}(y_n), F_{x_{n+1}}(y_{n+1}))$
In both series $p(y_n|x_n = \omega_1)$ is Gaussian, $N(0,1)$, and $p(y_n|x_n = \omega_2)$ is Gaussian $N(2,1)$. In the first experiment the copula is Gaussian, and it is a Student copula in the second one. The parameter $\nu = 10$ in the Student copula is assumed to be known.
The sampled realization $(X,U,Y) = (x,u,y)$ and the Bayesian MPM segmentation results based on the true model and the true parameters of the first series, in which we use the correlations $\rho_{\omega_1,\omega_1} = \rho_{\omega_2,\omega_2} = 0.9$ et $\rho_{\omega_1,\omega_2} = \rho_{\omega_2,\omega_1} = 0$, is presented in Figure 1. The values $\hat{X} = \hat{x}$ and $\hat{U} = \hat{u}$

are optimal, and thus they are the reference ones. Then $Y = y$ is segmented in unsupervised way using the true Gaussian copula, and the wrong Student copula. The aim of this experiment is study whether choosing the right copula is of importance or not. The results are presented in Figure 2 and we see that using Student copula instead of the right Gaussian one significantly biases the debases of the estimation of $X = x$ (error ratio of $\tau = 38{,}4\%$ instead of $\tau = 9{,}4\%$), while it moderately debases the quality of the estimation of $U = u$ (error ratio of $\tau = 0{,}7\%$ instead of $\tau = 0{,}5\%$). We also notice, according to the parameter estimations results presented in Tables 1 and 2, that the noise parameters are much better estimated when the true Gaussian copula based model is used.
To visualize the results as being images, we convert the mono-dimensional sequence into a bi-dimensional set of pixels via Hilbert-Peano scan, as specified in (Fjortoft *et al*. (2003)).

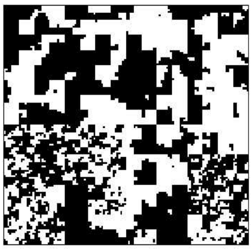
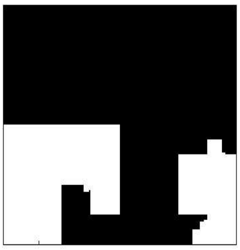
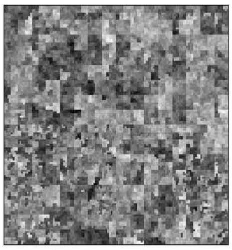
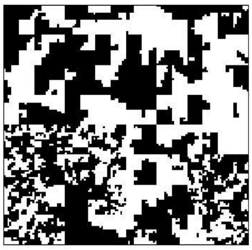
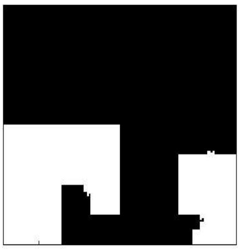
		
$X = x$	$U = u$	$Y = y$
		
$\hat{X} = \hat{x}$ (TMC), $\tau = 5{,}49\%$	$\hat{U} = \hat{u}$ (TMC), $\tau = 0{,}45\%$	

Figure 1 : Simulated $(X, U, Y) = (x, u, y)$ with Gaussian copula, and real model and real parameters based MPM segmentation segmentation.

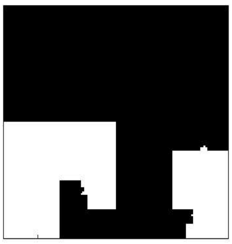
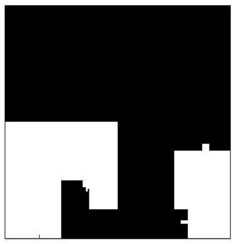
			
$\hat{X} = \hat{x}$ (GC), $\tau = 9{,}4\%$	$\hat{U} = \hat{u}$ (GC), $\tau = 0{,}5\%$	$\hat{X} = \hat{x}$ (SC), $\tau = 38{,}4\%$	$\hat{U} = \hat{u}$ (SC), $\tau = 0{,}7\%$

Figure 2 : Unsupervised segmentation of $Y = y$ in Figure 1, GC: Gaussian copula, SC: Student copula.

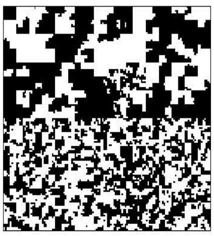
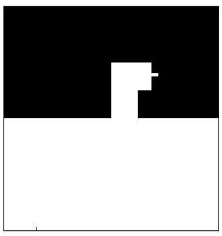
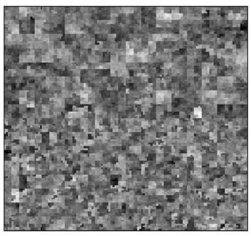
		GC		SC	
		ω_1	ω_2	ω_1	ω_2
m		-0.05	2.04	0.41	1.75
σ^2		0.91	0.96	1.63	1.04
ρ	ω_1	0.89	-0.05	0.93	0.18
	ω_2	-0.05	0.92	0.18	0.78

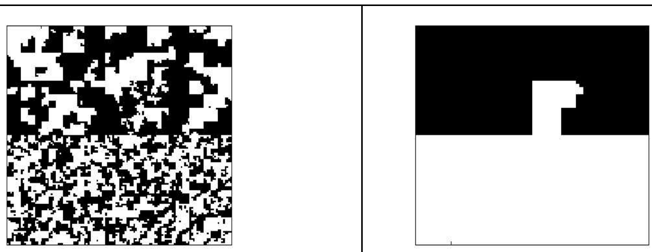
Table 1 : Parameter's estimation of the observation distribution

		GC				SC			
		λ_1		λ_2		λ_1		λ_2	
$p(u_n = \lambda_k, u_{n+1} = \lambda_l)$	λ_1	0.4999		0.0001		0.4999		0.0001	
	λ_2	0.0001		0.4999		0.0001		0.4999	
$p(x_n = \omega_i, x_{n+1} = \omega_j \mid u_{n+1} = \lambda_l)$		ω_1	ω_2	ω_1	ω_2	ω_1	ω_2	ω_1	ω_2
	ω_1	0.49	0.01	0.45	0.05	0.49	0.01	0.48	0.02
	ω_2	0.01	0.49	0.05	0.45	0.01	0.49	0.02	0.48

Table 2 : Parameter's estimation of the distribution $p(x,u)$

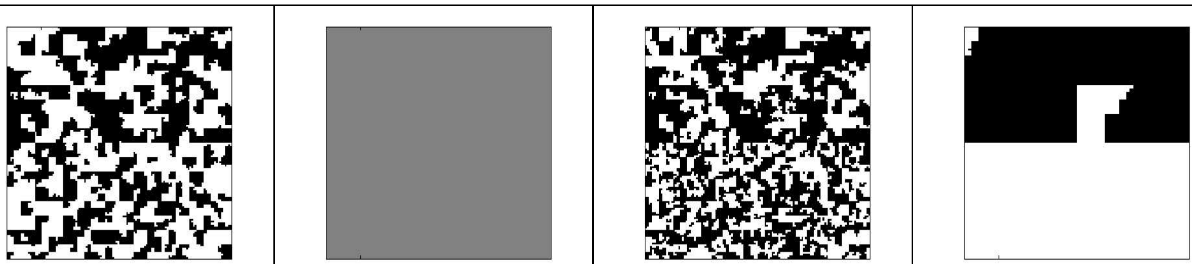
The sampled realization $(X,U,Y) = (x,u,y)$ corresponding to the second experiment, where the TMC distribution is based on Student copula, is presented in Figure 3. In the same figure are given the Bayesian MPM segmentation results based on the true model and the true parameters, in which we use the same correlations $\rho_{\omega_1,\omega_1} = \rho_{\omega_2,\omega_2} = 0.9$ et $\rho_{\omega_1,\omega_2} = \rho_{\omega_2,\omega_1} = 0$. The values $\hat{X} = \hat{x}$ and $\hat{U} = \hat{u}$ are optimal, and thus they are the reference ones. Then $Y = y$ is segmented in unsupervised way using the wrong Gaussian copula, and the right Student copula. As above, the aim of this experiment is study whether choosing the right copula is of importance or not and the difference is that the copulas have been inverted. The results are presented in Figure 4 and we see that using Gaussian copula instead of the right Student one significantly debases the quality of the estimation of $U = u$ (error ratio of $\tau = 54{,}1\%$ instead of $\tau = 0{,}8\%$), while it weakly debases the quality of the estimation of $X = x$ (error ratio of $\tau = 26{,}0\%$ instead of $\tau = 21{,}9\%$). We also notice, according to the parameter estimations results presented in Tables 3 and 4, that the noise parameters are much better estimated when the true Gaussian copula based model is used.

		
$X = x$	$U = u$	$Y = y$



$\hat{X}=\hat{x}$ (TMC), $\tau=15{,}4\%$	$\hat{U}=\hat{u}$ (TMC), $\tau=0{,}7\%$	

Figure 3 : Simulated $(X,U,Y)=(x,u,y)$ with Student copula, and real model and real parameters based MPM segmentation.



$\hat{X}=\hat{x}$ (GC) $\tau=26{,}0\%$	$\hat{U}=\hat{u}$ (GC), $\tau=54{,}1\%$	$\hat{X}=\hat{x}$ (SC), $\tau=21{,}9\%$	$\hat{U}=\hat{u}$ (SC), $\tau=0{,}8\%$

Figure 4 : Unsupervised segmentation, GC: Gaussian copula, SC: Student copula

		GC		SC	
		ω_1	ω_2	ω_1	ω_2
m		0.36	1.61	-0.13	1.94
σ^2		2.66	2.88	0.91	1.18
ρ	ω_1	0.92	-0.01	0.89	0.01
	ω_2	-0.01	0.94	0.01	0.91

Table 3 : Parameter's estimation of the observation distribution

		GC				SC			
		λ_1		λ_2		λ_1		λ_2	
$p(u_n=\lambda_k,u_{n+1}=\lambda_l)$	λ_1	~0.5		~0		0.4999		0.0001	
	λ_2	~0		~0.5		0.0001		0.4999	
$p(x_n=\omega_i,x_{n+1}=\omega_j \mid u_{n+1}=\lambda_l)$		ω_1	ω_2	ω_1	ω_2	ω_1	ω_2	ω_1	ω_2
	ω_1	0.46	0.04	-	-	0.49	0.01	0.45	0.05
	ω_2	0.04	0.46	-	-	0.01	0.49	0.05	0.45

Table 4 : Parameter's estimation of the distribution $p(x,u)$

7 Conclusion

We dealt in this paper with the problem of statistical segmentation of non stationary data hidden with correlated and non Gaussian noise. The first aspect of the problem was dealt with using the recent triplet Markov chains (TMC), as proposed in (Lanchantin and Pieczynski (2004)) in the context of Gaussian noise. The second aspect was dealt with using Copulas, as proposed in (Brunel and Pieczynski (2005)) in the hidden Markov chains (HMC) with correlated noise context. Setting these two ideas together, we arrive at a very rich model and some experiments show that they are workable, even in the unsupervised context.

References

Brunel N. and Pieczynski W., (2005). Unsupervised signal restoration using hidden Markov chains with copulas, *Signal Processing*, 85 (12), 2304-2315, Elsevier.

Cappé O., Moulines E., and Ryden T., (2005). Inference in hidden Markov models, Springer, Series in Statistics.

Delignon Y. and Pieczynski W.,(2002). Modeling non Rayleigh Speckle Distribution in SAR Images, *IEEE Trans. on Geoscience and Remote Sensing*, 40 (6) : 1430-1435, IEEE Computer Society.

Derrode S. and Pieczynski W., (2004). Signal and Image Segmentation using Pairwise Markov Chains, *IEEE Trans. on Signal Processing*, 52 (9) : 2477-2489, IEEE Computer Society.

Ephraim Y. and Merhaw N., (2002). Hidden Markov processes, *IEEE Trans. on Information Theory*, 48 (6), 1518-1569, IEEE Computer Society.

Fjørtoft R., Delignon Y., Pieczynski W., Sigelle M., and Tupin F., (2003). Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields, *IEEE Trans. on Geoscience and Remote Sensing*, 41 (3) : 675-686, IEEE Computer Society.

Koski T. (2001). Hidden Markov models for bioinformatics, Kluwer Academic Publishers.

Lanchantin P. and Pieczynski W., (2004). Unsupervised non stationary image segmentation using triplet Markov chains, Advanced Concepts for Intelligent Vision Systems (ACVIS 04), Aug. 31-Sept. 3, Brussels, Belgium.

JLapuyade-Lahorgue J. and Pieczynski W., (2006). Unsupervised segmentation of hidden semi-Markov non stationary chains, Twenty six International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt) 20 July 8-1306, Paris, France.

Nadarajah S. and Kotz S., (2008). Intensity models for non-Rayleigh speckle distributions, *International Journal of Remote Sensing*, 29 (2) : 529-541, Taylor & Francis.

Nelsen R. B. (1998). An introduction to Copulas. Number 139 in Lecture notes in Statistics, Springer-Verlag.

Pieczynski W., (2003). Pairwise Markov chains, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25 (5) : 634-639, IEEE Computer Society.

Pieczynski, W. (2007). Multisensor triplet Markov chains and theory of evidence, *International Journal of Approximate Reasoning*, 45 (1) : 1-16, Elsevier.

Pieczynski W. and Desbouvries F., (2005). On triplet Markov chains, International Symposium on Applied Stochastic Models and Data Analysis, (ASMDA 2005), May, Brest, France.

Willsky, A. S. (2002). Multiresolution Markov models for signal and image processing, Proceedings of IEEE, 90 (8) : 1396-1458, IEEE Computer Society