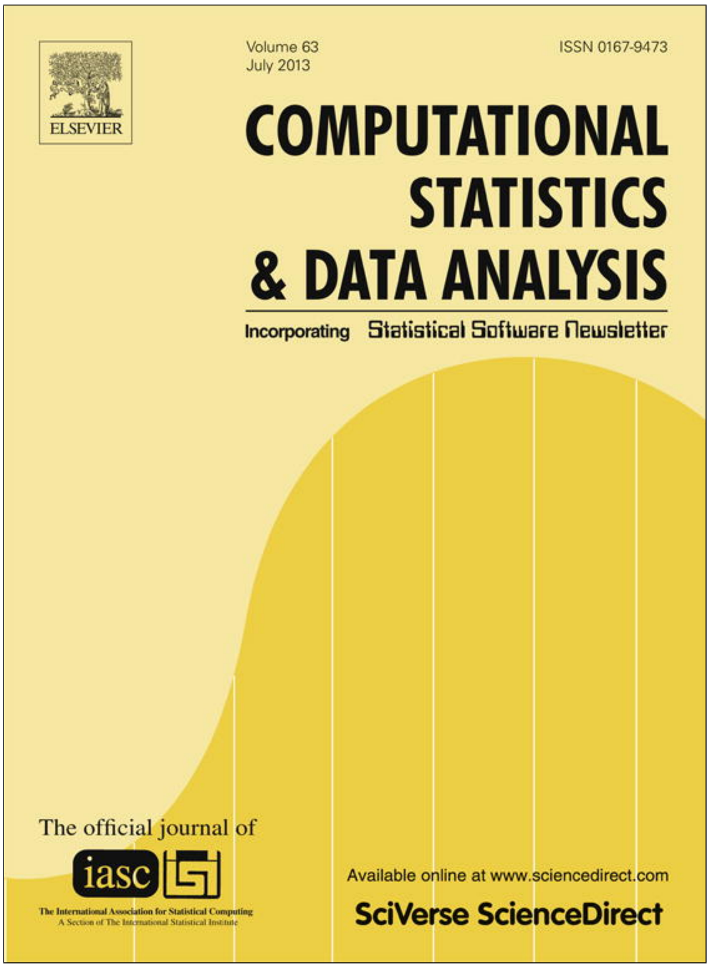

ELSEVIER
Volume 63
July 2013
ISSN 0167-9473
COMPUTATIONAL
STATISTICS
& DATA ANALYSIS
Incorporating Statistical Software Newsletter
The official journal of
iasc
The International Association for Statistical Computing
A Section of The International Statistical Institute
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Computational Statistics and Data Analysis 63 (2013) 81–98

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Unsupervised data classification using pairwise Markov chains with automatic copulas selection

Stéphane Derrode [a,*], Wojciech Pieczynski [b]

[a] Institut Fresnel (CNRS UMR 7249), Aix*Marseilles University and École Centrale Marseille, 38, rue Frédéric Joliot-Curie, 13451 Marseille cedex 20, France
[b] CITI Department (CNRS UMR 5157), TELECOM SudParis, Institut Télécom 9, Rue Charles Fourier, 91011 Evry cedex, France

ARTICLE INFO

Article history:
Received 28 November 2011
Received in revised form 30 January 2013
Accepted 30 January 2013
Available online 8 February 2013

Keywords:
Pairwise Markov chain
Copulas
Iterative conditional estimation

ABSTRACT

The Pairwise Markov Chain (PMC) model assumes the couple of observations and states processes to be a Markov chain. To extend the modeling capability of class-conditional densities involved in the PMC model, copulas are introduced and the influence of their shape on classification error rates is studied. In particular, systematic experiments show that the use of wrong copulas can degrade significantly classification performances. Then an algorithm is presented to identify automatically the right copulas from a finite set of admissible copulas, by extending the general "Iterative Conditional Estimation" (ICE) parameters estimation method to the context considered. The unsupervised segmentation of a radar image illustrates the nice behavior of the algorithm.

1. Introduction

Let $\boldsymbol{x}_{1:N} = (x_1, \ldots, x_N)$ and $\boldsymbol{y}_{1:N} = (y_1, \ldots, y_N)$ be two series of data. Each x_n takes its value in the finite set $\Omega = \{1, \ldots, K\}$ and each y_n in the set of real numbers $\mathbb{R}$. When looking for the unobservable data series $\boldsymbol{x}_{1:N}$ from the observable one $\boldsymbol{y}_{1:N}$, and when there is no deterministic link between them, probability theory provides a rigorous framework to lead to results that are generally effective and sometimes spectacular. The couple $(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N})$ is considered as a realization of two random processes $\boldsymbol{X}_{1:N} = (X_1, \ldots, X_N)$ and $\boldsymbol{Y}_{1:N} = (Y_1, \ldots, Y_N)$ and the stochastic links between the two series are modeled by a law $p(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N})$ of couple $(\boldsymbol{X}_{1:N}, \boldsymbol{Y}_{1:N})$. Despite the lack of deterministic relationship between $\boldsymbol{x}_{1:N}$ and $\boldsymbol{y}_{1:N}$, it is possible to propose optimal methods for finding $\boldsymbol{x}_{1:N}$, in mean or in long-term, when dealing with the problem a "large number" of times.

It is however impossible, when N increases, to consider the general law $p(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N})$ because of the high algorithmic complexity, and we are forced to consider specific laws. Among these the most spread law is the "Hidden Markov Chain" (HMC), which writes

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

This model was later generalized to "Pairwise Markov Chain" (PMC) (Pieczynski, 2003):

$$p(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N}) = p(x_1, y_1)\, p(x_2, y_2 | x_1, y_1) \ldots p(x_N, y_N | x_{N-1}, y_{N-1}). \quad (2)$$

The HMC model is very effective and commonly used, but the PMC model, which allows the same ease of processing as in the HMC, can improve performances significantly, even in an unsupervised way (Derrode and Pieczynski, 2004).

* Corresponding author. Tel.: +33 4 91 05 47 30; fax: +33 4 91 05 45 65.
E-mail addresses: stephane.derrode@centrale-marseille.fr (S. Derrode), wojciech.pieczynski@telecom-sudparis.eu (W. Pieczynski).
URLs: http://www.fresnel.fr/perso/derrode/ (S. Derrode), http://www-public.int-evry.fr/~pieczyn/ (W. Pieczynski).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2013.01.027

Consider a PMC such that the law $p(x_{n-1}, y_{n-1}, x_n, y_n)$ of $(X_{n-1}, Y_{n-1}, X_n, Y_n)$ does not depend on $n = 2, \ldots, N$. Law $p(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N})$ is then entirely characterized by

$$p(x_1, y_1, x_2, y_2) = p(x_1, x_2)\ p(y_1, y_2 | x_1, x_2). \tag{3}$$

This work deals with the modeling of laws $p(y_1, y_2 | x_1, x_2)$ by means of copulas (Nelsen, 2005). Copulas are used for a long time in the field of economy and finance (but without considering Markovianity) (Genest and Mackay, 1986; Nikoloulopoulos and Karlis, 2008; Genest, Rémillard et al., 2009; Genest, Masiello et al., 2009; Ntantamis, 2010), and only more recently in signal processing (Brunel and Pieczynski, 2005; Mercier et al., 2008; Le Cam et al., 2009; Sakji-Nsibi and Benazza-Benyahia, 2009; Stitou et al., 2009; Brunel et al., 2010; Iyengar et al., 2011; Sundaresan and Varshney, 2011).

The first work that combined copulas and Markov model was proposed in Brunel and Pieczynski (2005), where the process $\boldsymbol{X}_{1:N}$ is a Markov chain (such a model is called a HMC "with dependent noise"). In addition, a method for parameters estimation has been proposed, allowing unsupervised processings. Note that both HMC and copulas are known and used for several decades. It may then seem surprising that the two concepts have been considered in the same model only recently. This is likely due to the fact that in traditional models (1) the noise is "independent", which implies that $p(y_1, y_2 | x_1, x_2) = p(y_1 | x_1)\, p(y_2 | x_2)$, and therefore the problem of modeling the dependence of random variables Y_1 and Y_2 conditional on (X_1, X_2) does not arise. However, as discussed in this article, this dependence can have a significant influence on the quality of processings.

In this paper we extend work Brunel and Pieczynski (2005) by considering the problem of generalized mixtures estimation using the "Iterative Conditional Estimation" (ICE) method (Pieczynski, 1992): for all $(i, j) \in \Omega^2$, the copula associated to $p(y_1, y_2 | x_1 = i, x_2 = j)$ is unknown and is automatically searched for in a finite set of eligible copulas, from $\boldsymbol{Y}_{1:N} = \boldsymbol{y}_{1:N}$ only. Incidentally, we show the interest of the Markov models under consideration, by comparing unsupervised estimates of $\boldsymbol{X}_{1:N} = \boldsymbol{x}_{1:N}$ obtained with the structure of a PMC for the law of $(\boldsymbol{X}_{1:N}, \boldsymbol{Y}_{1:N})$ and with the structure of an i.i.d. pairwise mixture model for which pairs $(y_1, y_2), (y_3, y_4), \ldots$ are independent but each of them has the same distribution than in a PMC.

The study is focused on the impact of the choice of copulas in the PMC model, both for supervised and unsupervised data restoration. Also, we will consider in this work that the shapes of margins are known with the aim to not bias the results on copulas. Note that the automatic selection of marginals within a finite set of candidates has been studied in Giordana and Pieczynski (1997), and can be directly plugged in the proposed algorithm.

The experiments performed allow to affirm the importance of choosing the true copula. They also show the effectiveness of the automatic identification method of copulas, based on the Bayesian selection method proposed in Huard et al. (2006).

The remainder of the paper is organized as follows. Section 2 provides a brief overview of the PMC model and its various special cases, and the notion of copula. Section 3 is devoted to highlighting the importance of using the right copula for restoration in both Markovian and non Markovian contexts. The estimation of the generalized mixture, based on an extension of ICE, and the results of several Bayesian unsupervised restorations are proposed in Section 4. Section 5 presents comparative results regarding the segmentation of a radar image with automatic copulas selection. The final section contains conclusions and perspectives.

2. Copulas and PMC

The aim of this section is to present the Pairwise Markov Chain (PMC) model with copulas, including both supervised and unsupervised Bayesian data restoration according to this model.

2.1. PMC basics

Let $\boldsymbol{X}_{1:N} = (X_1, \ldots, X_N)$ and $\boldsymbol{Y}_{1:N} = (Y_1, \ldots, Y_N)$ be two random processes. Each X_n takes its value in the finite set $\Omega = (1, \ldots, K)$ and each Y_n in the set of real numbers $\mathbb{R}$. Let $\boldsymbol{Z}_{1:N} = (Z_1, \ldots, Z_N)$ with $\boldsymbol{Z}_n = (X_n, Y_n)$ for all $n = 1, \ldots, N$.

The process $\boldsymbol{Z}_{1:N}$ is said to be a "Pairwise Markov Chain" (PMC) (Pieczynski, 2003), if it is a Markov chain:

$$p(\boldsymbol{z}_{1:N}) = p(x_1, y_1)\ p(x_2, y_2 | x_1, y_1) \ldots p(x_N, y_N | x_{N-1}, y_{N-1}). \tag{4}$$

Transition probabilities can write in the following way

$$p(z_n | z_{n-1}) = p(x_n, y_n | x_{n-1}, y_{n-1}) = p(x_n | x_{n-1}, y_{n-1})\ p(y_n | x_{n-1}, y_{n-1}, x_n). \tag{5}$$

In PMC, the law $p(\boldsymbol{x}_{1:N} | \boldsymbol{y}_{1:N})$ is always Markovian, allowing the estimation of $\boldsymbol{X}_{1:N}$ from $\boldsymbol{Y}_{1:N} = \boldsymbol{y}_{1:N}$ (Ephraim and Merhav, 2002), while $\boldsymbol{X}_{1:N}$ being Markovian or not.

More precisely, according to laws $p(x_n | x_{n-1}, y_{n-1})$ and $p(y_n | x_{n-1}, y_{n-1}, x_n)$, we get four particular PMCs of interest:

(i) Process $\boldsymbol{Z}_{1:N}$ will be called "Hidden Markov Chain with Independent Noise" (HMC-IN) if $p(x_n | x_{n-1}, y_{n-1}) = p(x_n | x_{n-1})$ and $p(y_n | x_{n-1}, y_{n-1}, x_n) = p(y_n | x_n)$, so that transitions in Eq. (5) write

$$p(x_n, y_n | x_{n-1}, y_{n-1}) = p(x_n | x_{n-1})\ p(y_n | x_n). \tag{6}$$

Process $\boldsymbol{X}_{1:N}$ is then a Markov chain and random variables $Y_1, \ldots, Y_N$ are independent conditionally to $\boldsymbol{X}_{1:N}$. This classical model is traditionally called the hidden Markov chain.

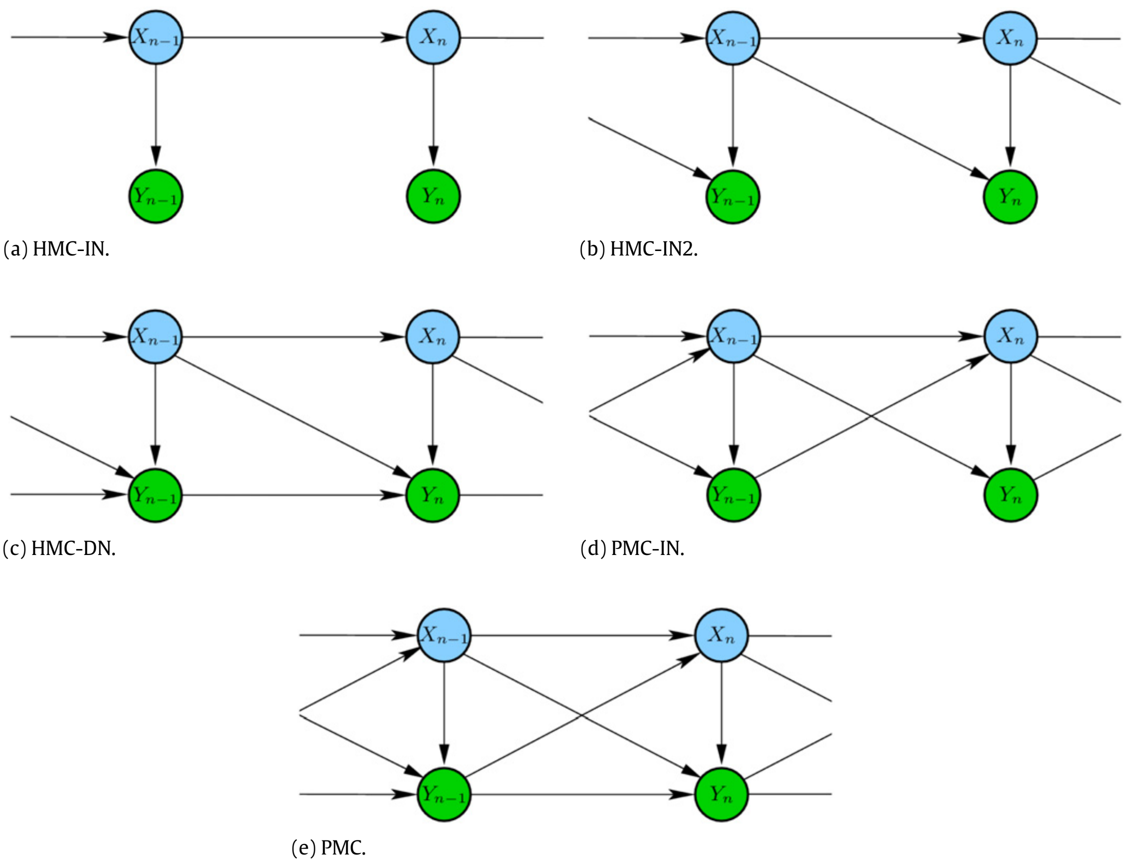


Fig. 1. Oriented dependence graphs for the HMC-IN, HMC-IN2, HMC-DN, PMC-IN and PMC models.

(ii) Process $\boldsymbol{Z}_{1:N}$ will be called "Hidden Markov Chain with Independent Noise of order 2" (HMC-IN2) if transitions in Eq. (5) write

$$p\left(x_n, y_n \mid x_{n-1}, y_{n-1}\right) = p\left(x_n \mid x_{n-1}\right)\; p\left(y_n \mid x_{n-1}, x_n\right). \tag{7}$$

Process $\boldsymbol{X}_{1:N}$ is then a Markov chain and random variables $Y_1, \ldots, Y_N$ are independent conditionally to $\boldsymbol{X}_{1:N}$. An HMC-IN is a particular case of HMC-IN2.

(iii) Process $\boldsymbol{Z}_{1:N}$ will be called "Hidden Markov Chain with Dependent Noise" (HMC-DN) if $p\left(x_n \mid x_{n-1}, y_{n-1}\right) = p\left(x_n \mid x_{n-1}\right)$, so that transitions in Eq. (5) write

$$p\left(x_n, y_n \mid x_{n-1}, y_{n-1}\right) = p\left(x_n \mid x_{n-1}\right)\; p\left(y_n \mid x_{n-1}, y_{n-1}, x_n\right). \tag{8}$$

Process $\boldsymbol{X}_{1:N}$ is again a Markov chain but $Y_1, \ldots, Y_N$ are (eventually) dependent conditionally to $\boldsymbol{X}_{1:N}$. An HMC-IN2 is a particular case of HMC-DN.

(iv) Process $\boldsymbol{Z}_{1:N}$ will be called "Pairwise Markov Chain with Independent Noise" (PMC-IN) if transitions in Eq. (5) write

$$p\left(x_n, y_n \mid x_{n-1}, y_{n-1}\right) = p\left(x_n \mid x_{n-1}, y_{n-1}\right)\; p\left(y_n \mid x_{n-1}, x_n\right). \tag{9}$$

Process $\boldsymbol{X}_{1:N}$ is no necessarily a Markov chain but $Y_1, \ldots, Y_N$ are independent conditionally to $\boldsymbol{X}_{1:N}$. Hence, a PMC-IN is a particular case of PMC-DN, the latter being abbreviated by PMC when no confusion is possible.

Oriented dependence graphs for the general PMC model and the four particular cases (i)–(iv) are reported in Fig. 1. Note also that models in Eqs. (4), (6) and (8) can be evaluated using on-line demonstrators at url www.fresnel.fr/perso/hmcext/index.php.

In this work, we will only consider the general PMC model given by Eq. (5) and classical HMC-IN model (i) given by Eq. (6). Furthermore, in the following, we consider Stationary and Reversible PMCs (SR-PMC). The first hypothesis means that $p\left(z_n, z_{n+1}\right)$ does not depend on $n = 1, \ldots, N-1$ and the second one means that the two families of conditional laws $p\left(z_{n+1} \mid z_n\right)$ and $p\left(z_n \mid z_{n+1}\right)$ are identical. Under the second hypothesis, models (7) and (8) are equivalent. We then get the following result, whose demonstration within a general framework can be consulted in Lanchantin et al. (2011):

Proposition. *Let $\boldsymbol{Z}_{1:N}$ be a SR-PMC, then the three following conditions are equivalent:*

(i) *$\boldsymbol{X}_{1:N}$ is a Markov chain;*
(ii) *$p\left(y_2 \mid x_1, x_2\right) = p\left(y_2 \mid x_2\right)$;*
(iii) *$p\left(y_n \mid x_{1:N}\right) = p\left(y_n \mid x_n\right)$, for all $n = 1, \ldots, N$.*

Given Eqs. (4)–(9) and previous proposition, we can assess how the PMC model generalizes the HMC-IN one satisfying Eq. (6). Note that this greater generality can result in greater efficiency in unsupervised image segmentation: as experimented in Derrode and Pieczynski (2004), the error rate can be halved.

The main purpose of this paper is to study various models for laws $p(y_1, y_2 | x_1, x_2)$ in Eq. (4) with copulas. This problem has been addressed in the case of HMC-DN, with transitions given by Eq. (8), in Brunel and Pieczynski (2005). Results obtained in unsupervised image segmentation are very encouraging. Here, we extend the work by studying the estimation of $p(z_1, z_2)$, including the automatic choice of copulas within a finite set of admissible copulas.

2.2. Copulas in PMC

From $p(z_2 | z_1) = p(z_1 | z_2)$, and so $p(x_1, x_2) = p(x_2, x_1)$ and $p(y_1 | x_1, x_2) = p(y_2 | x_2, x_1)$, a stationary and reversible PMC is characterized by

- $K(K+1)/2 - 1$ joint *a priori* probabilities $p(x_1, x_2)$;
- K^2 bi-dimensional densities $p(y_1, y_2 | x_1, x_2) = f_{x_1,x_2}(y_1, y_2)$, generally called "data-driven densities" since they model the variability of observations (e.g. noise in measurements), with only K^2 different margins.

Densities $p(y_1, y_2 | x_1, x_2)$ can be parameterized using copulas (Nelsen, 2005), introduced by Sklar (1959), which allow to define a 2D density f from its two marginal densities $f^{(1)}$ and $f^{(2)}$, and a dependence structure c, called "copula"

$$f(y_1, y_2; \boldsymbol{\theta}) = f^{(1)}(y_1; \boldsymbol{\theta}_1)\, f^{(2)}(y_2; \boldsymbol{\theta}_2)\, c\left(F^{(1)}(y_1; \boldsymbol{\theta}_1), F^{(2)}(y_2; \boldsymbol{\theta}_2); \boldsymbol{\theta}\right),$$

where $F^{(.)}$ denotes Cumulative Distribution Function (CDF) associated to $f^{(.)}$ and $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}\}$ denotes the set of parameters to characterize the parametric copula and the two margins. It is then possible to construct densities over $\mathbb{R}^2$ by considering various marginal distributions (Gaussian, gamma, beta of first and second kinds…) and various copulas (Gaussian, Student't, Clayton…) in an independent manner. A bi-dimensional Gaussian density is a particular case of Gaussian margins combined with a Gaussian copula.

For all experiments presented in this paper, we will consider the eight one-parameter copulas presented in Appendix A, and the zero-parameter product copula which gives the independence case (see below). These copulas can all be parameterized by Kendall's rank correlation (denoted by $\tau \in [-1, 1]$), allowing comparison between copula shapes with the same correlation. We will write either $c(.,.;\theta)$ or $c(.,.;\tau)$. Note that the range of possible value for τ is not the same for all copulas. Some of them do not allow $\tau = 0$, whereas some others do not allow $\tau < 0$. In the list considered here, Gaussian and Student't copulas are the only ones which cover the entire range of possible values for τ.

To simplify notations and when no confusion is possible, let for $x_1 = i$ and $x_2 = j$, $p(x_1, x_2) = p_{ij}$, $c_{x_1,x_2}(y_1, y_2; \tau_{x_1,x_2}) = c_{ij}(y_1, y_2; \tau_{ij})$, $f(y_1, y_2 | x_1, x_2) = f_{ij}(y_1, y_2)$, and so forth for marginal densities, e.g. $f(y_1 | x_1, x_2) = f_{ij}(y_1)$ and $f(y_1 | x_1) = f_i(y_1)$.

Following work in Brunel and Pieczynski (2005) on HMC-DN, copulas can be introduced in PMC and sub-models (i)–(iv) to parameterize the K^2 data-driven densities. The mixtures for the HMC-IN (6), HMC-DN (8) and PMC (5) models write

$$f_{\text{HMC-IN}}(y_1, y_2) = \sum_{i=1}^{K} \sum_{j=1}^{K} p_{ij}\, f_i(y_1) f_j(y_2), \tag{10}$$

$$f_{\text{HMC-DN}}(y_1, y_2) = \sum_{i=1}^{K} \sum_{j=1}^{K} p_{ij}\, f_i(y_1) f_j(y_2)\, c_{ij}\left(F_i(y_1), F_j(y_2)\right), \tag{11}$$

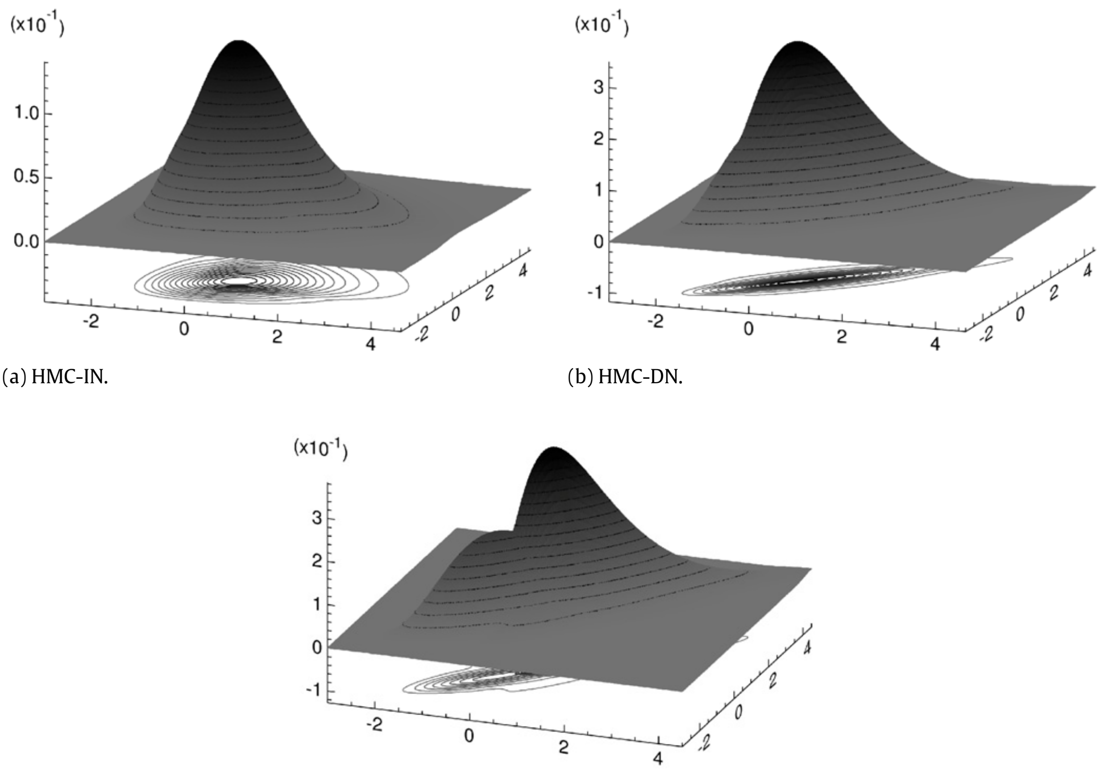
and

$$f_{\text{PMC}}(y_1, y_2) = \sum_{i=1}^{K} \sum_{j=1}^{K} p_{ij}\, f_{ij}(y_1) f_{ji}(y_2)\, c_{ij}\left(F_{ij}(y_1), F_{ji}(y_2)\right). \tag{12}$$

Indeed, one can note that:

- in Eq. (11), we write $f_j(y_2)$ instead of $f_{ij}(y_2)$ because, given x_2, y_2 is independent of x_1;
- in Eq. (12), inversion of indices in $f_{ji}(y_2) = f(y_2 | x_1 = j, x_2 = i)$ comes from the reversibility hypothesis of PMC models considered here.

An example of a $K = 2$ classes mixture is shown in Fig. 2 for the three models. We set $p(1,1) = 0.50$, $p(1,2) = p(2,1) = 0.05$ and $p(2,2) = 0.40$. For the PMC and HMC-DN models, we set the 4 copulas to be of Gumbel–Hougaard type (c^3, see Table A.11) with parameter $\theta = 3.33$ ($\tau = 0.7$). For the HMC-DN and HMC-IN models, we set the two margins to be Gamma distributions $\mathcal{G}(\lambda, \alpha, \theta)$: $f_1 \rightsquigarrow \mathcal{G}(-2.83, 8, 0.13)$ and $f_2 \rightsquigarrow \mathcal{G}(-1.63, 2.67, 0.38)$. The four margins for the PMC model are given by the Gamma distributions in Table 1.



(a) HMC-IN.

(b) HMC-DN.

(c) PMC.

Fig. 2. An example of mixture for the three models with $K = 2$. Parameter values are reported in the text.

Table 1
Margins parameters used for PMC simulation and restoration with two classes $\Omega = \{1, 2\}$. Gamma distributions have been chosen so that their standard deviations are identical to those of Gaussian distributions.

	Margin i, j
Gaussian margins $\mathcal{N}(\mu, \sigma)$ $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$f_{11} \rightsquigarrow \mathcal{N}(0.0, 1.00)$ $f_{12} \rightsquigarrow \mathcal{N}(0.3, 1.60)$ $f_{21} \rightsquigarrow \mathcal{N}(1.1, 1.40)$ $f_{22} \rightsquigarrow \mathcal{N}(1.5, 1.00)$
Gamma margins $\mathcal{G}(\lambda, \alpha, \theta)$ $p(x) = \frac{\theta^\alpha}{\Gamma(\alpha)} (x-\lambda)^{\alpha-1} e^{-\theta(x-\lambda)}$	$f_{11} \rightsquigarrow \mathcal{G}(-2.83, 8.00, 0.13), \mu = 0.0$ $f_{12} \rightsquigarrow \mathcal{G}(-3.65, 8.89, 0.18), \mu = 0.3$ $f_{21} \rightsquigarrow \mathcal{G}(-2.08, 3.08, 0.46), \mu = 1.1$ $f_{22} \rightsquigarrow \mathcal{G}(-1.63, 2.67, 0.38), \mu = 1.5$

3. PMC supervised data restoration

The aim of this section is to evaluate the influence of copula shapes in the supervised restoration (i.e. based on true parameters) of PMC data. We start by providing a method to simulate PMC data whatever the copula shapes involved. Then systematic results of data restoration are presented, with varying Kendall's rank correlation and margin shapes. The last subsection is intended to measure the influence of Markovianity on results, by mean of experiments regarding an original model called "Pairwise Mixture Model" (PMM) with copulas.

3.1. Simulation and supervised restoration of PMC data

According to Derrode and Pieczynski (2004), PMC data can be generated using the following procedure. Starting data ($n = 1$) are simulated according to

$$p(x_1) = \sum_{x_2=1}^{K} p(x_1, x_2), \qquad p(y_1 | x_1) = \sum_{x_2=1}^{K} p(x_1 | x_2)\, f_{x_1, x_2}(y_1).$$

Simulation of y_1 requires a sampling from a finite mixture of, possibly non-Gaussian, 1D densities.

Then, next data ($n > 1$) are generated by alternating the simulation of x_{n+1} (conditionally to $\boldsymbol{z}_n = (x_n, y_n)$) and the simulation of y_{n+1} (conditionally to $\boldsymbol{z}_n$ and x_{n+1}) according to

$$p(x_{n+1} | x_n = i, y_n) \propto p(i, x_{n+1}) \; f_{i,x_{n+1}}(y_n), \tag{13}$$

$$\begin{aligned} p(y_{n+1} | x_{n+1} = j, x_n = i, y_n) &= \frac{f_{ij}(y_n, y_{n+1})}{f_{ij}(y_n)} \\ &= f_{ji}(y_{n+1}) \; c_{ij}\left(F_{ij}(y_n), F_{ji}(y_{n+1}); \tau_{ij}\right). \end{aligned} \tag{14}$$

Note again that inversion of indices in $f_{ji}(y_{n+1})$ and $F_{ji}(y_{n+1})$ in Eq. (14) comes from the reversibility hypothesis of PMC models.

The simulation of y_{n+1} can be performed using the rejection principle, as presented in Appendix B. Although very general, the method can be computer demanding since it involves an acceptance criterion which can result in many rejected draws for every accepted one (the rejection rate depends on the copula as detailed in Appendix A). When dealing with a specific copula, it can be of interest to replace the rejection principle by methods suited for the copula, e.g. the conditional distribution method to simulate Franck's copula (Genest and Mackay, 1986) or Gumbel's copula using a numerical root finding algorithm, or the specific algorithms designed for Clayton's copula (Devroye, 1986), Archimedean copulas (Genest and Mackay, 1986) or t-copula (Demarta and McNeil, 2005). Some of algorithms are reported in Nelsen (2005).

The Bayesian restoration of $\boldsymbol{X}$ according to the MPM (Maximization of Posterior Marginals) criterion writes

$$\forall n \in [1, N], \quad \hat{x}_n = \arg\max_{x_n \in \Omega} p(x_n | y_{1:N}), \tag{15}$$

with $p(x_n | y_{1:N}) \propto \alpha_n(x_n) \; \beta_n(x_n)$ the marginal *a posteriori* distributions computed from the forward-like $\alpha_n(x_n) = p(x_n, y_{1:n})$ and the backward-like $\beta_n(x_n) = p(y_{n+1:N} | x_n, y_n)$ probabilities suited to the PMC model. These probabilities can be computed recursively

$$\begin{aligned} &\alpha_1(x_1) = p(x_1)\, p(y_1 | x_1), \\ &\alpha_{n+1}(x_{n+1}) = \sum_{x_n \in \Omega} \alpha_n(x_n) p(z_{n+1} | z_n), \quad \text{for } 1 \le n < N, \end{aligned} \tag{16}$$

and

$$\begin{aligned} &\beta_1(x_N) = 1, \\ &\beta_n(x_n) = \sum_{x_{n+1} \in \Omega} \beta_{n+1}(x_{n+1}) p(z_{n+1} | z_n), \quad \text{for } 1 \le n < N, \end{aligned} \tag{17}$$

see Derrode and Pieczynski (2004) for details. For latter use, see Eq. (23), let us consider joint *a posteriori* probabilities $p(x_n, x_{n+1} | y_{1:N})$ which write

$$p(x_n, x_{n+1} | y_{1:N}) = \frac{\alpha_n(x_n) \; p(z_{n+1} | z_n) \; \beta_{n+1}(x_{n+1})}{\sum_{a \in \Omega} \sum_{b \in \Omega} \alpha_n(b) \; p(b, y_{n+1} | a, y_n) \; \beta_{n+1}(b)}. \tag{18}$$

Bayesian restoration according to the Maximum A Posteriori (MAP) criterion is also available for the PMC model (Derrode and Pieczynski, 2004). But according to our experience in signal and image processing, these two criteria show similar behavior. This was also confirmed for experiments conducted below, so that we decided not to report MAP results for sake of clarity.

3.2. Impact of copula shapes on supervised PMC data restoration

In order to account for the numerical influence of the copula shapes only, we conducted the following experiment

1. We simulated N PMC data with $K = 2$ classes ($\Omega = \{1, 2\}$), according to some *a priori* probabilities p_{ij}, to some given copulas c_{ij} with parameter τ_{ij}, and to some margins f_{ij}.
2. Then we restored simulated observations according to MPM, using all simulation parameters except the copula shape which is either the true one or one in the list in Appendix A.

We set $K = 2$, $N = 2000$ and the same copula shape for the four copulas c_{ij} involved. Joint *a priori* probabilities were set to $p(1, 1) = 0.5$, $p(1, 2) = p(2, 1) = 0.05$ and $p(1, 2) = 0.4$. We conducted systematic experiments for all available copulas according to Gaussian and to Gamma margin (see Table 1 for parameter values) and for two Kendall's rank correlation values:

- $\tau_{\#1} = 0.16$. The set of eligible copulas is noted $\Pi_{\#1} = \left\{c^0, c^1, c^2, c^3, c^4, c^5, c^6\right\}$.
- $\tau_{\#2} = 0.70$. The set of eligible copulas is noted $\Pi_{\#2} = \left\{c^0, c^1, c^2, c^3, c^6, c^7, c^8\right\}$.

Table 2
PMC model. Mean classification error rates using $\tau_{\#1} = 0.16$ and Gaussian (a) or Gamma (b) margins. Example of reading: for table (a), when the true copula is c^3 and the tested copula is c^5, we get an error of 14.11%. The minimal error is underlined in bold for each column.

Copula	c^0	c^1	c^2	c^3	c^4	c^5	c^6
(a) Gaussian margins							
c^0	**11.09** (**0.9**)	14.30 (1.2)	13.72 (1.3)	13.90 (1.2)	14.64 (1.2)	14.73 (1.2)	13.88 (1.2)
c^1	11.28 (0.9)	**14.09** (**1.2**)	13.81 (1.4)	13.84 (1.2)	14.39 (1.3)	14.43 (1.3)	13.78 (1.3)
c^2	11.89 (0.9)	14.65 (1.2)	**13.13** (**1.3**)	14.03 (1.2)	15.31 (1.3)	15.60 (1.2)	14.23 (1.3)
c^3	11.71 (0.9)	14.36 (1.2)	13.59 (1.3)	**13.67** (**1.2**)	14.93 (1.3)	15.05 (1.2)	14.38 (1.3)
c^4	11.40 (1.0)	14.13 (1.2)	14.06 (1.4)	13.95 (1.2)	**14.26** (**1.3**)	14.28 (1.3)	13.82 (1.3)
c^5	11.69 (1.0)	14.22 (1.2)	14.51 (1.4)	14.11 (1.2)	14.34 (1.3)	**14.27** (**1.2**)	13.94 (1.3)
c^6	11.81 (1.0)	14.39 (1.2)	13.90 (1.4)	14.46 (1.2)	14.67 (1.3)	14.83 (1.3)	**13.50** (**1.3**)
(b) Gamma margins							
c^0	**8.54** (**0.7**)	10.58 (1.0)	10.89 (1.0)	10.95 (1.0)	10.87 (0.9)	10.82 (0.9)	10.46 (1.0)
c^1	8.75 (0.8)	**10.26** (**1.0**)	10.78 (1.0)	10.75 (1.0)	10.41 (1.0)	10.33 (0.9)	9.93 (1.0)
c^2	9.12 (0.8)	10.65 (1.0)	**10.18** (**0.9**)	10.92 (1.0)	11.04 (1.0)	11.07 (0.9)	10.12 (1.0)
c^3	8.78 (0.8)	10.36 (1.0)	10.58 (1.0)	**10.57** (**0.9**)	10.65 (1.0)	10.63 (0.9)	10.24 (1.0)
c^4	8.81 (0.8)	10.29 (1.0)	10.86 (1.0)	10.81 (1.0)	**10.33** (**1.0**)	10.24 (0.9)	9.93 (0.9)
c^5	8.89 (0.8)	10.33 (1.0)	11.09 (1.0)	10.88 (1.0)	10.38 (1.0)	**10.22** (**0.9**)	9.98 (1.0)
c^6	9.17 (0.8)	10.48 (1.0)	10.73 (1.0)	11.23 (1.0)	10.69 (0.9)	10.65 (0.9)	**9.73** (**0.9**)

Table 3
PMC model. Mean classification error rates using $\tau_{\#2} = 0.70$ and Gaussian (a) or Gamma (b) margins.

Copula	c^0	c^1	c^2	c^3	c^6	c^7	c^8
(a) Gaussian margins							
c^0	**11.06** (**1.0**)	26.27 (2.2)	25.77 (2.3)	26.27 (2.5)	27.50 (2.3)	25.63 (2.2)	25.71 (2.1)
c^1	41.55 (1.1)	**14.95** (**2.1**)	17.84 (2.2)	17.13 (2.1)	19.43 (2.5)	16.90 (2.2)	16.70 (2.2)
c^2	36.16 (1.0)	16.01 (2.1)	**16.86** (**2.2**)	15.69 (2.0)	17.07 (2.4)	16.27 (2.1)	15.97 (2.1)
c^3	40.50 (1.0)	18.14 (2.2)	18.98 (2.4)	**12.43** (**1.9**)	29.37 (2.7)	20.48 (2.4)	15.78 (2.2)
c^6	44.41 (1.9)	21.09 (2.3)	21.84 (2.4)	27.97 (2.4)	**6.31** (**1.1**)	18.68 (2.2)	24.21 (2.5)
c^7	37.97 (1.0)	16.56 (2.2)	17.64 (2.2)	19.12 (2.2)	13.62 (2.1)	**15.25** (**2.0**)	17.55 (2.3)
c^8	39.25 (1.0)	16.59 (2.1)	17.80 (2.5)	13.71 (1.9)	24.98 (2.8)	17.94 (2.3)	**14.99** (**2.0**)
(b) Gamma margins							
c^0	**8.69** (**0.7**)	22.48 (2.0)	21.94 (2.2)	22.18 (2.1)	23.45 (2.3)	21.97 (2.2)	21.95 (2.2)
c^1	44.89 (2.0)	**13.37** (**1.8**)	16.09 (2.0)	19.39 (2.2)	12.12 (1.7)	14.05 (1.9)	16.72 (1.9)
c^2	31.04 (2.0)	14.60 (1.8)	**14.91** (**2.0**)	20.14 (2.1)	9.23 (1.3)	13.36 (1.7)	16.72 (1.9)
c^3	35.84 (1.1)	15.71 (1.9)	16.98 (2.1)	**18.03** (**2.1**)	17.69 (2.2)	15.87 (2.0)	16.68 (1.9)
c^6	45.64 (2.9)	17.66 (2.0)	17.32 (1.9)	23.23 (2.1)	**6.71** (**1.1**)	14.85 (1.8)	20.08 (2.1)
c^7	37.56 (2.0)	14.59 (1.8)	15.24 (1.9)	20.56 (2.1)	8.60 (1.2)	**13.09** (**1.7**)	16.98 (1.9)
c^8	35.46 (1.4)	14.43 (1.8)	15.62 (2.0)	18.75 (2.2)	12.97 (1.7)	13.87 (1.8)	**16.05** (**1.9**)

A copula is said to be eligible if the value for Kendall's tau belongs to the admissible range for that copula, see last column of Table A.11.

For all shapes in $\Pi_{\#i}$, we simulated noisy data according to the corresponding PMC model. Then we restored data, providing all parameters used at simulation time and trying each of the copulas in $\Pi_{\#i}$. The error rates reported below are means of 300 independent experiments (standard deviation are reported between parenthesis). Experimental results are reported in Table 2 for $\tau_{\#1}$ and in Table 3 for $\tau_{\#2}$, using Gaussian and Gamma margins. Bold results underline the minimum value in each column.

Comments on results can be summarized as follows:

- Whatever Kendall's correlation value and margins shapes, the restoration with the right copula always gives the lowest mean error rate;
- When τ is low, the mean error rates can be very close (e.g. Table 2(a) column c^1 and Table 2(b) column c^5), showing that confusion can appear in experiments if correlation is low;
- When τ is large, the mean error rates are very different. This is especially true for Clayton's copula c^6, where the rate is divided up to 5 when compared to Gumbel copula c^3 in Table 3(a);
- Copulas with nearly the same density shapes show the same behavior and very small error rate differences (e.g. copulas c^7 and c^8).

The main conclusion is that using a wrong copula can result in poor results when correlation is high.

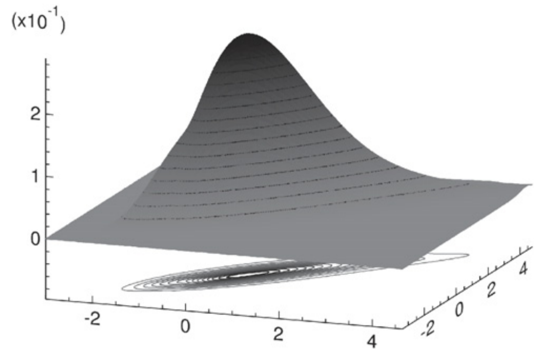


Fig. 3. Example of mixture for the PMM with $K = 2$ classes and the same parameters used to draw the mixture in Fig. 2(c).

3.3. Impact of Markovianity on results

To evaluate the influence of Markovianity on classification results, we conducted the same kind of experiments as before, but considering an i.i.d. Pairwise Mixture Model (PMM), in which (y_1, y_2) are the pairwise observations and (x_1, x_2) the pairwise hidden states. In this model, pairs $(y_1, y_2), (y_3, y_4), \ldots$ are independent but each of them have the same distribution than in the PMC. The PMM is presented in details in Derrode and Pieczynski (2011) for Gaussian densities. Let us extend it to the copula context considered in this paper.

In the PMM, hidden states are simulated using drawings from $p(x_1)$ and $p(x_2 | x_1)$ respectively, whereas observations are simulated using drawings from the densities $p(y_1 | x_1, x_2)$ and $p(y_2 | x_1, x_2, y_1)$ in Eq. (14) (replacing n by 1), using the same rejection principle used to simulate PMC data. The mixture model writes the same than the PMC model in Eq. (12) and the mixture corresponding to Fig. 2(c) is shown in Fig. 3 (we kept exactly the same parameter values).

The Bayesian MPM restoration of x_1 from observations (y_1, y_2) and model parameters is done according to

$$\hat{x}_1 = \arg \max_{i \in [1,\, K]} p(x_1 = i | y_1, y_2), \tag{19}$$

with

$$p(x_1 = i | y_1, y_2) = \sum_{j=1}^{K} p(i, x_2 = j | y_1, y_2) \propto \sum_{j=1}^{K} p_{ij} f_{ij}(y_1, y_2).$$

The same kind of equations applies to restore x_2 (summing over X_1).

For experiments, we computed the mean MPM error rate of 300 independent simulations and restorations for $N = 1000$ couple of data (y_1, y_2) (which corresponds to $N = 2000$ in PMC experiments), keeping the same parameter values than the ones used in PMC experiments. Results for Gaussian and Gamma margins are reported in Tables 4 and 5 for $\tau_{\#1}$ and $\tau_{\#2}$ respectively. These tables can directly be compared to Tables 2 and 3 respectively.

One can first notice that several copulas are confused when τ is low (c^3 with c^0 and c^5 with c^1 in Table 4(a)). Such confusions no more appear for large τ. Also, as for the PMC model, mean error rates are more distinguishable when Kendall's tau is high. Comparing results from PMC and PMM models, it is noticeable that mean error rates for PMM are reduced from 30% to 40% when compared to the PMC model: Markovianity has a strong influence on data restoration.

4. Unsupervised PMC data restoration with copula selection

One of the very interesting properties of the PMC model is the ability to estimate parameters from observations only. Automatic parameters estimation has already been experimented in the "full-Gaussian" case in Derrode and Pieczynski (2004) (using ICE), and in the HMC-DN model with Gaussian copulas and non-Gaussian margins (using the Stochastic EM Celeux and Diebolt, 1985) in Lanchantin et al. (2011), with application in image and signal processing. Given the full PMC model, one would like to know if it is possible to automatically recover the proper shape of the copulas involved in simulated data. To that goal, we incorporated the Bayesian copula selection method introduced by Huard et al. (2006) in an ICE-based parameter estimation scheme, allowing to select the "best shape" for each of the K^2 copulas at each ICE iteration from observations only. Several experiments finally illustrate the nice behavior of the entire algorithm.

4.1. Bayesian copula selection

Bayesian identification of marginal and joint CDFs, and copula selection are the subjects of numerous recent papers, among them (Nikoloulopoulos and Karlis, 2008; Genest, Rémillard et al., 2009; Genest, Masiello et al., 2009; Xiaomei et al., 2010; Noh et al., 2010). In this work, we used the Bayesian copula selection method (Huard et al., 2006) (i) for its simplicity and low computational burden, and (ii) since all copulas considered in this paper can be parameterized by Kendall's tau.

Table 4
PMM model. Mean classification error rates using $\tau_{\#1} = 0.16$ and Gaussian (a) or Gamma (b) margins.

Copula	c^0	c^1	c^2	c^3	c^4	c^5	c^6
(a) Gaussian margins							
c^0	**17.11** (**0.8**)	19.79 (0.8)	18.48 (0.8)	19.30 (0.7)	20.42 (0.8)	20.60 (0.8)	19.20 (0.8)
c^1	17.18 (0.7)	**19.74** (**0.8**)	18.56 (0.8)	19.34 (0.7)	20.32 (0.8)	20.49 (0.8)	19.15 (0.8)
c^2	17.12 (0.8)	19.78 (0.8)	**18.45** (**0.7**)	**19.27** (**0.7**)	20.37 (0.8)	20.60 (0.8)	19.20 (0.8)
c^3	17.26 (0.7)	19.86 (0.8)	18.57 (0.7)	19.29 (0.7)	20.47 (0.8)	20.65 (0.8)	19.48 (0.8)
c^4	17.30 (0.7)	19.75 (0.8)	18.63 (0.8)	19.41 (0.7)	**20.31** (0.8)	20.44 (0.8)	19.17 (0.8)
c^5	17.56 (0.7)	19.83 (0.8)	18.94 (0.8)	19.55 (0.7)	20.35 (0.8)	**20.43** (**0.8**)	19.26 (0.8)
c^6	17.35 (0.7)	19.78 (0.8)	18.70 (0.8)	19.62 (0.8)	20.31 (0.8)	20.46 (0.8)	**18.89** (**0.8**)
(b) Gamma margins							
c^0	**16.99** (**0.7**)	19.22 (0.7)	18.54 (0.7)	18.78 (0.6)	19.51 (0.6)	19.63 (0.7)	19.23 (0.6)
c^1	17.43 (0.6)	**18.74** (**0.6**)	18.45 (0.6)	18.65 (0.6)	18.84 (0.6)	18.91 (0.6)	18.67 (0.6)
c^2	17.40 (0.7)	19.06 (0.6)	**18.08** (**0.7**)	18.82 (0.6)	19.29 (0.6)	19.48 (0.6)	18.78 (0.7)
c^3	17.18 (0.6)	18.82 (0.6)	18.33 (0.6)	**18.58** (**0.6**)	18.94 (0.6)	19.06 (0.6)	18.78 (0.6)
c^4	17.65 (0.7)	18.74 (0.6)	18.56 (0.6)	18.73 (0.6)	**18.78** (**0.6**)	**18.83** (**0.6**)	18.73 (0.6)
c^5	17.74 (0.7)	18.80 (0.7)	18.74 (0.6)	18.77 (0.6)	18.82 (0.6)	18.84 (0.6)	18.81 (0.6)
c^6	17.75 (0.7)	18.90 (0.6)	18.36 (0.6)	18.96 (0.6)	19.06 (0.6)	19.20 (0.6)	**18.49** (**0.6**)

Table 5
PMM model. Mean classification error rates using $\tau_{\#2} = 0.70$ and Gaussian (a) or Gamma (b) margins.

Copula	c^0	c^1	c^2	c^3	c^6	c^7	c^8
(a) Gaussian margins							
c^0	**17.06** (**0.8**)	24.33 (0.7)	23.99 (0.7)	24.30 (0.7)	24.95 (0.7)	23.97 (0.8)	24.12 (0.7)
c^1	29.90 (0.8)	**22.47** (**0.7**)	22.59 (0.7)	22.41 (0.6)	22.52 (0.7)	22.93 (0.7)	22.67 (0.7)
c^2	20.25 (0.8)	22.66 (0.7)	**22.09** (**0.7**)	22.25 (0.7)	22.99 (0.7)	22.49 (0.7)	22.29 (0.7)
c^3	25.60 (0.8)	23.30 (0.7)	22.76 (0.8)	**21.80** (**0.8**)	25.72 (0.8)	23.47 (0.8)	22.43 (0.7)
c^6	29.13 (1.0)	25.73 (0.8)	24.93 (0.8)	27.38 (0.8)	**19.50** (**0.8**)	24.49 (0.8)	26.19 (0.8)
c^7	20.11 (0.8)	22.75 (0.7)	22.21 (0.7)	22.76 (0.7)	22.08 (0.7)	**22.38** (**0.7**)	22.61 (0.7)
c^8	21.52 (0.7)	22.81 (0.7)	22.26 (0.7)	21.93 (0.7)	24.28 (0.7)	22.78 (0.7)	**22.19** (**0.7**)
(b) Gamma margins							
c^0	**17.04** (**0.6**)	23.75 (0.6)	23.54 (0.6)	23.70 (0.6)	24.06 (0.6)	23.58 (0.6)	23.57 (0.6)
c^1	26.64 (0.9)	**12.02** (**0.6**)	11.14 (0.6)	11.24 (0.7)	12.89 (0.6)	11.69 (0.6)	11.35 (0.7)
c^2	26.26 (0.9)	12.25 (0.7)	**10.73** (**0.6**)	10.36 (0.7)	13.37 (0.6)	11.44 (0.6)	10.72 (0.7)
c^3	27.36 (0.9)	12.52 (0.7)	10.93 (0.6)	**10.07** (**0.7**)	14.05 (0.6)	11.78 (0.6)	10.74 (0.7)
c^6	28.97 (1.0)	14.36 (0.6)	13.43 (0.6)	15.15 (0.6)	**10.94** (**0.5**)	13.25 (0.6)	14.31 (0.6)
c^7	26.51 (0.9)	12.29 (0.7)	10.81 (0.6)	10.83 (0.7)	12.98 (0.6)	**11.34** (**0.6**)	10.94 (0.7)
c^8	26.78 (0.9)	12.30 (0.7)	10.78 (0.6)	10.17 (0.7)	13.72 (0.6)	11.54 (0.6)	**10.67** (**0.7**)

For short, given a set of 2D observations $\boldsymbol{y} = \{\boldsymbol{y}^1, \boldsymbol{y}^2\}$ with $\boldsymbol{y}^1 = \{y_1^1, \ldots, y_N^1\}$ and $\boldsymbol{y}^2 = \{y_1^2, \ldots, y_N^2\}$, the "best copula" c^s within a finite set of copula shapes $\Pi = \{c^1, \ldots, c^R\}$ is selected according to

$$s = \arg\max_{r\in[1,R]} \frac{1}{\tau_M^r - \tau_m^r} \int_{\tau_m^r}^{\tau_M^r} \prod_{n=1}^{N} c^r\left(F^1(y_n^1), F^2(y_n^2); \tau\right)\, d\tau, \tag{20}$$

where F^1 and F^2 are the CDF of marginal data series $\boldsymbol{y}^1$ and $\boldsymbol{y}^2$ (their shapes are supposed known). Coefficients τ_m^r and τ_M^r for copula c^r represent the minimal and maximal admissible values for Kendall's tau (see Table A.11 for copulas considered in this work). One interesting specificity of the method is that it does not rely on the estimation of Kendall's tau. This "Bayesian copula selection" method has been applied with success in Sakji-Nsibi and Benazza-Benyahia (2009) for wavelet-based multicomponent image retrieval. Let us now discuss how to integrate this method into ICE parameters estimation.

4.2. Automatic copula selection in ICE-based parameters estimation

Consider a stationary and reversible PMC whose law is given by $p_{\boldsymbol{\theta}}(z_1, z_2)$, with $\boldsymbol{\theta}$ a set of real parameters. When one wishes to estimate $\boldsymbol{\theta}$ from $\boldsymbol{y}_{1:N}$, we can consider at least two general methods that produce series of estimates $\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^q, \ldots$:

(i) "Expectation–Maximization" (EM) method: from $\boldsymbol{\theta}^0$, $\boldsymbol{\theta}^{q+1}$ is defined from $\boldsymbol{\theta}^q$ using

$$\boldsymbol{\theta}^{q+1}(\boldsymbol{y}_{1:N}) = \arg\max_{\boldsymbol{\theta}} E\Big[p_{\boldsymbol{\theta}}(\boldsymbol{X}_{1:N}, \boldsymbol{Y}_{1:N})\,\Big|\boldsymbol{Y}_{1:N} = \boldsymbol{y}_{1:N}, \boldsymbol{\theta}^q(\boldsymbol{y}_{1:N})\Big]. \tag{21}$$

(ii) "Iterative Conditional Estimation" (ICE) method (Pieczynski, 1992): from $\boldsymbol{\theta}^0$ and an estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N})$ of $\boldsymbol{\theta}$ from complete data $(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N})$, $\boldsymbol{\theta}^{q+1}$ is defined from $\boldsymbol{\theta}^q$ using

$$\boldsymbol{\theta}^{q+1}(\boldsymbol{y}_{1:N}) = E\left[\hat{\boldsymbol{\theta}}(\boldsymbol{X}_{1:N}, \boldsymbol{Y}_{1:N}) \,\middle|\boldsymbol{Y}_{1:N} = \boldsymbol{y}_{1:N}, \boldsymbol{\theta}^q(\boldsymbol{y}_{1:N})\right]. \tag{22}$$

ICE, which is more recent and less popular than algorithms from the EM family, has been used successfully in different contexts (Destrempes and Mignotte, 2004; Salzenstein et al., 2007; Wua et al., 2011). Possible equivalence of sequences of estimated parameters between EM and ICE has been stated in the case of an exponential family of distributions (Delmas, 1997). In the case we are interested in this paper, likelihood is difficult to handle and thus we choose to work with ICE.

When conditional expectation in Eq. (22) is not computable for some components θ_m in $\boldsymbol{\theta}$, we estimate them by simulating L realizations $\boldsymbol{x}^1_{1:N}, \ldots, \boldsymbol{x}^L_{1:N}$ of $\boldsymbol{x}_{1:N}$ according to $\boldsymbol{x}_{1:N} \,|\boldsymbol{y}_{1:N}$. To do this, we use the following transition

$$p_{\boldsymbol{\theta}^q}(x_{n+1} \,|x_n, y_{1:N}) = \frac{p_{\boldsymbol{\theta}^q}(x_n, x_{n+1} \,|y_{1:N})}{p_{\boldsymbol{\theta}^q}(x_n \,|y_{1:N})}, \tag{23}$$

see Eq. (18), and

$$\theta_m^{q+1}(\boldsymbol{y}_{1:N}) = \frac{\hat{\theta}_m(\boldsymbol{x}^{1,q}_{1:N}, \boldsymbol{y}_{1:N}) + \cdots + \hat{\theta}_m(\boldsymbol{x}^{L,q}_{1:N}, \boldsymbol{y}_{1:N})}{L}. \tag{24}$$

In the case considered in this paper, let

$$p_{\boldsymbol{\theta}}(z_1, z_2) = p_{\boldsymbol{\theta}}(x_1, x_2)\, p_{\boldsymbol{\theta}}(y_1, y_2 \,|x_1, x_2). \tag{25}$$

As our main objective being to study the importance of copulas in the estimation of $\boldsymbol{x}_{1:N}$ from $\boldsymbol{y}_{1:N}$, we assume that the marginal distributions $p_{\boldsymbol{\theta}}(y_1 \,|x_1, x_2)$ and $p_{\boldsymbol{\theta}}(y_2 \,|x_1, x_2)$ are entirely known, i.e. both the laws family and their shape parameters are known. Laws $p_{\boldsymbol{\theta}}(y_1, y_2 \,|x_1, x_2)$ are then determined by their copula. In order to simplify notations, let $p_{\boldsymbol{\theta}}(x_1 = i, x_2 = j) = p_{ij}$ and τ_{ij} denotes the unique parameter for copula c_{ij}.

We solve the estimation problem using complete data $(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N})$ in the following way:

- Parameters p_{ij} can be estimated by the following empirical estimate:

$$\hat{p}_{ij} = \frac{1}{N-1} \sum_{n=1}^{N-1} \boldsymbol{1}_{x_n=i, x_{n+1}=j}. \tag{26}$$

- We divide the sample $\boldsymbol{y}_{1:N}$ in K^2 sub-samples $(\boldsymbol{y}^{ij}_{1:N})$, $i, j \in \Omega$ such that for $n = 1, \ldots, N-1$, $y_n \in \boldsymbol{y}^{ij}_{1:N}$ if $(x_n, x_{n+1}) = (i, j)$. For all $(i, j) \in \Omega^2$, we select the "best copula" c^s_{ij} corresponding to $p_{\boldsymbol{\theta}}(y_1, y_2 \,|x_1 = i, x_2 = j)$ among $\Pi_{ij} = \left\{c^1_{ij}, \ldots, c^P_{ij}\right\}$, according to criterion in Eq. (20), and then estimate its parameter τ^s_{ij} from $\boldsymbol{y}^{ij}_{1:N}$.

  The principle of ICE is then applied according to:

- At first iteration ($q = 0$), we set initial values p^0_{ij} for p_{ij} and initial copulas $c^{s,0}_{ij}$, based on a kmeans classification.
- For next iterations, p^{q+1}_{ij} are estimated from p^q_{ij} and $c^{s,q}_{ij}$ by taking the conditional expectation of Eq. (26) (see Eqs. (21) and (22) in Derrode and Pieczynski (2004)), whereas $c^{ij,q+1}_s$ are estimated using the complete data procedure described above, replacing $\boldsymbol{x}_{1:N}$ by $\boldsymbol{x}^q_{1:N}$ ($L = 1$).

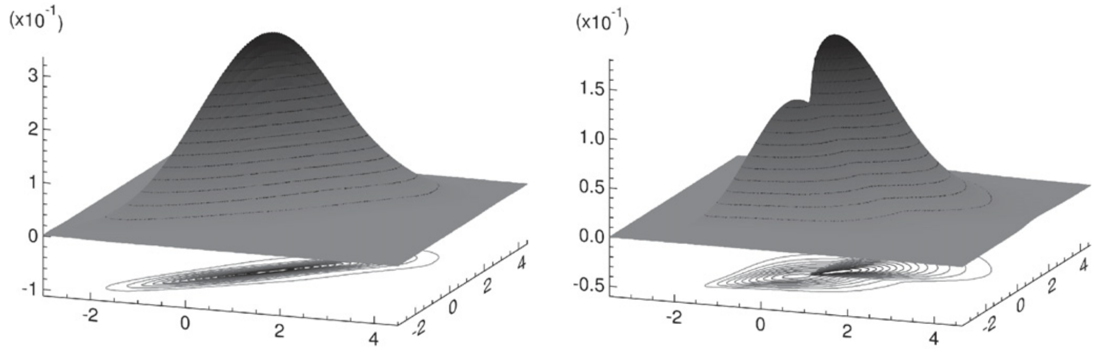
Finally, the K^2 best copulas $c^{s,Q}_{ij}$ involved in an estimated PMC model are the copulas selected when ICE has converged ($q = Q$).

4.3. Experiments on copula selection in the PMC model

This section intends to evaluate the combination of ICE and Bayesian copula selection method for the reliable identification of copulas used in PMC.

For data simulation, common parameters of all experiments presented below are: $K = 2$ ($\Omega = \{1, 2\}$), $N = 2500$, $Q = 30$ (number of ICE iterations), and $p(1, 1) = 0.50$, $p(1, 2) = p(2, 1) = 0.05$, $p(2, 2) = 0.40$. Specific parameters used for the two experiments are:

1. Experiment #1
   - Set of copulas: $\forall(i, j) \in \Omega^2$, $\Pi_{ij} = \Pi_{\#1} = \left\{c^1, c^3, c^6\right\}$ (resp. Gaussian, Gumbel, Clayton).
   - $c_{11} = c^1$ with $\tau_{11} = 0.7$, $c_{12} = c^3$ with $\tau_{12} = 0.4$, $c_{21} = c^3$ with $\tau_{21} = 0.4$ and $c_{22} = c^6$ with $\tau_{22} = 0.7$.
   - Gaussian margins from Table 1.
2. Experiment #2
   - Set of copulas: $\forall(i, j) \in \Omega^2$, $\Pi_{ij} = \Pi_{\#2} = \left\{c^2, c^3, c^5, c^6\right\}$ (resp. Student't, Gumbel, Cubic section, Clayton).
   - $c_{11} = c^2$ with $\tau_{11} = 0.25$, $c_{12} = c^5$ with $\tau_{12} = 0.10$, $c_{21} = c^5$ with $\tau_{21} = 0.10$ and $c_{22} = c^6$ with $\tau_{22} = 0.20$.
   - Gamma margins from Table 1.



(a) Experiment #1. (b) Experiment #2.

Fig. 4. Examples of simulated mixtures for experiments conducted in Section 4.3, when $p(1, 1) = 0.50$, $p(1, 2) = p(2, 1) = 0.05$, $p(2, 2) = 0.40$.

Table 6
Results of automatic copula selection for experiment in Section 4.3. The integer value gives the number of times the right copula has been chosen for the 10 experiments. The mean Kendall's tau is given between parenthesis (true values are recalled in bold for ease of comparison).

Experiment	c_{11}	c_{12}	c_{21}	c_{22}
#1	10 (0.69–**0.70**)	9 (0.29–**0.40**)	7 (0.32–**0.40**)	10 (0.71–**0.70**)
#2	10 (0.23–**0.25**)	9 (0.10–**0.10**)	9 (0.11–**0.10**)	10 (0.22–**0.20**)

Table 7
Idem as Table 6, but with joint *a priori* probabilities set to $p(1, 1) = p(2, 2) = 0.35$ and $p(1, 2) = p(2, 1) = 0.15$ for data simulation.

Experiment	c_{11}	c_{12}	c_{21}	c_{22}
#1	10 (0.69–**0.70**)	10 (0.40–**0.40**)	10 (0.40–**0.40**)	10 (0.69–**0.70**)
#2	10 (0.29–**0.25**)	10 (0.11–**0.10**)	10 (0.11–**0.10**)	8 (0.17–**0.20**)

Examples of mixture are shown in Fig. 4. The sets of candidates were build on the basis of their shape diversity.

Regarding unsupervised restoration, all parameters involved in the PMC model are estimated except the marginal density families, e.g. Gaussian, Gamma, which are supposed known (but parameters of margins are also estimated). The automatic selection of margin laws family has been studied in Giordana and Pieczynski (1997) in a vectorial HMC, and could have been adapted here.

Table 6 gives the number of times the right copula were selected for the $K^2 = 4$ copulas on 10 independent simulations and unsupervised restorations for the two experiments. Note that parameters estimated for margin shapes are not reported.

Whatever the margin and Kendall's tau involved, the right copulas are always selected for copulas c_{11} and c_{22}. One can note a few confusions for copulas c_{12} and c_{21}, which can be explained by the low number of samples available for copulas estimation (about $p\,(1, 2)\,N = p\,(2, 1)\,N = 0.05 * 2500 = 125$). This remark is confirmed by results reported in Table 7, where the only difference with previous experiments concerns the *a priori* probabilities which were set to $p(1, 1) = 0.35, p(1, 2) = p(2, 1) = 0.15$ and $p(2, 2) = 0.35$ for data simulation. In that case, the automatic selection of copulas c_{12} and c_{21} shows no confusion. Nevertheless, these few confusions have a limited impact on the error rates:

- Experiment #1: the mean error rate is 13.74% for the unsupervised case and 12.16% for the supervised case (i.e. without parameters estimation, see Section 3.2).
- Experiment #2: the mean error rate is 12.89% for the unsupervised case and 11.35% for the supervised case.

5. Unsupervised image segmentation

This section is intended to illustrate the use of copula selection in PMC for unsupervised segmentation of images. We will focus on the JERS1 Synthetic Aperture Radar image of Rondonia, Brazil, in Fig. 5(a). It is a 3 looks amplitude image with 256 by 256 pixels, and 25 m by 25 m soil resolution. SAR images are known to be very challenging due to the speckle that noises the image.

Rondonia is a part of the Amazon where cultivation displaces the forest. In the Amazon, the prevalent method of cultivation, called "slash and burn" is made in the following way: first plots of dense forest are cut and burned. Then, plots of burnt land are put into cultivation and converted into meadows after two or three years. Then other plots of forest are cut, burned, and transformed into cultivated land. Finally, $K = 3$ classes should be considered in the extract 5(a): burn plot, cultivation, and dense forest, cf. Fig. 5(b).

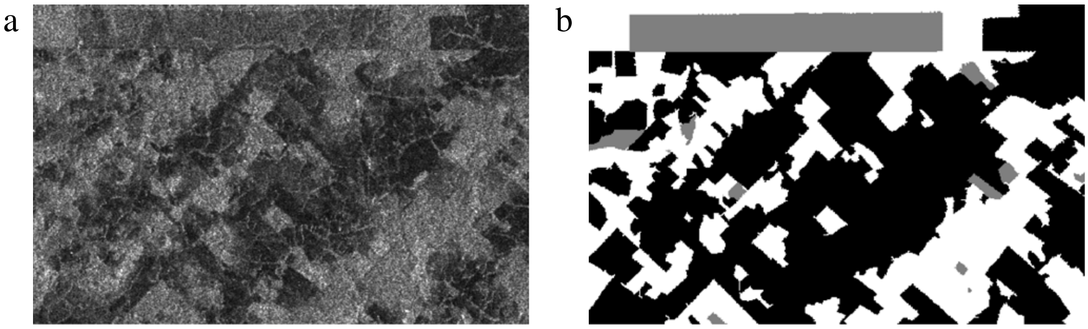


Fig. 5. (a) Three-look JERS1 image of Rondonia. (b) Image of classes, manually segmented by an expert.

Table 8
Detail of margin shapes. Types IV and VI refer to the Pearson' system of distributions.

	$c_{1,1}$		$c_{1,2}$		$c_{1,3}$		$c_{2,1}$		$c_{2,2}$		$c_{2,3}$		$c_{3,1}$		$c_{3,2}$		$c_{3,3}$	
Margin	f_{11}	f_{11}	f_{12}	f_{21}	f_{13}	f_{31}	f_{21}	f_{12}	f_{22}	f_{22}	f_{23}	f_{32}	f_{31}	f_{13}	f_{32}	f_{23}	f_{33}	f_{33}
Type	IV	IV	IV	VI	IV	VI	VI	IV	VI	VI	VI	VI	VI	IV	VI	VI	VI	VI

Table 9
Selected copula shapes and Kendall's tau for the image in Fig. 5(a). Notation c^i refers to Table A.11.

	$c_{1,1}$	$c_{1,2}$	$c_{1,3}$	$c_{2,1}$	$c_{2,2}$	$c_{2,3}$	$c_{3,1}$	$c_{3,2}$	$c_{3,3}$
Copula	c^1	c^4	c^1	c^4	c^1	c^4	c^1	c^4	c^1
Kendall's tau	0.29	0.15	0.21	0.17	0.25	−0.19	0.20	−0.17	0.21

As the PMC model is one dimensional, it is first required to convert the bi-dimensional image into a one dimensional vector of data. Following previous works Giordana and Pieczynski (1997) and Derrode and Pieczynski (2004), the transformation is made through the Hilbert–Peano (Skarbek, 1992) which is known to best preserve the neighborhood of pixels within the vector when compared to other classical scan such as the zig-zag one. Here are the main steps to process an image with the PMC model:

- The bi-dimensional lattice of pixels is first converted into a 1D sequence of observations through the Hilbert–Peano scan;
- Then, parameter estimation and Bayesian restoration techniques described in previous sections can be applied to obtain a restored sequence of class data;
- Finally, the restored sequence is converted back to a class image by applying the inverse Peano scan.

In this experiment, we compared the segmentation of the image 5(a) by a classical HMC-IN model and by the PMC one when different families of copulas were considered, using the error rate with respect to the ground-truth 5(b). Especially, we experimented the situations where all $K^2 = 9$ copulas are (a) Gaussian c^1, (b) Student c^2, (c) FGM c^4, (d) Clayton c^6, and (e) each of them are automatically selected within $\Pi = \left\{c^1, c^2, c^4, c^6\right\}$.

Regarding the choice of K^2 margin's shapes, and following results obtained by automatic selection in Delignon (2002), we set the margin laws to be of type IV and VI from Pearson' system of distributions (Johnson and Kotz, 1994), see details in Table 8. Recalling that due to the reversibility hypothesis of PMC models, the number of independent margins is divided by 2. Parameters of all margins were estimated by ICE. Also, the number of ICE iterations were set to 150 for all experiments, and the MPM criterion were used for classification.

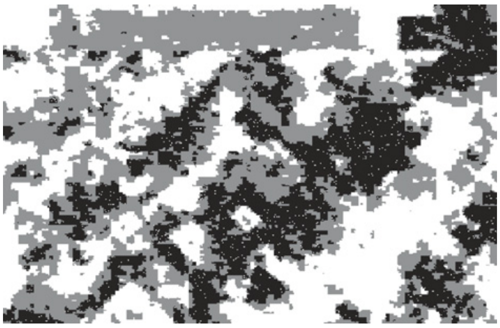
Remark. The Bayesian criterion presented in Section 4.1 allows only to deal with small sample size because of numeric overflow. In order to cope with this difficulty for image segmentation, the selection of copulas at each ICE iteration were performed on a sub-sample of maximum 1200 samples for each of the K^2 configurations.

Results of segmentation for the five cases described above plus the HMC-IN case are reported in Fig. 6. The lowest error rate is obtained when copulas are automatically selected (37.6%). The selected copulas are reported in Table 9. As can be seen, copulas for c_{ii} classes are Gaussian whereas copulas for c_{ij}, $i \neq j$ classes are Gaussian and FGM. Margin parameters are reported in Table 10, matrix of joint *a priori* probabilities $p(x_1, x_2)$ were estimated as

$$\begin{pmatrix} 0.22 & 7.90\ 10^{-3} & 9.13\ 10^{-3} \\ 7.90\ 10^{-3} & 0.37 & 4.04\ 10^{-3} \\ 9.13\ 10^{-3} & 4.04\ 10^{-3} & 0.37 \end{pmatrix},$$

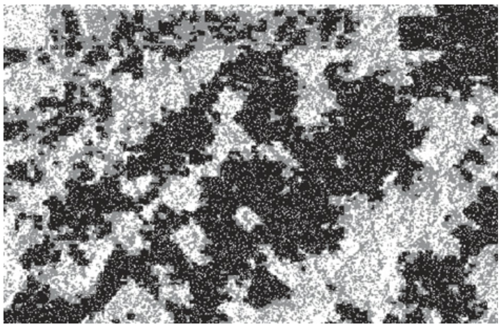
giving the estimated mixture in Fig. 7, according to Eq. (12).

	1	2	3
1	**43.7**	14.2	42.1
2	2.0	**78.0**	20.0
2	1.5	12.5	**86.1**



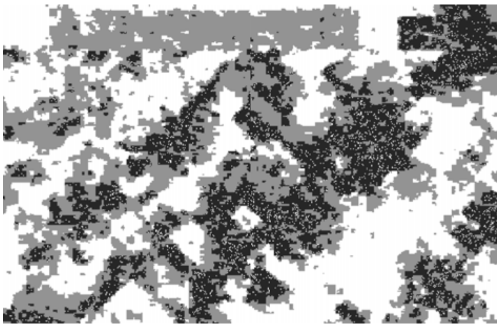
(a) PMC Gaussian 39.9%.

	1	2	3
1	**64.8**	16.0	19.1
2	8.0	**47.9**	44.1
2	35.2	14.6	**50.2**



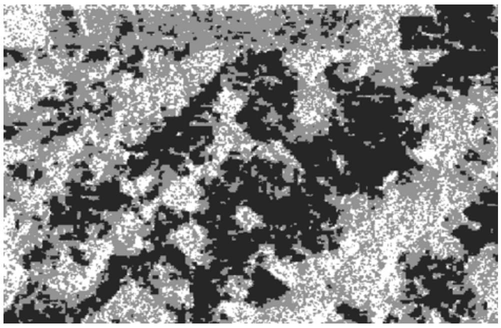
(b) PMC Student 42.7%.

	1	2	3
1	**39.7**	16.8	43.4
2	2.0	**78.4**	19.6
3	1.9	14.6	**83.5**



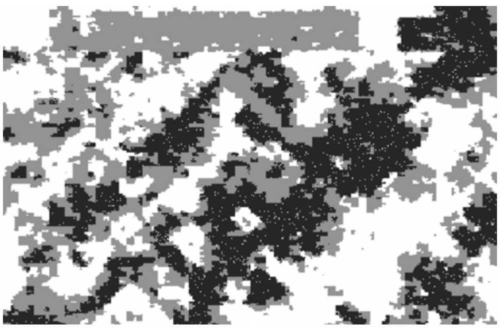
(c) PMC FGM 42.1%.

	1	2	3
1	**58.1**	7.6	34.3
2	4.7	**41.5**	53.8
3	14.8	7.7	**77.4**



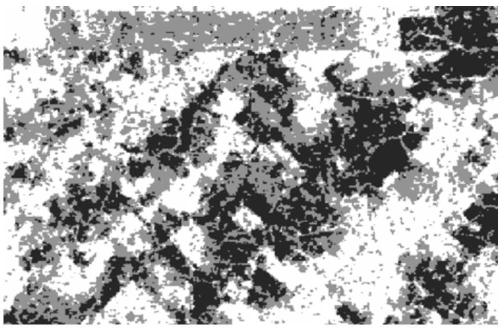
(d) PMC Clayton 46.2%.

	1	2	3
1	**47.6**	14.7	37.6
2	2.4	**78.6**	19.0
3	1.9	12.9	**85.2**



(e) PMC Selected copulas 37.6%.

	1	2	3
1	**41.5**	15.6	42.9
2	2.6	**73.7**	23.8
3	4.7	19.3	**76.0**



(f) HMC-IN 43.6%.

Fig. 6. Confusion matrices and segmentation results for the 5 PMC configurations (a–e) and the HMC-IN model (f).

As expected, the choice of copulas is critical since the quality of segmentation depends on them. A wrong choice of copulas can even give segmentation with lowest quality than a classical HMC-IN model (e.g. 46.2% for PMC with Clayton

Table 10
Estimated margin parameters for the image in Fig. 5(a). μ_i: four moments; β_1: skewness; β_2: kurtosis.

	μ_1	μ_2	μ_3	μ_4	β_1	β_2
f_{11}	49	$3.9\ 10^2$	$1.4\ 10^4$	$1.5\ 10^6$	3.00	9.6
f_{12}	61	$3.9\ 10^2$	$1.1\ 10^4$	$1.2\ 10^6$	2.20	7.6
f_{21}	33	97	$5\ 10^2$	$3.2\ 10^4$	0.27	3.4
f_{13}	40	$2.3\ 10^2$	$3.3\ 10^3$	$2.5\ 10^5$	0.91	4.8
f_{31}	18	24	-39	$1.8\ 10^3$	0.11	3.2
f_{22}	74	$6.5\ 10^2$	$1.3\ 10^4$	$1.7\ 10^6$	0.58	4.1
f_{23}	73	$4.2\ 10^2$	$-1.2\ 10^2$	$5.3\ 10^5$	0.00	3.0
f_{32}	110	$5.5\ 10^2$	$1.6\ 10^4$	$1.6\ 10^6$	1.40	5.2
f_{33}	112	$1.1\ 10^3$	$2.7\ 10^4$	$4.8\ 10^6$	0.53	3.9

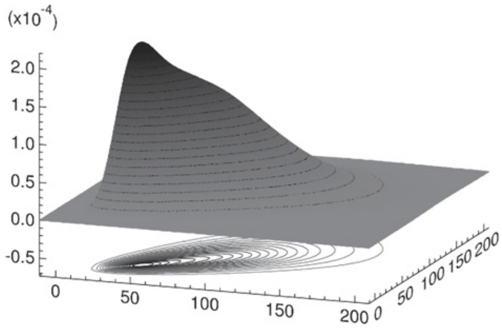


Fig. 7. Mixture estimated by ICE with copulas selected within $\Pi = \left\{c^1, c^2, c^4, c^6\right\}$ after 150 iterations.

copulas versus 43.6% for HMC-IN). In real-word applications, of course, we either do not know the true copula or, even, the copula is used as an approximation to the unknown probability law governing the process. So what happen if the finite set of admissible copulas does not include the true copula or even include a close approximation? As experimented in Section 3.2 and confirmed here, we can expect a severe degradation of results. To prevent from such a situation, it is important to construct admissible sets of copulas rich enough in term of shape diversity.

6. Conclusion

This paper examined the influence of copula shapes in the pairwise Markov chain model. We have first presented supervised restoration results when systematically interchanging copulas used for PMC data simulation with other copula families. All things being equal, the use of the false copula can degrade significantly segmentation results, both in the Markovian and non Markovian contexts. In the case of a noise with strong correlation, the use of an independent noise model can produce poor results. Also the Markovianity allows significant improvements when compared to a model that takes into account a single neighbor (i.e. Pairwise Mixture Model).

We also presented an algorithm for the automatic selection of copulas involved in PMC within a finite set of admissible copulas. According to experiments, the method of copula identification and parameters estimation, based on the general ''Iterated Conditional Expectation'' (ICE) method, allows effective unsupervised classifications, as illustrated by the segmentation of a SAR image. In real-word applications, as copula is mainly used as an approximation to the unknown probability law governing the process, it is important that the set of candidate copulas being diversified in term of shapes.

As perspective, let us mention the possibilities of extension of the copulas selection method proposed to more complex Markov models, like triplet Markov Chains (Lanchantin et al., 2011), hidden non stationary semi-Markov chains (Lapuyade-Lahorgue and Pieczynski, 2012), or still triplet Markov chains hidden with long correlation noise (Lanchantin et al., 2008), which are recent extensions of the HMC and in which the use of classical ICE gave interesting results.

Appendix A. Copulas used in this work

Table A.11 gives details about the one-parameter copulas considered for experiments: probability density function (pdf), range for parameter θ and its closed-form solution to Kendall's tau. We write either $c^p\,(u_1, u_2; \theta)$ or $c^p\,(u_1, u_2; \tau)$. For the Student't copula (c^2) the degree of freedom ν is supposed to be known.

Table A.11
One parameter (named θ) copulas $c^p\,(u_1, u_2; \theta)$ used in this report (FGM stands for Farlie–Gumbel–Morgenstern).

$\#p$	Name	cdf C^p	pdf c^p	Kendall's tau	$[\tau_m^p, \tau_M^p]$
0	Product	$C^0 = u_1\, u_2$	$c^0 = 1$	0	
1	Gauss[a]	$C^1 = \int_0^{u_1} \phi\left(\frac{\phi^{-1}(u_2)-\rho\phi^{-1}(u)}{\sqrt{1-\rho^2}}\right) du$	$c^1 = \frac{1}{\sqrt{1-\theta^2}}\ \exp\left(-\frac{1}{2}\ \boldsymbol{\xi}^T\ (\boldsymbol{\rho}-\boldsymbol{I})\ \boldsymbol{\xi}\right)$	$\frac{2}{\pi}\operatorname{asin}\theta$	$[-1, 1]$
		where $\xi_i = \phi^{-1}(u_i)$ with ϕ the standard normal distribution, $\boldsymbol{\rho} = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}$ and $\boldsymbol{I}$ are the 2×2 correlation and identity matrices.			
2	Student[a]	$C^2 = \int_0^{u_1} t_{\nu+1}\left(\sqrt{\frac{\nu+1}{\nu+(t_\nu^{-1}(u_1))^2}}\ \frac{t_\nu^{-1}(u_2)-\rho t_\nu^{-1}(u)}{\sqrt{1-\rho^2}}\right) du$	$c^2 = \frac{1}{\sqrt{1-\theta^2}}\ \frac{\Gamma\left(\frac{\nu}{2}+1\right)\Gamma\left(\frac{\nu}{2}\right)}{\Gamma^2\left(\frac{\nu+1}{2}\right)}\ \left(1+\frac{1}{\nu}\boldsymbol{\xi}^T\ \boldsymbol{\rho}^{-1}\ \boldsymbol{\xi}\right)^{-\frac{\nu+2}{2}}\ \prod_{i=1}^2 \left(1+\frac{\xi_i^2}{\nu}\right)^{\frac{\nu+1}{2}}$	$\frac{2}{\pi}\operatorname{asin}\theta$	$[-1, 1]$
		where $\xi_i = t_\nu^{-1}(u_i)$ with t_ν the t distribution with ν degrees of freedom, $\boldsymbol{\rho} = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}$ is the 2×2 correlation matrix and $\Gamma(.)$ the Gamma function.			
3	Gumbel[b]	$C^3 = \exp(-(U_1+U_2)^{\frac{1}{\theta}})$ where $U_1 = (-\ln(u_1))^\theta$ and $U_2 = (-\ln(u_2))^\theta$.	$c^3 = \frac{U_1}{u_1\ \ln(u_1)}\ \frac{U_2}{u_2\ \ln(u_2)}\ (\theta-1+U_1+U_2)^{\frac{1}{\theta}}\ (U_1+U_2)^{\frac{1}{\theta}-2}\ \exp(-(U_1+U_2)^{\frac{1}{\theta}})$	$1-\frac{1}{\theta}$	$[0, 1]$
4	FGM	$C^4 = u_1 u_2\,(1+\theta(1-u_1)(1-u_2))$	$c^4 = 1+\theta\,(1-2u_1)\ (1-2u_2)$	$\frac{2\theta}{9}$	$\left[-\frac{2}{9}, \frac{2}{9}\right]$
5	Cubic section	$C^5 = u_1 u_2(1+2\theta(1-u_1)(1-u_2)(1+u_1+u_2-2u_1u_2))$	$c^5 = 1+2\theta\Big((1-u_1)(1-u_2)(-8u_2u_1+2u_1+2u_2+1)+u_1(1-u_2)(4u_2u_1-u_1-2u_2-1)+(1-u_1)u_2(4u_2u_1-2u_1-u_2-1)+u_1u_2(-2u_2u_1+u_1+u_2+1)\Big)$	$\frac{2}{3}\theta-\frac{6}{225}\theta^2$	$\left[0, \frac{33}{200}\right]$
6	Clayton[b]	$C^6 = \left(u_1^{-\theta}+u_2^{-\theta}-1\right)^{-\frac{1}{\theta}}$	$c^6 = (1+\theta)\ u_1^{-1-\theta}\ u_2^{-1-\theta}\ \left(-1+u_1^{-\theta}+u_2^{-\theta}\right)^{-\frac{1}{\theta}-2}$	$\frac{\theta}{\theta+2}$	$]0, 1]$
7	Arch12[b,c]	$C^7 = \left(1+(U_1+U_2)^{\frac{1}{\theta}}\right)^{-1}$	$c^7 = \frac{U_1}{u_1(u_1-1)}\ \frac{U_2}{u_2(u_2-1)}\ \left(\theta-1+(\theta+1)\ (U_1+U_2)^{\frac{1}{\theta}}\right)\ \frac{(U_1+U_2)^{\frac{1}{\theta}-2}}{\left(1+(U_1+U_2)^{\frac{1}{\theta}}\right)^3}$	$1-\frac{2}{3\theta}$	$\left[\frac{1}{3}, 1\right]$
		where $U_1 = \left(\frac{1}{u_1}-1\right)^\theta$ and $U_2 = \left(\frac{1}{u_2}-1\right)^\theta$			
8	Arch14[b,c]	$C^8 = \left(1+(U_1+U_2)^{\frac{1}{\theta}}\right)^{-\theta}$	$c^8 = U_1\ U_2\ \ (U_1+U_2)^{\frac{1}{\theta}-2}\ \left(1+(U_1+U_2)^{\frac{1}{\theta}}\right)^{-2-\theta}\ \frac{\left(\theta-1+2\theta(U_1+U_2)^{\frac{1}{\theta}}\right)}{\theta u_1 u_2\left(u_1^{\frac{1}{\theta}}-1\right)\left(u_2^{\frac{1}{\theta}}-1\right)}$	$1-\frac{2}{3\theta}$	$\left[\frac{1}{3}, 1\right]$
		where $U_1 = \left(u_1^{-\frac{1}{\theta}}-1\right)^\theta$ and $U_2 = \left(u_2^{-\frac{1}{\theta}}-1\right)^\theta$			

[a] Family of elliptical copulas.
[b] Family of Archimedean copulas.
[c] Coin from the order of appearance in Nelsen (2005).

Table B.12
Closed-form solutions for $u_2* = \arg\max_{u_2 \in [0,1]} c(u_1, u_2)$ and $M = c(u_1, u_2*)$ for several copulas in Table A.11 ($u_1 \in [0, 1]$); see also Fig. B.8.

#i	Name	u_2*	M
0	Product	–	1
1	Gaussian	$\phi\left(\frac{\phi^{-1}(u_1)}{\theta}\right)$	$\frac{1}{\sqrt{1-\theta^2}} \exp\left(\frac{1}{2}\left[\phi^{-1}(u_1)\right]^2\right)$
2	Student't	cf. Appendix C	
3	Gumbel	No closed-form solution found → numerical solution	
4	FGM	$\begin{cases} 0 & \text{if } (1-2u_1)\theta > 0 \\ 1 & \text{if } (1-2u_1)\theta < 0 \\ - & \text{else} \end{cases}$	$1 + \max\left((1-2u_1)\,\theta, -(1-2u_1)\,\theta\right)$
5	Cubic section	$\begin{cases} 0 & \text{if } 0 \le u_1 < \frac{1}{2} \\ - & \text{if } u_1 = \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} < u_1 \le 1 \end{cases}$	$\begin{cases} 1 + 2\theta\left(-3u_1^2 + 1\right) & \text{if } 0 \le u_1 \le \frac{1}{2} \\ 1 + 2\theta\left(-3u_1^2 + 6u_1 - 2\right) & \text{if } \frac{1}{2} \le u_1 \le 1 \end{cases}$
6	Clayton	$\min\left(1, \left(\frac{\theta+1}{\theta}\left(u_1^{-\theta} - 1\right)\right)^{-\frac{1}{\theta}}\right)$	$\begin{cases} (1+\theta)\, u_1^{\theta} & \text{if } u_2* = 1 \\ \left(\frac{\theta+1}{2\theta+1}\right)^{\frac{2\theta+1}{\theta}} \frac{\theta}{u_1\left(1-u_1^{\theta}\right)} & \text{else} \end{cases}$
7	Arch12	No closed-form solution found → numerical solution	
8	Arch14	No closed-form solution found → numerical solution	

Appendix B. Rejection algorithm to simulate $Y_2|Y_1 = y_1$ in a copula framework

Let Y_1 and Y_2 two real-valued random variables with probability density function $f^{(1)}(.)$ and $f^{(2)}(.)$, and cumulative distribution functions $F^{(1)}(.)$ and $F^{(2)}(.)$. Assuming a copula representation for $f(y_1, y_2)$, we can write

$$p(y_2 | y_1) = f_{y_1}(y_2) = f^{(2)}(y_2)\; c\left(F^{(1)}(y_1), F^{(2)}(y_2); \boldsymbol{\theta}\right). \tag{B.1}$$

Assuming a real number M and a density g such that $\forall x \in \mathbb{R}, f_{y_1}(x) \le M\, g(x)$, the simulation of Y_2 conditionally to $Y_1 = y_1$ can be performed using the rejection algorithm (Devroye, 1986):

1. Sample $X = x$ according to g and $V = v$ according to $\mathcal{U}([0, 1])$, the uniform law.
2. Accept $y_2 = x$ if

$$v \le \frac{f_{y_1}(x)}{M\, g(x)}. \tag{B.2}$$

3. Else, go back to 1.

Choosing $g = f^{(2)}$ and $M = \max_{u_2 \in [0,1]} c(u_1, u_2)$, Eq. (B.2) now writes

$$v \le \frac{c\left(u_1, F^{(2)}(x); \boldsymbol{\theta}\right)}{\max\limits_{u_2 \in [0,1]} c(u_1, u_2; \boldsymbol{\theta})},$$

where $u_1 = F^{(1)}(y_1)$.

So, it is possible to generate drawings from $Y_2|Y_1 = y_1$ whatever the shape of the copula, once we know the pdf of the copula and how to generate a random variate for Y_2. Algorithm efficiency, i.e. number of rejections before an acceptance occurs, depends on the value of M and so on the copula shape. Table B.12 shows analytical solutions for M for several copulas studied here. Appendix C details calculations for the Student't copula. When no closed-form solution is available, a numerical method can be easily implemented to find the maximum value.

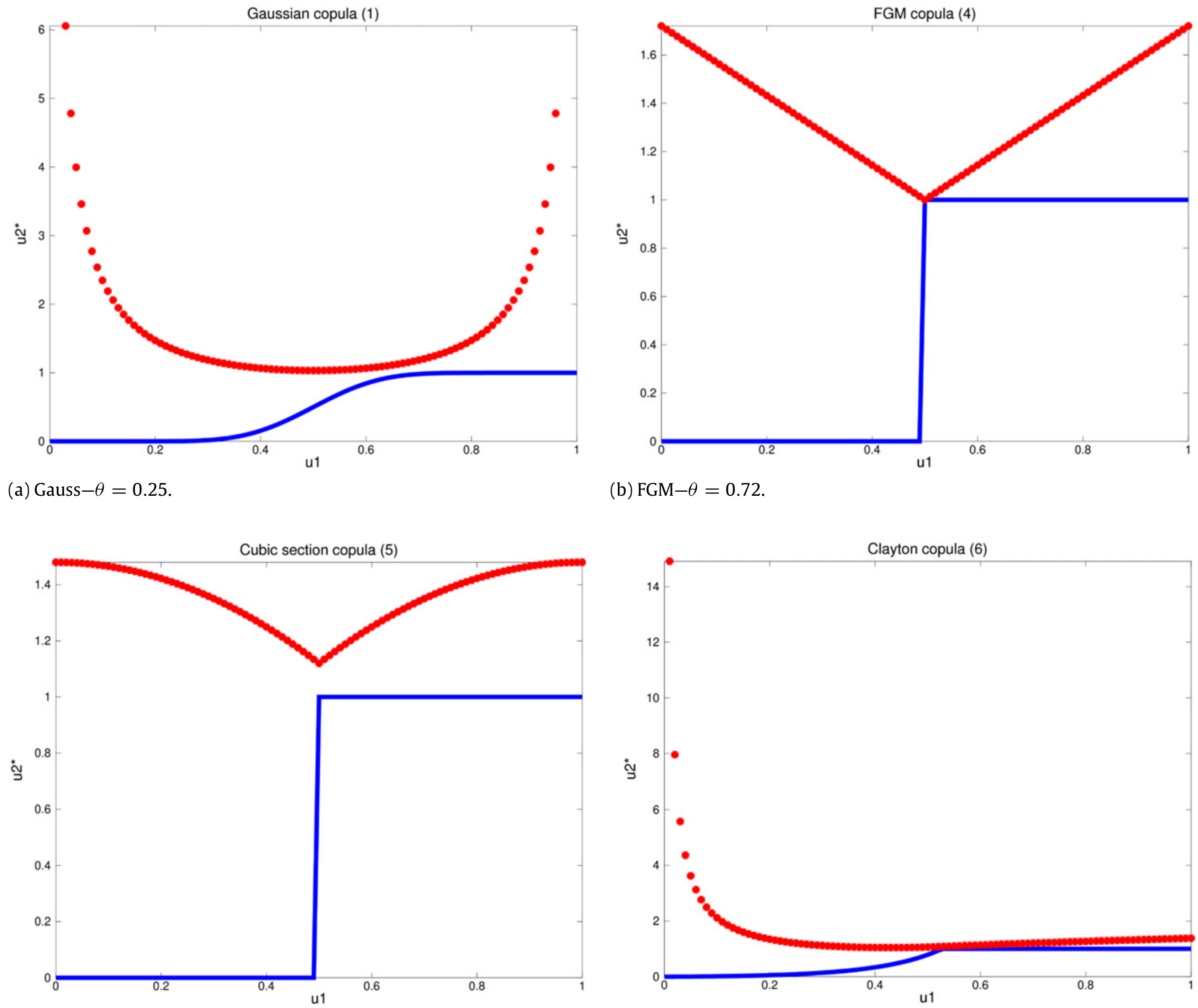
Appendix C. Calculation of $u_2* = \arg\max_{u_2 \in [0,1]} c^2(u_1, u_2)$

This appendix details the calculation of $u_2* = \arg\max_{u_2 \in [0,1]} c^2(u_1, u_2)$ for the Student't copula. From $\frac{\partial c^2(u_1, u_2)}{\partial u_2} = 0$, we get

$$-\xi_2^3 - \theta \xi_1 \nu\, \xi_2^2 + \left((\nu+1)\left(\nu\,(1-\theta^2) + \xi_1^2\right) - \nu\,(\nu+2)\right)\xi_2 + \theta \nu \xi_1\,(\nu+2) = 0, \tag{C.1}$$

where $\xi_i = t_\nu^{-1}(u_i)$. Using Tartaglia–Cardan method for solving third order equations,

$$a = -1, \qquad b = -\theta \xi_1 \nu, \qquad c = (\nu+1)\left(\nu\,(1-\theta^2) + \xi_1^2\right) - \nu\,(\nu+2), \qquad d = \theta \nu \xi_1\,(\nu+2)$$



(a) Gauss–$\theta = 0.25$.

(b) FGM–$\theta = 0.72$.

(c) Cubic section–$\theta = 0.24$.

(d) Clayton–$\theta = 0.38$.

Fig. B.8. Curves $u_2 * (u_1)$ (solid, in blue) and $M(u_1)$ (sign 'o', in red) for some copulas in Table B.12. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Eq. (C.1) writes $z^3 + pz + q = 0$, with $\xi_2 = z - \frac{b}{3a}$ and

$$p = -\frac{b^2}{3a^2} + \frac{c}{a}, \qquad q = \frac{b}{27a}\left(\frac{2b^2}{a^2} - \frac{9c}{a}\right) + \frac{d}{a}, \qquad \Delta = q^2 + \frac{4}{27}p^3.$$

The solution to this third order equation depends on the sign of Δ:

- If $\Delta > 0$, there is an unique real-valued solution:

$$z_0 = \sqrt[3]{\frac{-q + \sqrt{\Delta}}{2}} + \sqrt[3]{\frac{-q - \sqrt{\Delta}}{2}}.$$

  Note that each term of the sum can be complex, but the sum of the two terms is real-valued. The solution finally writes

$$u_2 * = t_\nu\left(z_0 - \frac{b}{3a}\right).$$

- if $\Delta \leq 0$, there are three real-valued solutions:

$$\forall k \in \{0, 1, 2\}, \quad z_k = 2\sqrt{\frac{-p}{3}} \cos\left(\frac{1}{3}\arccos\left(\frac{-q}{2}\sqrt{\frac{27}{-p^3}}\right) + \frac{2k\pi}{3}\right).$$

The solution to be retained among the three possible values $s_k = t_\nu\left(z_k - \frac{b}{3a}\right)$ is the one which gives the higher value

$$u_2{*} = \arg\max_{k\in\{0,1,2\}} c^2\left(u_1, s_k\right).$$

The solution is not tractable analytically but the highest value can be easily computed.

In both cases, the analytical solution for $M = c^2\left(u_1, u_2{*}\right)$ is not tractable but can also be easily computed.

References

Brunel, N., Lapuyade-Lahorgue, J., Pieczynski, W., 2010. Modeling and unsupervised classification of multivariate hidden Markov chains with copulas. IEEE Transactions on Automatic Control 55 (2), 338–349.

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

Celeux, G., Diebolt, J., 1985. The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. Computational Statistics Quarterly 2, 73–82.

Delignon, Y., 2002. Approche statistique en traitement du signal. Applications à l'imagerie radar et aux systèmes de communications sans fil. Ph.D. Thesis. Lille University, IEMN, September 10.

Delmas, J., 1997. An equivalence of the EM and ICE algorithms for exponential family. IEEE Transactions on Signal Processing 45 (10), 2613–2615.

Demarta, S., McNeil, A., 2005. The t-copula and related copulas. International Statistical Review 73 (1), 111–129.

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

Derrode, S., Pieczynski, W., 2011. Unsupervised restoration in Gaussian pairwise mixture model. In: Proc. of the European Signal Processing Conf., EUSIPCO'11, Barcelona, Spain.

Destrempes, F., Mignotte, M., 2004. A statistical model for contours in images. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (5), 626–638.

Devroye, L., 1986. Non-Uniform Random Variate Generation. Springer-Verlag, New York, USA.

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

Genest, C., Mackay, J., 1986. Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. Canadian Journal of Statistics 14 (2), 145–159.

Genest, C., Masiello, E., Tribouley, K., 2009. Estimating copula densities through wavelets. Insurance: Mathematics and Economics 44, 170–181.

Genest, C., Rémillard, B., Beaudoin, D., 2009. Goodness-of-fit tests for copulas: a review and a power study. Insurance: Mathematics and Economics 44, 199–213.

Giordana, N., Pieczynski, W., 1997. Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (5), 465–475.

Huard, D., Évin, G., Favre, A., 2006. Bayesian copula selection. Computational Statistics & Data Analysis 51, 809–822.

Iyengar, S., Varshney, P., Damarla, T., 2011. A parametric copula-based framework for hypothesis testing using heterogeneous data. IEEE Transactions on Signal Processing 59 (5), 2308–2319.

Johnson, N., Kotz, S., 1994. Continuous Univariate Distribution, Tome I and II, second ed. Wiley-Interscience.

Lanchantin, P., Lapuyade-Lahorgue, J., Pieczynski, W., 2008. Unsupervised segmentation of triplet Markov chains hidden with long-memory noise. Signal Processing 88 (5), 1134–1151.

Lanchantin, P., Lapuyade-Lahorgue, J., Pieczynski, W., 2011. Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise. Signal Processing 91 (2), 163–175.

Lapuyade-Lahorgue, J., Pieczynski, W., 2012. Unsupervised segmentation of hidden semi-Markov non stationary chains. Signal Processing 92 (1), 29–42.

Le Cam, S., Belghith, A., Salzenstein, F., Collet, C., 2009. Wheezing sounds detection using multivariate generalized Gaussian distributions. In: IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP'09, Taipei, Taiwan, 18–25 April.

Mercier, G., Moser, G., Serpico, S., 2008. Conditional copula for change detection on heterogeneous SAR data. IEEE Transactions on Geoscience and Remote Sensing 46 (5), 1428–1441.

Nelsen, R., 2005. An Introduction to Copulas, second ed. In: Springer Series in Statistics, Springer, New York, USA.

Nikoloulopoulos, A., Karlis, D., 2008. Copula model evaluation based on parametric bootstrap. Computational Statistics & Data Analysis 52, 3342–3353.

Noh, Y., Choi, K., Lee, I., 2010. Identification of marginal and joint CDFs using Bayesian method for RBDO. Structural and Multidisciplinary Optimization 40, 35–51.

Ntantamis, C., 2010. Detecting structural breaks using hidden Markov models. CREATES Research Papers 2010-52. School of Economics and Management, University of Aarhus. URL: http://ideas.repec.org/p/aah/create/2010-52.html.

Pieczynski, W., 1992. Statistical image segmentation. Machine Graphics and Vision 1 (1–2), 261–268.

Pieczynski, W., 2003. Pairwise Markov chains. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 634–639.

Sakji-Nsibi, S., Benazza-Benyahia, A., 2009. Copula-based statistical models for multicomponent image retrieval using a Bayesian copula selection. In: Proc. of the IEEE Int. Conf. on Image Processing, ICIP'09, Cairo, Egypt, 7–10 November, pp. 253–256.

Salzenstein, F., Collet, C., Le Cam, S., Hatt, M., 2007. Non stationnary fuzzy Markov chain. Pattern Recognition Letters 28 (16), 2201–2208.

Skarbek, W., 1992. Generalized Hilbert scan in image printing. In: Klette, R., Kropetsh, W.G. (Eds.), Theoretical Foundations of Computer Vision. Akademie Verlag, Berlin, Germany (Chapter).

Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris 8, 229–231.

Stitou, Y., Lasmar, N., Berthoumieu, Y., 2009. Copulas based multivariate gamma modeling for texture classification. In: IEEE Int. Conf on Acoustics, Speech and Signal Processing, ICASSP'09, pp. 1045–1048.

Sundaresan, A., Varshney, P., 2011. Location estimation of a random signal source based on correlated sensor observations. IEEE Transactions on Signal Processing 59 (2), 787–799.

Wua, Y., Li, M., Zhang, P., Zong, H., Xiao, P., Liu, C., 2011. Unsupervised multi-class segmentation of SAR images using triplet Markov fields models based on edge penalty. Pattern Recognition Letters 32, 1532–1540.

Xiaomei, Q., Jie, Z., Xiaojing, S., 2010. Archimedean copula estimation and model selection via l1-norm symmetric distribution. Insurance: Mathematics and Economics 46, 406–414.