Modeling and Unsupervised Classification of Multivariate Hidden Markov Chains With Copulas

N. J.-B. Brunel, Jerome Lapuyade-Lahorgue, and Wojciech Pieczynski

Abstract—Parametric modeling and estimation of non-Gaussian multidimensional probability density function is a difficult problem whose solution is required by many applications in signal and image processing. A lot of efforts have been devoted to escape the usual Gaussian assumption by developing perturbed Gaussian models such as Spherically Invariant Random Vectors (SIRVs). In this work, we introduce an alternative solution based on copulas that enables theoretically to represent any multivariate distribution. Estimation procedures are proposed for some mixtures of copula-based densities and are compared in the hidden Markov chain setting, in order to perform statistical unsupervised classification of signals or images. Useful copulas and SIRV for multivariate signal classification are particularly studied through experiments

Index Terms—Copulas, EM algorithm, hidden Markov chains, hidden Markov models, inference for margins, maximum likelihood, multivariate modeling, spherically invariant random vector (SIRV), statistical classification.

I. INTRODUCTION

ONE of the main problems in the statistical analysis of multi-component images and multi-dimensional signals is the choice of relevant statistical parametric laws. This difficulty is usually overcome by an assumption of Gaussianity, which is often justified by the use of the central limit theorem, the maximum entropy principle, or the interpretability of the parameters. As a last resort, the tractability of the formulae for estimation, filtering, classification, can be a self-justification for the use of Gaussian laws. Nevertheless, in numerous applications, the non-Gaussianity cannot be neglected and other laws have to be used. For example, different particular non-Gaussian laws have been relevantly proposed in radar signal processing [5], [13], [38]. Among others, the presence of heavy tailed distributions can present a serious limitation for applications of Gaussian models, and the use of the family of laws called "stable laws", which include various heavy tailed distributions, is an interesting alternative [1]. Another problem that can also be encountered is the choice of a multivariate law that is compatible with some previous knowledge. For instance, such a previous constraint can be due to physical knowledge of the phenomena involved, implying some knowledge of the statistical behavior of each component of the multivariate distribution considered.

In particular, this modeling problem occurs when wanting consider unsupervised Bayesian classification procedures, which is the subject of this paper. For the classification of scalar data, it is usually assumed that the expected classes differ from each other by a mean level and an inner degree of dispersion. The Gaussian hypothesis is an illustration of this a priori knowledge on the classes, and the use of Gaussian vectors for multivariate data shows that the same implicit assumption is made for multivariate signals. This idea is not limited to the multivariate Gaussian law, and wider families such as the "elliptical laws", which will be specified below, are indexed by mean and covariance parameters. However, this can be restrictive, because the difference between the classes can depend on the ways the different components are linked, independently of their mean level and of their inner dispersion. Moreover, the choice of multivariate models is far more restricted than for univariate one, especially when one wants to respect some constraints on the law of each component. Therefore, for a given class, the general problem is to model the corresponding multivariate distribution in such a way that the various components are correlated, are not necessarily Gaussian, and the marginal distributions of the different components can differ from each another. A very stimulating answer to this general problem is provided by the theory of copulas [31]. Indeed, copulas enable us to widen the ability of multivariate modeling by separating the problem of finding adequate forms of the marginal distributions, and finding adequate dependence structure of the vector. Therefore, it is possible to cross different classical margins and different dependence models in order to obtain a large variety of original multivariate models. Moreover, it also underlines the influence of dependence between the components for the characterization of the classes.

The aim of this paper is to propose some original models simultaneously using multivariate hidden Markov chains and copulas, with original parameter estimation methods. The new models and related new parameter estimation methods enable us to propose unsupervised image and signal classification, whose interest is validated by some experiments.

The paper is organized as follows. In Section II, we recall the definition and some properties of Spherically Invariant Random Vector (SIRV) models and give two major examples encountered in signal and image processing, with related parameter

estimation methods. The main properties of copulas are also recalled in the Section II. Section III is devoted to Bayesian unsupervised classification with hidden Markov chains, and we present original unsupervised learning methods for these models. Different experiments showing the interest of the new modeling and of the associated processing are presented in Section IV. Finally, conclusions are drawn in the final Section.

II. MULTIVARIATE MODELING AND ESTIMATION IN SINGLE CLASS CASE

There are two sub-sections in the present section. In the first one, we recall two classical multivariate parametric models and describe some associated estimation procedures in the case of independent observations. Some novelties concerning parameter estimation are also proposed for the K law. In the second one we introduce copulas and we recall some of their classical properties. Therefore there are neither Markov models nor numerous classes in this section, which will be studied in the third one.

A. Spherically Invariant Random Vector

Definitions and Examples: Let $Y = (Y^1, \ldots, Y^M)$ be a Gaussian random vector taking its values in R^M, $M \geq 2$, with mean $m = (m_1, \ldots, m_M)$ and covariance matrix $\Sigma = (\sigma_{ij})_{1 \leq i,j \leq M}$. This Gaussian law, which will be denoted by $N(m, \Sigma)$, is widely used in different statistical models and processing; in particular, in statistical signal and image processing. However, the stochastic behavior of some phenomena can deviate from the normal law, and, for instance, such is the case in remote sensing data. Spherically Invariant Random Vectors (SIRV) model is a possible generalization of the Gaussian one. Its law can be seen as a modification of the normal law $N(m, \Sigma)$ due to a random fluctuation, modeled by a strictly non-negative real random variable U, of the covariance matrix. More precisely, a vector Y is called a SIRV if there exists a strictly non-negative real random variable U and $V \sim N(0, \Sigma)$ such that $Y = m + U^{-1/2}V$. Let us notice that in radar signal processing, U and V are sometimes called "texture" and "speckle", respectively.

Let us recall some basic properties of SIRVs [35], [38]. We will assume that the distribution of U admits a density g with respect to the Lesbegue measure. Setting $q(y) = (1/2)(y - m)'\Sigma^{-1}(y - m)$, where $(y - m)'$ is the transpose of the vector $(y - m)$, one can see classically that the density of the distribution of the couple (Y, U) is $p(y, u) = g(u)p(y \mid u) = g(u)u^{M/2}|\Sigma|^{-1/2}/(2\pi)^{M/2}] \exp[-uq(y)]$. Thus the marginal density, which is the density of the distribution of a SIRV, has the following integral expression:

$$p(y) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{M/2}} \int_0^\infty u^{M/2} \exp[-uq(y)]g(u)du. \quad (1)$$

Therefore we see, according to (1), that the general expression of the density $p(y)$ is of the form $p(y) = |\Sigma|^{-1/2} h[(y - m)'\Sigma^{-1}(y - m)]$, with h being the integrable non-negative function defined from the density g and the integer M by $h(a) = (2\pi)^{-M/2} \int_0^{+\infty} u^{M/2} \exp[-au]g(u)du$. Thus the model distribution is indexed by a triplet (m, Σ, g). It is sometimes said that SIRVs belong to the family of "elliptical models" or "elliptically contoured densities models" since the densities have elliptical iso-contours, which means that they are constant for y such that $(y - m)'\Sigma^{-1}(y - m)$ is constant.

The parameter m is a location parameter and the matrix Σ is a scatter parameter; however, it is important to notice that they are, in general, neither a mean vector nor a covariance matrix. Let us notice that the classical SIRVs consist of the sub-family with mean equals to 0. The location parameter m is introduced for greater generality since it does not change the main features of centered SIRVs. Otherwise, SIRVs present the two following important properties:

(i) all the components $Y^1, \ldots, Y^M$ have the same marginal distribution;
(ii) the functional expression of the joint law of a SIRV Y is determined by its marginal laws [35].

The consequence of the second claim is that under the spherical invariance assumption, the marginal behaviors of channels imply the way that they are interacting. This assumption can be true for some kinds of data or signal—as coherent pulses for radar signals—but in the case of more general data, it can turn out to be questionable. We will see below how copulas make possible to deal with such more general cases.

Remark 2.1: Let us notice that the same density $p(y)$ can be defined by different triplets (m, Σ, g); for example, (m, Σ, g) and $(m, \lambda\Sigma, g_\lambda)$, where g_λ is defined from g by $g_\lambda(u) = 1/\lambda g(u/\lambda)$, give rise to the same density $p(y)$. In order to obtain a one-to-one correspondence between $p(y)$ and the parameters, we must consider some constraint on the mean of U, or on the scale of Σ. For instance, one possible condition is $Tr(\Sigma) = M$.

A closed-form expression for the density $p(y)$ can be derived for particular laws of U. In this paper, we consider two such (well-known) cases: the K law and the Student's T law. These distributions are related to the G distribution. We recall that the G distribution $\gamma(a, b)$ has a density equal to $f(y) = (1/\Gamma(a)b^a) \exp[-y/b]y^{a-1}1_{y \geq 0}$ where $a, b > 0$ and Γ is the Euler's function [2]. We have the following classical result: if the distribution of U is $\gamma(\nu/2, 2/\nu)$—which will be denoted by $U \sim \gamma(\nu/2, 2/\nu)$—then the distribution of Y is the following "Student law with ν degrees of freedom" (also called "T law"):

$$p(y) = \frac{2\Gamma\left(\frac{\nu+M}{2}\right)|\Sigma|^{-1/2}}{(\pi\nu)^{M/2}\Gamma(\frac{\nu}{2})} \left(1 + \frac{2q(y)}{\nu}\right)^{-(\nu+M)/2}. \quad (2)$$

Let us note that if $\nu > 2$, m is the mean of the T law and $\nu\Sigma/(\nu - 2)$ is its covariance matrix. The lower ν is, the thicker the tails are and, for the extreme case $\nu = 2$, we obtain the Cauchy law which does not even admit a mean, and which is one of the few stable laws with a closed-form density.
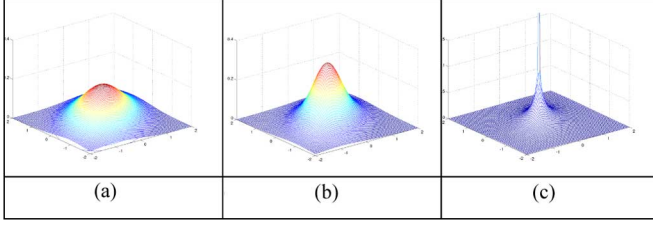
Fig. 1.   2-D and contour plots of the Normal, T and K density with zero mean and $\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ (a) Normal density (b) T density with $\nu = 10$ (c) K density with $\nu = 2$.

Concerning the second case we deal with, it is possible to show that if $U^{-1} \sim \gamma(\nu/2, 2/\nu)$, then Y has the following " K distribution":

$$p(y) = \frac{2 \left(\frac{\nu}{2}\right)^{\nu/2} |\Sigma|^{-1/2}}{(2\pi)^{M/2}\Gamma(\frac{\nu}{2})} \left(\sqrt{\frac{2q(y)}{\nu}}\right)^{(\nu-M)/2}$$
$$\times K_{(\nu-M)/2}\left(\sqrt{2\nu q(y)}\right) \quad (3)$$

where K_α is the modified Bessel function of the second kind [2].

The K law behaves better than the T law because it has mean m and covariance matrix Σ for all $\nu > 0$. When $\nu = 2$, we have a generalized Laplace law and K tends to a Dirac distribution at m when ν goes to zero. Classical asymptotic approximations of the K function enable to show that the tails are equivalent to $\exp[-x]/x$ for x going to infinity, which implies heavier tails than the Gaussian law [2]. Some plots of the densities in dimension 2 are given in Fig. 1.

In conclusion, the K and T laws are interesting SIRV models because they have closed form densities and they can still be interpreted in the same manner as the Gaussian law. However, they are richer because of an additional tuning parameter for the heaviness of the tails, which enables us to cope more easily with outliers. For the two families, the Gaussian law is a limit case when ν tend to infinity, but they differ significantly from each other by their behavior around the mean when ν decreases: the K law becomes sharper while the T law becomes flatter.

1) Parameter Estimation With EM Algorithm: The estimation of SIRVs has been addressed in number of papers, especially for the T law (see [28], [29], and references therein) and the K law (for radar detection [5], [6], [12], [18]). For the latter, the method of moments is widely used, as recalled in [23]. The same method can also furnish estimators for the KUBW family obtained by applying the «compound generating principle» to wider families of densities for texture [13].

We propose to estimate $\theta = (m, \Sigma, \nu)$ by the Maximum Likelihood Estimator (MLE). The main problem is the effective computation of the maximum of the log-likelihood. To solve it, we use the classical Expectation-Maximization (EM) principle to build a sequence of parameters $(\theta_n)_{n \geq 1}$ converging to a stationary point of the log-likelihood. EM has been successfully applied to the T law [28] and more recently to the univariate K law [37], and we refer to these papers for the explicit computations of the E-steps, which are the only difficulty here.

By the way, we give the general algorithm for the computation of the MLE by EM because it can be used in practice for every SIRVs: the E-steps—which is the only challenge for SIRV—can be computed either analytically (as in this paper) or by simulation, *e.g.* Monte Carlo EM (based on empirical estimates of the posterior moments), Stochastic-EM [11] or Bayesian estimators with MCMC sampling methods [36].

Let $\left(y_1^N, u_1^N\right) = \left((y_i)_{1 \leq i \leq N}, (u_i)_{1 \leq i \leq N}\right)$ be N independent realizations of (Y, U) whose density is

$$p(y, u) = g(u) \frac{|\Sigma|^{-1/2}}{(2\pi)^{M/2}} u^{M/2} \exp[-uq(y)].$$

The sample (y_1^N, u_1^N) is called "complete data" in contrast with the observations y_1^N, which are sometimes termed "incomplete data". The log-likelihood of the complete data is $L_c^N(\theta) = \sum_{i=1}^N \log(p(y_i, u_i))$. According to the EM principle, a sequence $(\theta_n)_{n \geq 1}$ is defined by starting from an initial value θ_0 and by computing $\theta_{n+1}(y_1^N) = \arg\max_\theta E[L_c^N(\theta) | y_1^N, \theta_n]$. Given the form of $p(y, u)$, we have

$$E\left[L_c^N(\theta) \middle| y_1^N, \theta_n\right] = -\frac{N}{2} \log((2\pi)^M |\Sigma|)$$
$$+ \sum_{i=1}^N E\left[\frac{M}{2} \log(U_i) - U_i q(y_i) + \log(g(U_i)) \middle| y_1^N, \theta_n\right] \quad (4)$$

which is the general "E step" in SIRVs. Its achievement requires the calculation of the three following conditional expectations: $w_i^n = E[U_i | y_i, \theta_n]$, $E[\log(U_i) | y_i, \theta_n]$ and $E[\log(g(U_i)) | y_i, \theta_n]$. The solution of the "$M$ step" admits a generic form that corresponds to the standard expression of robust M-estimators of location and scatter parameters [27], i.e.,

$$\begin{cases} m^{n+1} = \dfrac{\sum\limits_{i=1}^N w_i^n y_i}{\sum\limits_{i=1}^N w_i^n} \\ \Sigma^{n+1} = \frac{1}{N} \sum_{i=1}^N w_i^n \left(y_i - m^{n+1}\right) \left(y_i - m^{n+1}\right)' \end{cases} \quad (5)$$

with $w_i^n = E[U_i | y_i, \theta_n]$.

The posterior expectation necessitates the closed-form expression of the integrals $\int_0^\infty up(y, u)du$ (divided by $p(y)$). These expressions are directly derived from classical tables of integrals transforms [2]. Indeed, for the T-distribution $up(y, u) \propto u^{(M+\nu)/2} \exp[-u(2q(y) + \nu)/2]$, and for the K-distribution, $up(y, u) \propto u^{(M+\nu)/2} \exp[-uq(y) - \nu/2u]$. It finally gives the following weights:

$$\begin{cases} T : w_i^n = \frac{M+\nu^n}{2q_i^n+\nu^n} \\ K : w_i^n = \frac{\left[K_{(M-\nu^n)/2+1}(\sqrt{2\nu^n q_i^n})\right]}{\left[\sqrt{\frac{2q_i^n}{\nu^n}} K_{(M-\nu^n)/2}(\sqrt{2\nu^n q_i^n})\right]} \end{cases} \cdot \quad (6)$$

The tail parameter ν is found by solving the maximization problem $\arg\max_\theta \sum_{i=1}^N E[\log(g(U_i)) | y_1^N, \theta_n]$, which is done by solving $\frac{d}{d\nu} E[\log(g(U)) | y_1^N, \theta] = 0$. If the derivative of $\log(g(u))$ w.r.t to ν is dominated with respect to the conditional distribution $p(u | y_1^N, \theta)$ (which is the case of the gamma and gamma inverse distribution), the Lebesgue's

dominated convergence theorem enables to permute integration and differentiation so that we have only to compute $(1/2)E[\log(U_i) - U_i | y_i, \theta_n]$ when g is G distribution, and $-(1/2)E[\log(U_i) + 1/U_i | y_i, \theta_n]$ when g is inverse G distribution. The updating formula are

$$\begin{cases} T : \phi\left(\frac{\nu^{n+1}}{2}\right) = \phi\left(\frac{\nu^n}{2}\right) + 1 + \frac{1}{N} \sum_{i=1}^{N} \left(\log(w_i^n) - w_i^n\right) \\ K : \phi\left(\frac{\nu^{n+1}}{2}\right) = 1 - \frac{1}{N} \sum_{i=1}^{N} E[\frac{1}{U_i} + \log(U_i) | y_i, \theta_n] \end{cases}$$ (7)

where $\phi(x) = \log(x) - \Gamma'(x)/\Gamma(x)$. Indeed in both cases, the logarithm of the normalizing constant of the density g equals $(\nu/2)\log(\nu/2) - \log\Gamma(\nu/2)$ that we have to differentiate also w.r.t ν. The only difficulty is the computation of the right-hand side expression of (7) for the K-distribution, which is done in [37]. Finally, we have

$$\begin{cases} E[\log(U_i) | y_i, \theta_n] = -\log(\sqrt{\frac{2q_i^n}{\nu^n}}) + \frac{d}{d\alpha}\log\left(K_\alpha\sqrt{2\nu^n q_i^n}\right) \\ E\left[\frac{1}{U_i}\Big| y_i, \theta_n\right] = \frac{\left[\sqrt{\frac{2q_i^n}{\nu^n}}K_{(M-\nu^n)/2 - 1}(\sqrt{2\nu^n q_i^n})\right]}{\left[K_{(M-\nu^n)/2}(\sqrt{2\nu^n q_i^n})\right]}. \end{cases}$$ (8)

B. Copulas

Let us consider a random vector $Y = (Y_1, \ldots, Y_M)$ taking its values in R^M. The distribution of Y defines the distributions of the components $Y_1, \ldots, Y_M$ and the converse is false in general. However, there are some situations in which the Marginal distributions of Y do determine its distribution. The most known case, which is out of interest in this paper, is the case of independent components. More surprisingly, Rangaswamy and *al.* have shown that the law of one marginal is sufficient to determine the whole distribution of a SIRV [35]. In spite of the interest of SIRVs described above, such a property appears as a rather strong limitation. In fact, we have seen that SIRV are of interest because they can be viewed as a generalization of Gaussian vectors that allows heavier tails. Copulas, which we describe below, will permit to build more general models that can have different marginal laws. In fact, we will see that copulas allow one to dissociate marginal distributions modeling from dependence modeling, and therefore make possible building unusual and varied multivariate densities.

1) Generality on Copulas: Let $Y_1, \ldots, Y_M$ be M real random variables, $F_1, \ldots, F_M$ the cumulative distribution functions (cdf) of their laws, and F the cdf of the law of $Y = (Y_1, \ldots, Y_M)$. If $F_1, \ldots, F_M$ are continuous, which will be assumed in this paper, then according to Sklar theorem there exists an unique function $C : [0, 1]^M \to [0, 1]$, called "copula," such that [31]:

$$\begin{aligned} &\forall (y_1, \ldots, y_M) \in R^M, \\ &F(y_1, \ldots, y_M) = C(F_1(y_1), \ldots, F_M(y_M)). \end{aligned}$$ (9)

An important property is that for a random vector $Y = (Y_1, \ldots, Y_M)$ the associated copula models the dependence of its components in an intrinsic way, independently of the marginal distributions of these components. More precisely, the random vector $Y^* = (Y_1^*, \ldots, Y_M^*)$ defined by $Y_1^* = \varphi_1(Y_1), \ldots, Y_M^* = \varphi_M(Y_M)$, with any $\varphi_1, \ldots, \varphi_M$ continuous strictly increasing functions from R to R, has the same copula as Y. In particular, we can use the vector Y to obtain a vector Y^* whose components are any desired marginal distributions, and whose copula is the same that the copula defined by Y. More precisely, let $F_1^*, \ldots, F_M^*$ be the M desired cdf. Then the vector $Y^* = (Y_1^*, \ldots, Y_M^*)$ defined with $\varphi_1 = (F_1^*)^{-1} \circ F_1, \ldots, \varphi_M = (F_M^*)^{-1} \circ F_M$ has the same copula C as Y, and has the desired $F_1^*, \ldots, F_M^*$ as marginal cdf. Otherwise, we can see that taking for $F_1^*, \ldots, F_M^*$ the identity from $[0, 1]$ to $[0, 1]$, the copula defined by Y is the cdf of the vector $Y^* = (Y_1^*, \ldots, Y_M^*)$. Therefore, a copula can also be seen as a cdf on the hypercube $[0, 1]^M$ with uniform marginal distributions. Finally, we can summarize the properties of copulas useful for this paper as follows:

(i) a copula $C : [0, 1]^M \to [0, 1]$ is a cdf with uniform marginal distributions;

(ii) each cdf F on R^M is given by a copula and M cdfs $F_1, \ldots, F_M$ on R with (9);

(iii) considering (9), one can either make the distributions $F_1, \ldots, F_M,.$ vary and keep the same C, or make the copula C vary and keep the same $F_1, \ldots, F_M$. In the first case, it is possible to obtain any desired marginal distributions for a random vector with correlated components. In the second case, it is possible to obtain a wide family of different random vectors with correlated components having the same marginal distributions.

To illustrate the flexibility offered by copulas, let us consider the models of the previous section: Gaussian, Student, and K model. According to (9), we have three different copulas, denoted respectively by C_G, C_T, and C_K. According to (iii), each of these copulas can be used with any other marginal distributions. For example, we can take C_G with Gaussian margins—which gives the classical Gaussian distribution—, with Student margins, or with K margins. The same can be made for C_T and C_K, which provides, assuming that for a given distribution all components are of a same nature, nine different models. If we assume that for a given model the components can be of different nature, we obtain 3×3^M different models. Let us notice that the last hypothesis is not purely an academic one and it can occur in real situations, especially when the multivariate data are provided by sensors of different nature.

Copulas are generally introduced by (9), but in practice it is more useful to deal with the densities. When the different cdfs considered above are differentiable, we can use the densities $p(y_1, \ldots, y_M) = \partial^M/\partial u_1 \ldots \partial u_M F(y_1, \ldots, y_M)$ with respect to the Lebesgue measure on R and R^M, and (9) can be rewritten by introducing $c(u_1, \ldots, u_M) = \partial^M/\partial u_1 \ldots \partial u_M C(u_1, \ldots, u_M)$:

$$p(y_1, \ldots, y_M) = \left[\prod_{i=1}^{M} f_i(y_i)\right] c[F_1(y_1), \ldots, F_M(y_M)].$$ (10)

Conditional densities are then written

$$p(y_1 | y_2 \ldots, y_M) = f_1(y_1) \frac{c[F_1(y_1), \ldots, F_M(y_M)]}{c_{2,M}[F_2(y_2), \ldots, F_M(y_M)]}$$ (11)
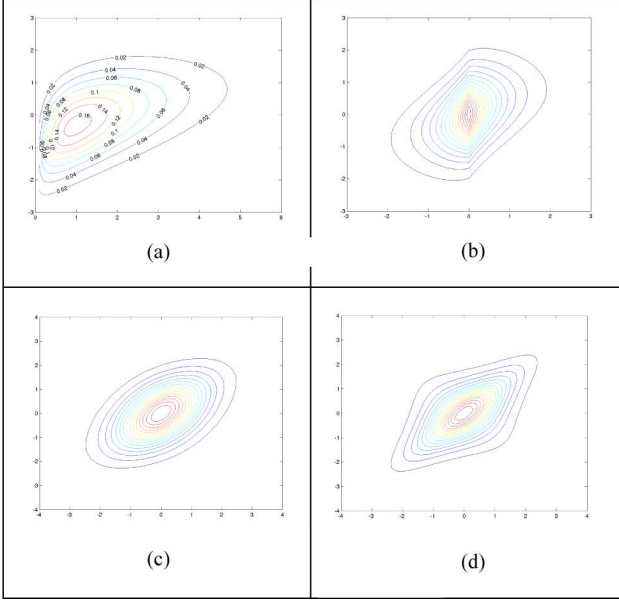
Fig. 2. Bivariate densities with various copulas and various marginal distributions. (a) Gamma and Normal margins (b) K margins with different tail parameters (c) Student margins with same tail parameters (d) Student margins with different tail parameters.

where $c_{2,M}$ is the density of the sub-copula $C_{2,M}$ obtained from C by $C_{2,M}(u_2,\ldots,u_M) = C(1,u_2,\ldots,u_M)$ ($C_{2,M}$ is also the copula of the vector $(Y_2,\ldots,Y_M)$).

2) Gaussian and Student Copulas: Numerous families of parametric copulas enable one to explore different kinds of stochastic dependence; one can find in [31] a very complete overview. In this paper, we deal only with the Gaussian and Student copulas whose densities are easy to compute and enable us to recover known models. Nevertheless, we can define densities with very different shapes, with particular symmetry or completely asymmetric. Some bivariate densities exemplifying the versatility of copulas are presented in Fig. 2.

We invert (10) to compute the density of Gaussian copula, which gives

$$c_{G,\rho}(u_1,\ldots,u_M) = |\rho|^{-1/2} \exp\left[-\frac{1}{2}\zeta'(\rho^{-1} - I_M)\zeta\right] \quad (12)$$

where ρ is a correlation matrix, $\zeta = (\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_M))$, and Φ is the cdf of a centered and standardized Gaussian distribution on R. Similarly, we obtain the density of the Student copula

$$c_T(u_1,\ldots,u_M) = |\rho|^{-1/2}$$
$$\times \frac{\Gamma\left(\frac{\nu+M}{2}\right)\Gamma\left(\frac{\nu}{2}\right)^{M-1}\left(1 + \frac{1}{\nu}\zeta'\rho^{-1}\zeta\right)^{-(\nu+M)/2}}{\Gamma\left(\frac{\nu+1}{2}\right)^M \prod_{i=1}^M \left(1 + \frac{\zeta_i^2}{\nu}\right)^{-(\nu+1)/2}}. \quad (13)$$

Here $\zeta = (t_\nu^{-1}(u_1),\ldots,t_\nu^{-1}(u_M))$, where t_ν^{-1} is the cdf of a centered and standardized univariate Student law, with ν degrees of freedom.

Let us note that the matrix ρ involved in these copulas is no more the correlation matrix of Y but it is the correlation matrix of the transformed random vector ζ.

3) Parameter Estimation: Let $Y = (Y_1,\ldots,Y_M)$ be a random vector in R^M with cdf F and marginal cdfs $F_1(.\,|\theta_1),\ldots,F_M(.\,|\theta_M)$ depending on parameters $\theta_1,\ldots,\theta_M$, and let $f_1(.\,|\theta_1),\ldots,f_M(.\,|\theta_M)$ be the corresponding densities. Let C_η be the copula of F indexed by a parameter η, and c_η its density. The problem we address is the estimation of $\theta = (\theta_1,\ldots,\theta_M)$ and η from an independent sample $y_N = (y_i)_{1\le i\le N}$, with $y_i = (y_i^1,\ldots,y_i^M)$. If one wants to use the MLE, the log-likelihood to be maximized is

$$L^N(\theta,\eta) = \sum_{i=1}^N \sum_{j=1}^M \log\left[f_j\left(y_i^j\,\Big|\,\theta_j\right)\right]$$
$$+ \sum_{i=1}^N \log\left[c_\eta\left[F_1\left(y_i^1\,|\,\theta_1\right),\ldots,F_M\left(y_i^M,|\,\theta_M\right)\right]\right]. \quad (14)$$

The search of the global maximum of L is difficult in general, since we do not have closed form solutions. We propose to use instead the *Inference Functions for Margins* method (IFM, described by Joe in [21]) whose idea is to perform two (easy) maximizations instead of a single difficult one. One first maximizes the first term of the right-hand side in (14), which gives $\hat{\theta} = (\hat{\theta}_1,\ldots,\hat{\theta}_M)$: each $\hat{\theta}_k$ is then the MLE of θ_k based on the data $(y_1^k,\ldots,y_N^k)$. Then we search η that maximizes $L^N(\hat{\theta},\eta)$, which defines an estimator $\hat{\eta}$. Under the classical regularity conditions for the consistency of the MLE, this procedure furnishes consistent estimators of the parameters (θ,η); hence, it suffices to know how to estimate copulas with i.i.d samples in the hypercube $[0,1]^M$ to estimate a copula-based densities.

In order to obtain the MLE, we differentiate the function L^M which give the normal equations to solve, and we give in this paper the corresponding solutions for the copulas C_G and C_T.

In the case of the Gaussian copula, the normal equation derived from (12) gives a closed form estimator for the matrix ρ:

$$\hat{\rho} = \frac{1}{N}\sum_{i=1}^N Z_i Z_i' \quad (15)$$

with $Z_i = (\Phi^{-1}(u_i^1),\ldots,\Phi^{-1}(u_i^M))$. So the ML estimator is the covariance matrix of the "Gaussianized" sample $(Z_i)_{1\le i\le N}$

Considering the Student copula C_T, let us first assume that the tail parameter ν of C_T is known. The likelihood (13) in ρ of the Student copula boils down to the likelihood of the random vector $(Z_1,\ldots,Z_M) = (t_\nu^{-1}(F_1(Y_1|\theta_1)),\ldots,t_\nu^{-1}(F_M(Y_M|\theta_M))$ computed with the Student density (2) and parameters $0,\rho,\nu$. In that case, we have only to estimate the scatter parameter ρ which has only 1s on the diagonal (because it is a correlation matrix). Hence, the MLE of the copula is simply the MLE of the T distribution defined by

$$\hat{\rho} = \arg\min_\rho \sum_{i=1}^N (\nu + M)\log\left(\frac{\nu + z_i'\rho^{-1}z_i}{\nu}\right) + N\log|\rho|. \quad (16)$$

This MLE is a consistent estimator for elliptic distributions under broad conditions, as it has been shown by Maronna [27] (since the transformed vector $(Z_1,\ldots,Z_M)$ is elliptic, it furnishes also a consistent estimator for the copula). The MLE (16)

can be computed by remarking that the first order condition implies that $\hat{\rho}$ satisfies the following implicit equation:

$$\hat{\rho} = \frac{1}{N} \sum_{i=1}^{N} \frac{\nu + M}{\nu + z_i' \hat{\rho}^{-1} z_i} z_i z_i' \qquad (17)$$

$\hat{\rho}$ is then a fixed point of $f : \rho \rightarrow 1/N \sum_{i=1}^{N} (\nu + M) z_i z_i' / (\nu + z_i' \rho^{-1} z_i)$, which can be approximated by computing the sequence of iterates $\forall t > 0, \rho_{t+1} = f(\rho_t)$. For all $\nu > 0$ and $N \geq M$, Kent and Tyler [24] have shown that $\hat{\rho}$ in (17) exists and is unique and that the iterative search reaches the minimum. An initial value providing fast convergence is then the Gaussian estimator $\rho_0 = 1/N \sum_{i=1}^{N} \zeta_i \zeta_i'$. Nevertheless, the obtained matrix is not necessarily a correlation matrix, and we suggest to normalize the matrices ρ_t by the following transformation $\forall i, j, \rho_{ij} = \rho_{ij} / \sqrt{\rho_{ii} \rho_{jj}}$. This slight modification of the fixed point algorithm does not modify the convergence, and it still provides a correct estimate, as we have empirically observed.

When the parameter ν is unknown, the iterative computation cannot be used and we must perform a multidimensional search of the solution (ν, ρ). We avoid the global optimization by splitting again the computation of the different estimators. We suggest using a moment estimator of ρ and then doing a numerical optimization for ν. Indeed, there is a relationship between the Kendall's Tau [13] (which is a rank statistic) and the matrix ρ, i.e., $\rho_{ij} = \sin(\pi \tau_{ij}/2)$ for $1 \leq i, j \leq M$ where τ_{ij} stands for the Kendall's Tau between Y_i and Y_j.

III. UNSUPERVISED CLASSIFICATION USING MULTIVARIATE HIDDEN MARKOV CHAINS

The multivariate modeling (SIRV and copulas) considered in the previous section is introduced in order to propose suitable multivariate statistical models for Bayesian classification of multidimensional data. More precisely, we focus on classical hidden Markov chains (HMC) which enable to deal with dependent data encountered in signal and image processing problems, but the main point is the modeling of the conditional distributions of the observations and the ability to estimate mixtures of multivariate densities for unsupervised classification. Hence, we show how the different parameter estimation methods described above can be extended to HMC.

A. Classification With Multivariate Hidden Markov Chains

For notational convenience and to make the paper self-contained, let us first briefly recall main principles of Bayesian classification with multivariate hidden Markov chains (MHMC, also see [9], [34]). Let $X = (X_1, \ldots, X_N)$ be a Markov chain, each X_i taking its values in a finite set of classes $\Omega = \{1, \ldots, K\}$, and let $Y = (Y_1, \ldots, Y_N)$ be a random multivariate process, each $Y_i = (Y_i^1, \ldots, Y_i^M)$ taking its values in R^M. The pair (X, Y) is a called multivariate HMC (MHMC) if its joint distribution is

$$p(x, y) = p(x_1) \prod_{i=2}^{N} p(x_i | x_{i-1}) \prod_{i=1}^{N} p(y_i | x_i). \qquad (18)$$

The classification problem is thus the estimation of the unobserved process $X = x$ from the observations $Y = y$. The MHMC model is often used because we can easily compute the Bayesian estimators Maximum A Posteriori (MAP) and Maximum Posterior Mode (MPM). In this paper we will focus on MPM, denoted by $\hat{x}_{MPM} = (\hat{x}_{1,MPM}, \ldots, \hat{x}_{N,MPM})$. The MPM, which is optimal in that it minimizes the mean ratio of wrongly classified points, is defined by

$$\hat{x}_{i,MPM} = \arg \max_{1 \leq k \leq K} p(x_i = k | y). \qquad (19)$$

The computation of the posterior marginal distributions is feasible thanks to the "forward-backward" method, which is a fast and exact algorithm allowing one to calculate, among others, the marginal posterior distributions used in (19).

B. Parameter Estimation and Unsupervised Classification

In practical cases, we need to use a parametric modeling of the distributions involved in (18) and to estimate the parameters by using only the observations $Y = y$. We assume then that the multivariate HMC considered is stationary, so that neither $p(x_i, x_{i+1})$ nor $p(y_i | x_i)$ depend on $1 \leq i \leq N$. Moreover, we assume that $p(y_1 | x_1 = j)$ belongs to a parametric family of densities $P(\Theta_j)$ indexed by a parameter $\theta_j \in \Theta_j$. We have to estimate the global parameter ϕ formed by the joint probability matrix $p(x_1, x_2)$, noted $P = (p_{ij})_{1 \leq j, k \leq K}$, and $\theta_1, \ldots, \theta_K$. As in the independent case considered in the previous section, we propose to use the general EM method. When the laws are assumed to be Gaussian or to belong to the exponential family, the EM principle gives simple algorithms for the approximation of the MLE of a finite mixture of densities. Despite some well-known drawbacks like sensibility to initial conditions and low speed of convergence, the obtained EM sequences of parameters present, in general, satisfying qualities. For estimation of a mixture of multivariate laws considered in this paper, the EM principle usually stumbles across the maximization step so that we propose some adaptations of the classical EM for finite mixture SIRV and copula-based models. Let us notice that the methods below can be seen as extensions of the methods discussed in the previous section to the multivariate HMC.

1) SIRV Models With K and T Laws: We present in this subsection the EM formulas in the case of MHMC defined by (18), where $p(y_i | x_i)$ are either K or T distributions. In both cases, it is useful to introduce the "texture" random chain involved in the definition of SIRV $U = (U_1, \ldots, U_N)$ such that $p(y_i | x_i)$ is the marginal distribution of $p(y_i, u_i | x_i)$. Thus let us consider the complete process $T = (X_i, U_i, Y_i)_{1 \leq i \leq N}$, with the law given by $p(x, u, y) = p(x_1) \prod_{i=2}^{N} p(x_i | x_{i-1}) \prod_{i=1}^{N} p(u_i, y_i | x_i)$. We can notice that we have a "double" hidden chain (X, U), which is such that the r. v. $U_1, \ldots, U_N$ are independent conditionally on (X, U). Otherwise, the distribution of Y conditional on (X, U) is the very classical independent Gaussian distribution.

Finally, the problem is to estimate from $Y = y$ the parameters $P = (p_{ij})_{1 \leq j, k \leq K}$, and $\theta_1, \ldots, \theta_K$, where $\theta_i = (m_i, \Sigma_i, \nu_i)$

After having chosen an initial value $\phi^0 = (P^0, \theta_1^0, \ldots, \theta_K^0)$, we derive from the EM principle an iterative algorithm, where $\phi^{n+1} = (P^{n+1}, \theta_1^{n+1}, \ldots, \theta_K^{n+1})$ is obtained from $Y = y$,

$\phi^n = (P^n, \theta_1^n, \ldots, \theta_K^n)$ and the posterior probabilities $\pi_i^{k,n} = p(x_i = k \,|y, \phi^n)$ and $\pi_i^{jk,n} = p(x_i = j, x_{i+1} = k \,|y, \phi^n)$ in the following manner:

$$
\begin{cases}
p_{jk}^{n+1} = \frac{1}{N} \sum_{i=1}^{N} \pi_i^{jk,n} \\[2mm]
m_k^{n+1} = \dfrac{\sum_{i=1}^{N} w_i^{k,n} \pi_i^{k,n} y_i}{\sum_{i=1}^{N} w_i^{k,n} \pi_i^{k,n}} \\[4mm]
\Sigma_k^{n+1} = \frac{1}{N} \sum_{i=1}^{N} w_i^{k,n} \pi_i^{k,n} (y_i - m_k^{n+1})(y_i - m_k^{n+1})'
\end{cases}
\tag{20}
$$

The updating formulas for the location and dispersion parameters are similar to those of (5), which are adapted to the mixture context with $w_i^{jk,n} = E[U_i \,|y_i, x_i = k, \phi^n]$ and the posterior probabilities $\pi_i^{k,n}$. In the same fashion, the equation for the tail parameters $(\nu_k)_{1 \le k \le K}$ is the adaptation of (7) by using the posterior expectations $w_i^{jk,n} = E[(1/U_i) \,|y_i, x_i = k, \phi^n]$, $w_i^{jk,n} = E[\log(U_i) \,|y_i, x_i = k, \phi^n]$ and the posterior probabilities $\pi_i^{k,n}$

$$
\begin{cases}
T : \phi\left(\frac{\nu_k^{n+1}}{2}\right) = \\[2mm]
\phi(\frac{\nu_k^n}{2}) + 1 + \frac{1}{\frac{1}{N} \sum_i \pi_i^{k,n}} \sum_{i=1}^{N} \pi_i^{k,n} \left(\log\left(w_i^{k,n}\right) - w_i^{k,n}\right) \\[4mm]
K : \phi(\frac{\nu_k^{n+1}}{2}) = \\[2mm]
1 - \frac{1}{\frac{1}{N} \sum_i \pi_i^{k,n}} \sum_{i=1}^{N} \pi_i^{k,n} E\left[\frac{1}{U_i} + \log(U_i)\,\Big|\, y_i, x_i = k, \theta_n\right].
\end{cases}
\tag{21}
$$

The difference between the mixture and the single class estimation procedures is that in the latter case, we did not need the expression of the density in (1) unlike in the mixture case, where we need it for the computation of the posterior probabilities $\pi_i^{k,n}$. This can be a serious limitation to the use of general SIRV models in the classification setting; however, we can calculate the needed density for the T and K law which are among the most used models.

2) Copula-based Models: We will use the following notations. For the K classes, we have for each $k = 1, \ldots, K$ a copula $C(u_1, \ldots u_M \,|\eta_k)$ and M marginal cdfs $F(.\,|\theta_{1,k}), \ldots, F(.\,|\theta_{M,k})$, whose densities will be denoted by $F(.\,|\theta_{1,k}), \ldots, F(.\,|\theta_{M,k})$. Thus the parameters to be estimated are $\phi = ((\theta_{i,k})_{1 \le i \le M}, \eta_k)_{1 \le k \le K}$. Let us first consider the complete data (X, Y). According to (18) and (14), the log-likelihood can be written as

$$
L_c^N(\phi) = \log(p(x)) + \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{M} \log\left[f\left(y_i^j \,\Big|\, \theta_{i,k}\right)\right] 1_{\{x_i=k\}} +
$$
$$
\sum_{k=1}^{K} \sum_{i=1}^{N} \log\left[c\left[F\left(y_i^1|\theta_{1,k}\right), \ldots, F\left(y_i^M \,|\, \theta_{M,k}\right) \,|\, \eta_k\right]\right] 1_{\{x_i=k\}}.
\tag{22}
$$

The E-step consists in replacing the function $1_{\{x_i=k\}}$ by the posterior probability $\pi_i^{k,n}$. Hence, as in Section II, we have to come up against a difficult maximumization program on $\phi = ((\theta_{i,k})_{1 \le i \le M}, \eta_k)_{1 \le k \le K}$, and we propose then the following two step estimation at each M-step:

1) Compute the maximization step for the M mixtures of marginals, so that we have $\hat{\theta}_{i,k}$ [i.e., maximize the second term of the right-hand formula of (22)];
2) Plug the estimators into $\sum_{k=1}^{K} \sum_{i=1}^{N} \log\left[c\left[F\left(y_i^1|\theta_{1,k}\right), \ldots, F\left(y_i^M|\theta_{M,k}\right) \,|\, \eta_k\right]\right] \pi_i^{k,n}$ and maximize on $(\eta_k)_{1 \le k \le K}$.

The estimation of the transition parameter of the Markov chain X is the same as in the previous subsection. The parameters are updated until the sequence stabilizes around a particular value occurs.

The motivation of this algorithm is that we can estimate independently the parameters of the M univariate HMC $(X, Y^j)_{1 \le j \le M}$ indexed with parameters $(P, (\theta_{i,k})_{1 \le k \le K})$ and then plug these values in the log-likelihood $L_c^N(\phi)$ for computing the dependence parameters of the multivariate HMC. This is the application of the Inference Function for Margins method to the vector $Y = (Y_1, \ldots, Y_N)$. Indeed, we estimate first the marginal distribution of Y (which is the mixture $\sum_{k=1}^{K} \pi_k f(y, (\theta_{i,k})_{1 \le i \le M})$, and plug the estimated parameters in the complete log-likelihood. Hence, the algorithm links the $M + 1$ models in order to have the same estimation for the parameter P. In particular, the interest of this approach is that it enables to use already known maximization procedures for the marginal laws and also known estimators for the copulas, as the ones given in Section II-B.

IV. EXPERIMENTS

The first subsection is devoted to the unsupervised classification in five different test beds. The models used are SIRVs, and the parameters are estimated with EM. In the second subsection, we compare the estimates of SIRV parameters when they are computed by IFM in the context of the new copula based multivariate HMC.

A. Estimation and Classification of HMC With EM

Let us consider three SIRV models, denoted by G (for Gaussian), T (for Student) and K (for K law). Otherwise, let us consider a finite Markov chain $X = (X_1, \ldots, X_N)$ with three classes $K = 3$), a transition matrix $A = (a_{ij})_{1 \le i,j \le K}$, with $a_{ii} = 0.8$ and $a_{ij} = 0.1$ for $i \ne j$. Finally, a HMC is defined by three $p(y_n \,|x_n)$, $x_n = 1, \ldots, x_n = 3$ distributions on R^2. The location (mean) parameters of these three distributions are $m_1 = (0,0)'$, $m_2 = (1.5, 1.5)'$, $m_3 = (3,3)'$ and the scatter matrices are $\Sigma_k = \begin{bmatrix} 1 & \rho_k \\ \rho_k & 1 \end{bmatrix}$, with $\rho_1 = 0.4$, $\rho_2 = 0.2$ and $\rho_3 = 0.5$. All these parameters being fixed, we then study five kinds of models with various tail parameters: a Gaussian model (noted Γ), a Student model with identical tail parameters $\nu_1 = \nu_2 = \nu_3 = 10$ (noted T_1), an another Student model with different tail parameters $\nu_1 = 5$, $\nu_2 = 10$, $\nu_3 = 15$ (noted T_2), a K model with identical tail parameters $\nu_1 = \nu_2 = \nu_3 = 10$ (noted K_1), and an another K model with different tail parameters $\nu_1 = 2$, $\nu_2 = 4$, $\nu_3 = 8$ (noted K_2). The theoretical error rates of the five models are evaluated by Monte-Carlo and presented in the first line of Table I. Despite of different tail parameters, we can observe that the models T_1 and T_2, as well as the models K_1 and K_2 give nearly the same
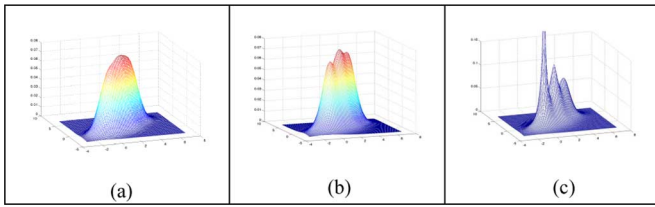
Fig. 3. Three examples of SIRV models (a) mixture Γ. (b) mixture T_1. (c) mixture K_2.

error rate, so that the difference is then mainly due to the shape of the densities, as illustrated in Fig. 3.

We simulate data with models $\Gamma, T_1, T_2, K_1, K_2$ and the error rates are presented columnwise in Table I. We show also in Fig. 3 the densities Γ, T_1, K_1. The parameters are estimated with EM applied to G, T, and K models, and the data are classified according to the Bayesian MPM method. The unsupervised error rates are presented in lines 3, 4, and 4 of Table I. The EM algorithm uses 100 iterations, which was sufficient to obtain convergence.

We can observe that for each model the worst case is that of Student laws and that the easiest is that of K laws. Moreover, we obtain the best performance in unsupervised classification with the true model (except for T_1 but the difference is small), but we have a difference between the models with equal or unequal tail parameters in the case of the K law. While there is no difference for T_1 and T_2 (except when we use the model G), we have a noticeable difference for the K distribution, and it appears that different tail parameters (or different textures in radar language) help for retrieving the different classes. We can also remark that the K distribution seems to be well-adapted for the classification of G and T models since we have the best performance with the K model in nearly each configuration.

In order to evaluate the robustness of the estimation and classification procedures, we have performed the same comparison of the 3 models on mixtures of Cauchy distributions (i.e., mixture of Student with $\nu_1 = \nu_2 = \nu_3 = 2$). This is the most difficult case since the supervised error rate equals 20%, nevertheless the T and K distributions behave analogously to the situations described in Table I. Indeed, we report unsupervised error rates equal to 21,5% for the T-model and 22% for the K-model. As one could expect, the Gaussian model gives very poor results since two thirds of the experiments gave an error rate higher than 40% (and one quarter have an error rate higher than 50%).

B. Estimation and Classification of HMC With Copulas and IFM

The aim of this subsection is to compare the EM and IFM efficiencies in the case of Γ and T_1 models. For Γ, we interpret a Gaussian distribution as a density obtained with a Gaussian copula and Gaussian marginals, so that the copula-based model is exactly the parametric model G. In the case of the mixture T_1, instead of considering the model T, we use a generalized Student model obtained with a copula C_T and Student marginals, so that each marginal can have different tail parameter. However, we use a simpler model by imposing the same tail parameter ν_m for all the margins. Hence, each component of the mixture is

indexed by one additional parameter ν_C (by comparison with a T law), which is the tail parameter of the Student copula (see (13)). We note this model (C_{T,ν_C}, T_{ν_m}) and when $\nu_C = \nu_m$ for all the components, we obtain the T model, which is then nested in (C_{T,ν_C}, T_{ν_m}); but if the $\nu_C \neq \nu_m$ then the estimated model is out of the SIRVs family, and we are in the same case as the distributions drawn in Fig. 2. In this paper, we fix $\nu_C = 10$ for all the classes, so that the mixture T_1 belongs to $(C_{T,10}, T_{\nu_m})$. We then compare the mean estimation of the parameters and error rates when $N = 500$ observations by Monte-Carlo (with 500 simulations), and by using 300 iterations for the EM and the IFM algorithms. The results are presented in Tables II–V.

To initialize both EM and IFM algorithms we used a preliminary classification based on a HMC with T distributions whose parameters are estimated by EM. This EM algorithm is itself initialized with a classification based on a K-means clustering. When we consider data coming from heavy tailed distribution as Cauchy, it is necessary to use the l_1 norm and not the Euclidean l_2 norm at this stage, or else a bad initialization can make the IFM diverge. In order to reduce the risk of spurious local minima (whose number is increased with heavy-tailed distribution), we use also a multi-start approach (10 different initial conditions), by keeping the parameters with the higher likelihood with class probabilities $P(X_i = k)$, $k = 1, 2, 3$, higher than 5% (see Tables VI–IX).

The conclusion of these experiments is that the efficiencies of EM and IFM are similar. However, we remark that EM better estimates ρ_2, while IFM better estimates the tail parameters. We remark also that since the parameters ν_k are different from the parameter of the copula ν_C, the estimated model tends to differ from the initial mixture of student distribution: we get then generalized elliptic models. Otherwise, we obtain similar results for unsupervised classification. In fact, we have an error rate of 14.2% for the model Γ estimated by EM, and 14.4% when the model is estimated by IFM. In the case of T_1, these error rates are 17.1% and 16.9%, respectively.

These experiments show that the copula-based multivariate models, which are extensions of the classical SIRV models, can be efficiently learnt with the proposed IFM method.

We complete this study with a robustness analysis by considering the extreme case of 3 Cauchy distributions, as in Section IV–A. The conclusion is that EM algorithm is more robust than IFM (the error rate for IFM is 22.5% versus 21.5% for EM) in this particular extreme case. This comes from difficulties in the estimation of the second class for which there is a bias in the position and the scatter parameters. The difference is essentially due to the fact that IFM lies on a model that does not contain the generating process. Indeed, we recall that the copula parameter ν_C is different from the parameters ν_k of the margins. This discrepancy might be amplified when the fluctuations become important, as it is the case with the Cauchy distribution. But for classification and estimation purpose, the difference between the margins and the dependency matrices ρ are still clearly identified.

The proposed algorithm for computing the Inference For Margins estimator is a slight modification of the EM algorithm, and hence it possesses the same local property, i.e., it depends on the choice of the initial value. It is difficult to discuss and

TABLE I
ERROR RATES FOR SUPERVISED AND UNSUPERVISED CLASSIFICATION (%)

	Γ	T_1	K_1	T_2	K_2
True models	12.4	14.5	10.9	14.5	10.9
G	13.4	16	13.1	16.9	13
T	13.4	15.6	13.8	15.7	11.7
K	13.5	15.5	13.1	15.7	11.7

Error rates for supervised and unsupervised classification (%)

TABLE II
ESTIMATIONS OF m_1, m_2, m_3 WITH EM AND IFM FOR Γ
AND $T_1 \cdot C1$ CLASS; TRUE PAR: TRUE PARAMETERS

cl	true par	Γ		T_1	
		EM	IFM	EM	IFM
1	(0, 0)	(-0.01, -0.01)	(0, 0)	(0, -0.01)	(0, 0)
2	(1.5, 1.5)	(1.5, 1.5)	(1.47, 1.47)	(1.49, 1.51)	(1.5, 1.52)
3	(3, 3)	(3.01, 3.05)	(3, 3)	(3, 2.99)	(3, 2.99)

TABLE III
ESTIMATIONS OF THE VARIANCES (DIAGONAL OF Σ_k) WITH EM AND
IFM FOR Γ AND $T_1 \cdot$ C1: CLASS; TRUE PAR: TRUE PARAMETERS

Cl	True par	Γ		T_1	
		EM	IFM	EM	IFM
1	(1, 1)	(0.96, 0.97)	(0.98, 0.96)	(1, 0.96)	(0.99, 0.96)
2	(1, 1)	(0.96, 0.98)	(0.96, 0.98)	(0.97, 0.99)	(0.92, 0.94)
3	(1, 1)	(0.96, 0.96)	(0.99, 0.99)	(1, 1.02)	(0.97, 0.99)

TABLE IV
ESTIMATIONS OF ρ_1, ρ_2, ρ_3 WITH EM AND IFM FOR Γ
AND $T_1 \cdot$ C1: CLASS; TRUE PAR: TRUE PARAMETERS

Cl	True par	Γ		T_1	
		EM	IFM	EM	IFM
1	0.4	0.4	0.39	0.38	0.38
2	0.2	0.2	0.19	0.18	0.17
3	0.5	0.5	0.5	0.49	0.5

TABLE V
ESTIMATIONS OF ν_1, ν_2, ν_3 WITH EM AND IFM FOR Γ
AND $T_1 \cdot$; TRUE PAR: TRUE PARAMETERS

class	true parameters	EM	IFM
1	10	18.4	15.8
2	10	23.9	17.2
3	10	18.3	14.3

TABLE VI
ESTIMATIONS OF ν_1, ν_2, ν_3 WITH EM AND IFM FOR T_1

class	true parameters	EM	IFM
1	(0, 0)	(-0.01, -0.01)	(0.01, -0.01)
2	(1.5, 1.5)	(1.48, 1.5)	(1.28, 1.35)
3	(3, 3)	(3.01, 3.01)	(2.99, 2.97)

TABLE VII
ESTIMATIONS OF THE VARIANCES (DIAGONAL OF Σ_k) WITH EM AND IFM
FOR CAUCHY MIXTURE

class	true parameters	EM	IFM
1	(1, 1)	(0.95, 0.95)	(0.96, 0.96)
2	(1, 1)	(1, 0.99)	(0.92, 0.88)
3	(1, 1)	(0.93, 0.95)	(0.94, 0.95)

TABLE VIII
ESTIMATIONS OF ρ_1, ρ_2, ρ_3 WITH EM AND IFM FOR CAUCHY MIXTURE

Class	True parameters	EM	IFM
1	0.4	0.37	0.37
2	0.2	0.19	0.11
3	0.5	0.45	0.47

TABLE IX
ESTIMATIONS OF ν_1, ν_2, ν_3 WITH EM AND IFM FOR CAUCHY

Class	True parameters	EM	IFM
1	2	1.8	2.
2	2	2.2	7.3
3	2	1.8	1.9

errors as the estimation of the parameters of each marginal does not depend on the models (possibly wrong) used for the other marginals or the joint distribution. Moreover, the robustness can be increased for the estimation of the parameters of the copulas, by using a semi-parametric version of the IFM, called the omnibus estimator which estimates the copula without the explicit use of the parametric assumptions made on the marginals. In the particular robustness analysis we dealt with, we can check that the copula-based model with 3 additional parameters (the tail parameters of the margins) can still be correctly estimated by IFM (and parameters remain close to the asymptotically optimal estimator) despite its increased complexity.

C. Real Data Processing

1) Radar Doppler Segmentation: Motivation and Modeling: In this section we present an application of the non-Gaussian modelling to the segmentation of radar Doppler data. An important aim in radar signal processing is the estimation of the velocity field of the radar environment [10], [26]. Indeed, the difference Δf between frequencies of the transmitted signal and the received signal is given by $\Delta f = 2v/\lambda_0$, where v is the speed of the reflector and λ_0 the wave length of the transmitted wave. The frequencies Δf are usually called "Doppler frequencies." In practice, the signals received are structured on a distance-azimuth map; azimuth representing the angle formed by the radar and the normal direction. For a given distance and azimuth, we observe a complex vector $Y = (Y^1, \ldots, Y^M)$ whose the fast Fourier transform provides us with the distribution of the signal power through the Doppler frequencies, called usually a "Doppler spectrum"

$$\forall \Delta f \in \left[-\frac{M}{2T_{coh}}, \frac{M}{2T_{coh}} \right], \ S(\Delta f)$$

$$= \frac{1}{2\pi M} \left| \sum_{n=1}^{M} Y^n \exp\left(2i\pi n \frac{T_{coh}}{M} \Delta f \right) \right|^2$$

evaluate the robustness of IFM with respect to the modeling errors and the different realizations of the data, because it is related to the finite sample statistical properties of IFM, which are different from the MLE. It has been shown that IFM gives better results than MLE on small samples [16] (obviously MLE tends to be the best estimator as the sample size tends to infinity). Nevertheless, we advocate that there is a lower sensitivity (greater statistical robustness) with respect to modeling

where T_{coh} is the coherent time and corresponds to the time needed to get a suitable Doppler spectrum, in practise, this time is about $T_{coh} = 1$ ms. $S(\Delta f)$ represents the Doppler intensity due to the Doppler frequency Δf; if the reflector has a speed equal to v, $S(\Delta f)$ is proportional to a Dirac distribution centered at $\Delta f = 2v/\lambda_0$. However, a reflector such as a fluid can have several speed components, that corresponds to different peaks in the Doppler Spectrum.

It is to say, in a range-azimuth map, one can be interested in finding and characterizing the spatial changes in the Doppler spectra. A possible way to construct such a map is to perform unsupervised segmentation of the radar returns based on the Doppler (frequency) content of the range-azimuth cells [25].

Let us detail some of the statistical assumptions about the radar signal. It is assumed that each vector $Y = (Y^1, \ldots, Y^M)$ is centered and stationary, so that the Doppler information is contained in the Toeplitz covariance matrix $\Sigma = (\gamma(k))_{0 \le k \le M-1}$ of the sample $Y = (Y^1, \ldots, Y^M)$ where $\gamma(k) = E(Y^n \bar{Y}^{n-k})$ and the spectrum can be estimated by Fast Fourier Transform of this covariance matrix

$$\forall \Delta f \in \left[-\frac{M}{2T_{coh}}, \frac{M}{2T_{coh}}\right], \ S(\Delta f)$$
$$= \frac{1}{2\pi} \sum_{k=0}^{M-1} \gamma(k) \exp\left(2i\pi k \frac{T_{coh}}{M} \Delta f\right).$$

Under the auto-regressive assumption, the Doppler spectrum can be parametrized by a finite number q of reflexion coefficients $\mu^{(n)} = \left(\mu_1^{(n)}, \ldots, \mu_q^{(n)}\right)$ and the back-scattered intensity $\varepsilon_q^{(n)}$, which are easily computed via the Durbin-Watson algorithm, [20]. More precisely, we propose then to construct automatically a map of the velocity field by applying the unsupervised classification algorithm defined in Section III-A to the reflexion coefficients μ instead of to the returns $Y = (Y^1, \ldots, Y^M)$.

Let $p(\mu^{(n)} | x_n)$ be the distribution of the nth cell on C^q (C being the set of complex numbers), to be defined. We can not consider a classical SIRV model because the reflexion coefficients are in the unit disc. Let $z_n = \left(\left|\mu_1^{(n)}\right|, \arg\left(\mu_1^{(n)}\right), \ldots, \left|\mu_q^{(n)}\right|, \arg\left(\mu_q^{(n)}\right)\right)$ be the observation from the distance-azimuth cell n. In the following, we will suppose that $q = 10$. In order to reduce the complexity of $p\left(\mu^{(n)} | x_n\right)$, we will suppose that the angles $\arg(\mu_1^{(n)}), \ldots, \arg(\mu_{10}^{(n)})$ are independent of each other conditionally on the states x_n. Moreover, as the arguments lie on $[0, 2\pi]$ and are likely not to be uniformly distributed, we suppose that the angles have a Von Mises Fisher distribution with direction parameters $\tilde{\mu}_{1,x_n}, \ldots, \tilde{\mu}_{q,x_n}$ and concentration parameters $\kappa_{1,x_n}, \ldots, \kappa_{q,x_n}$ [3], hence $p\left(\arg\left(\mu_1^{(n)}\right), \ldots, \arg\left(\mu_q^{(n)}\right) | x_n\right) = \prod_{k=1}^{q} p\left(\arg\left(\mu_k^{(n)}\right) | x_n\right)$, with

$$p\left(\arg\left(\mu_k^{(n)}\right) | x_n\right) \propto \exp\left(\kappa\left(\cos\left(\arg\left(\mu_k^{(n)}\right)\right)\right.\right.$$
$$\left.\left. \times \text{Re}\left(\tilde{\mu}_{k,x_n}\right) + \sin\left(\arg\left(\mu_k^{(n)}\right)\right) \text{Im}\left(\tilde{\mu}_{k,x_n}\right)\right)\right).$$
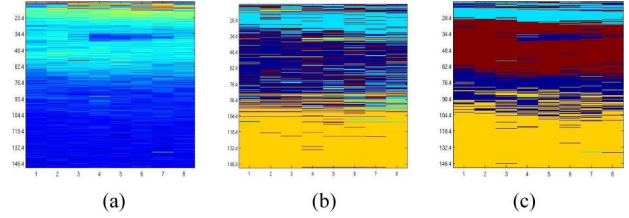

Fig. 4. (a): Range-azimuth map of returned intensities (8 azimuths and 1548 range cells); (b): classification of radar returns using independent modules; (c): classification of radar returns using dependent modules with Gaussian copulas.

Moreover, we suppose that $\left(\left|\mu_1^{(n)}\right|, \ldots, \left|\mu_q^{(n)}\right|\right)$ and $\left(\arg\left(\mu_1^{(n)}\right), \ldots, \arg\left(\mu_q^{(n)}\right)\right)$ are independent conditionally on the states x_n so that $p(z_n | x_n) = p\left(\left|\mu_1^{(n)}\right|, \ldots, \left|\mu_q^{(n)}\right| | x_n\right) \prod_{k=1}^{q} p\left(\arg\left(\mu_k^{(n)}\right) | x_n\right)$.

Finally, as $\left|\mu_k^{(n)}\right| \in [0,1]$ (see [4] for details) and are not necessarily uniformly distributed, we will suppose that $p\left(\left|\mu_1^{(n)}\right|, \ldots, \left|\mu_q^{(n)}\right| | x_n\right)$ has beta distributions for margins, i.e., $p\left(\left|\mu_k^{(n)}\right| | x_n\right) \propto \left|\mu_k^{(n)}\right|^{a_{x_n}-1} \left(1 - \left|\mu_k^{(n)}\right|\right)^{b_{x_n}-1}$, and we will consider two dependence structures between the modules:

Model 1 (independence)

$$p\left(\left|\mu_1^{(n)}\right|, \ldots, \left|\mu_{10}^{(n)}\right| | x_n\right) = \prod_{k=1}^{10} p\left(\left|\mu_k^{(n)}\right| | x_n\right). \quad (23)$$

Model 2 (dependence with Gaussian copula)

$$p\left(\left|\mu_1^{(n)}\right|, \ldots, \left|\mu_{10}^{(n)}\right| | x_n\right) = \prod_{k=1}^{10} p\left(\left|\mu_k^{(n)}\right| | x_n\right)$$
$$\times c\left(F_1^{x_n}\left(\left|\mu_1^{(n)}\right|\right), \ldots, F_{10}^{x_n}\left(\left|\mu_{10}^{(n)}\right|\right)\right) \quad (24)$$

where $p(\left|\mu_k^{(n)}\right| | x_n)$ is the beta distribution with f.d.r. $F_k^{x_n}$ and c is the Gaussian copula.

2) Data and Classification of Radar Returns: We use data kindly provided by THALES coming from measurement made with a coastal X-band radar whose transmitted frequency is $f0 = 10$ Ghz. The radar resolution is 300 meters (size of a cell), and we are provided with $M = 16$ pulses. The data environment is divided into 1548 distance cells and 8 azimuth cells which gives us a total of 8×1548 cells. During the measurement campaign, weather conditions were rainy and so we attempted to recognize in the radar environment four classes: a class corresponding to the dominancy of sea return, a class to the dominancy of rain return, another to thermal noise and the last probably being the blind area of the radar (close to the radar). In the Fig. 4(a), we have represented as a color graduation the return's intensities $\varepsilon_q^{(n)}$, even if we don't use these data as observations in the hidden Markov model, they let us visualize the nature of the radar signal in terms of intensity. The hidden Markov model was detailed in the previous paragraph with $q = 10$. In the Fig. 4(b), the segmentation using independent copulas is represented and the segmentation using Gaussian copulas is represented in the Fig. 4(c).

From Fig. 4, we can see that the Gaussian copula gives more homogeneous classes than the independent copula. So in guise of conclusion from these results, it can be important to notice that the radar return are widely likely to be spatially correlated. It can be interesting, in the future, to compare different copulas based models and select the most suitable dependence model in the context of this experiment. Regarding the different classes, in Fig. 4.(c), the light blue class seems to correspond to the blind area of the radar, the yellow one to thermal noise (targets being negligible compared to noise), the dark blue class to a dominancy of sea returns and the brown class to rain speckle.

V. CONCLUSION

In this paper we have introduced a general model of multivariate density and we have applied it in the setting of classification, in particular with hidden Markov chains. We also introduced a related parameter estimation method as well so that unsupervised classification is allowed with this model.

The model is based on copulas and multivariate "hidden Markov chains" (HMC), which enables one to consider correlated sensors and noise margins of any form. In particular, the model contains extensions of the classical "spherically invariant random vectors" (SIRV) models, such as asymmetric T and K laws, reusing the expression of Fang *et al.* in [15]. Hence, the model introduced should be of particular interest for the classification of radar signals, but the model is quite versatile, thanks to the generality and flexibility of copulas. Otherwise, the model proposed can also be seen as an extension of the models introduced in [19], [22], where the sensors are assumed to be independent.

The proposed parameter estimation method is an iterative algorithm based on the "inference functions for margins" (IFM) principle for the estimation of hidden Markov models. It does not correspond to the classical maximization of the likelihood, and it appears as a slight modification of the classical EM algorithm [23], [28]. In particular, despite a common feature with the general ECM algorithm proposed in [30] which consists in a sequential estimation of the parameters, the IFM algorithm does not perform a systematic increase of the log-likelihood. We perform instead a sequential search of the roots of the estimating equations defining the IFM estimator. Nevertheless, in practice, one can see a "nearly" monotone increase of the log-likelihood after each step of the IFM algorithm, as the normal equations of MLE remains quite close the IFM equations.

For the estimation of HMC with SIRV based on EM, we have shown that our algorithm enables each sensor to be decoupled, so that we could consider the case of different textures for different sensors. Nevertheless, the complexity of the algorithm is not increased significantly, and it gives similar estimates. The interest of the proposed algorithm also lies in the fact that it makes it possible to derive estimation methods for recent multivariate models proposed in [8], or for generalizations of HMCs such as pairwise Markov chains [7], [14], [32], or triplet Markov chains [33].

Possible extensions of this work concern the development of various copulas, and of criteria for the choice of copulas in image and signal processing, in order to fully exploit the generality of copula-based models. We considered HMC but other

hidden Markov models, like trees or fields, could be used in a similar way: the first results obtained with hidden Markov trees seem very promising [17].

Let us also mention that other general estimation methods would have been well-suited in the hidden data setting, as "iterative conditional estimation" (ICE [14], [19]).

REFERENCES

[1] , R. Adler, R. Feldman, and M. Taqqu, Eds., *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Boston, MA: Birkhauser, 1998.

[2] G. E. Andrews, R. Askey, and R. Roy, *Special Functions, Encyclopedia of Mathematics and Its Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[3] A. Banarjee, I. S. Dhillon, J. Ghosh, and S. Sra, "Clustering on the unit hypersphere using Von Mises-Fisher distributions," *J. Mach. Learning Res.*, vol. 6, pp. 1345–1382, 2005.

[4] F. Barbaresco, "Recursive eigendecomposition via autoregressive analysis and ago-antagonistic regularization," in *Proc. Int. Conf. Acoust., Speech Signal Processing (ICASS'97)*, 1997, pp. 3969–3972.

[5] T. J. Barnard and F. Khan, "Statistical normalization of spherically invariant non-Gaussian clutter," *IEEE J. Oceanic Eng.*, vol. 29, no. 2, pp. 303–309, Apr. 2004.

[6] T. J. Barnard and D. D. Weiner, "Non-Gaussian clutter modeling wit generalized spherically invariant random vectors," *IEEE Trans. Signal Processing*, vol. 44, no. 10, pp. 2384–2390, Oct. 1996.

[7] N. Brunel and W. Pieczynski, "Unsupervised signal restoration using hidden Markov chains and copulas," *Signal Processing*, vol. 85, pp. 2304–2315, 2005.

[8] N. Brunel, W. Pieczynski, and S. Derrode, "Copulas in vectorial hidden Markov chains for multicomponent image segmentation," in *In Proc. Int. Conf. Acoust., Speech Signal Processing (ICASSP'05)*, 2005, pp. 717–720.

[9] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Chains*. New York: Springer, 2005.

[10] M. Carpentier, *Principles of Modern Radar Systems*. New York: Wiley, 1988.

[11] G. Celeux and J. Diebolt, "The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem," *Comp. Statist. Quarterly*, vol. 2, pp. 73–82, 1985.

[12] E. Conte, A. De Maio, and G. Ricci, "Recursive estimation of the covariance matrix of a compound Gaussian process and its application to adaptive CFAR detection," *IEEE Trans. Signal Processing*, vol. 50, no. 8, pp. 1908–1915, Aug. 2002.

[13] Y. Delignon and W. Pieczynski, "Modeling non-Rayleigh speckle distribution in SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 40, no. 6, pp. 1430–1435, Jun. 2002.

[14] S. Derrode and W. Pieczynski, "Signal and image segmentation using pairwise Markov chains," *IEEE Trans. Signal Processing*, vol. 52, no. 9, pp. 2477–2489, Sep. 2004.

[15] H.-B. Fang, K.-T. Fang, and S. Kotz, "The meta-elliptical distributions with given marginals," *J. Multivariate Anal.*, vol. 82, pp. 1–16, 2002.

[16] J.-D. Fermanian and O. Scaillet, Some statistical pitfalls in copula modeling for financial applications FAME—Int. Ctr. Financial Asset Manag. Eng. Univ. Geneva, Tech. Rep. 108, 2004.

[17] F. Flitti, C. Collet, and A. Joannic-Chardin, "Unsupervised multiband image segmentation using hidden Markov quadtree and copulas," in *Proc. Int. Conf. Image Processing (ICIP'05)*, 2005, pp. 634–637.

[18] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing*, vol. 82, no. 12, pp. 1847–1859, 2002.

[19] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 5, pp. 465–475, May 1997.

[20] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall International Editions, 2000.

[21] H. Joe, *Multivariate Models and Dependence Concepts, Volume 73 of Monographs on Statistics and Applied Probability*. London, U.K.: Chapman and Hall, 1997.

[22] N. L. Johnson and S. Kotz, *Continuous Univariate Distributions*, second ed. New York: Wiley, 1994, vol. 1.

[23] I. R. Joughin, D. B. Percival, and D. P. Winebrenner, "Maximum likelihood estimation of K distribution parameters for SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 5, pp. 989–999, Sep. 1993.

[24] J. T. Kent and D. E. Tyler, "Redescending M-estimates of multivariate location and scatter," *Annals Stat.*, vol. 19, no. 4, pp. 2102–2119, 1991.

[25] J. Lapuyade-Lahorgue and F. Barbaresco, "Radar detection using siegel distance between auto-regressive processes, application to hf and X-band radar," in *Proc. IEEE Radar Conf.*, Rome, Italy, May 2008, pp. 1–6.

[26] F. Le Chevalier, *Principles of Radar and Sonar Signal Processing*. Boston, MA: Artech House, 2003.

[27] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Annals Stat.*, vol. 4, pp. 51–67, 1976.

[28] G. J. McLachlan and T. Khrishnan, *The EM Algorithm and Extensions*. New York: Wiley Interscience, 1996.

[29] G. J. McLachlan and D. Peel, *Finite Mixture Models. Wiley series in Probability and Statistics: Applied Probability and Statistics Section*. New York: Wiley, 2000.

[30] X. L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ECM algorithm—A general framework," *Biometrika*, vol. 80, no. 2, pp. 267–278, 1993.

[31] R. B. Nelsen, *An introduction to Copulas. Number 139 in Lecture Notes in Statistics*. Berlin, Germany: Springer-Verlag, 1998.

[32] W. Pieczynski, "Pairwise Markov chains," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 5, pp. 634–639, 2003.

[33] W. Pieczynski, "Multisensor triplet Markov chains and theory of evidence," *Int. J. Approximate Reasoning*, vol. 45, no. 1, pp. 1–16, 2007.

[34] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[35] M. Rangaswamy, D. Weiner, and A. Ozturk, "Non-Gaussian random vector identification using spherically invariant random process," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 111–124, 1993.

[36] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods. Springer texts in Statistics*. New York: Springer-Verlag, 1994.

[37] W. Roberts and S. Furui, "Maximum likelihood estimation of K-distribution parameters via the expectation-maximization algorithm," *IEEE Trans. Signal Processing*, vol. 48, no. 12, pp. 3303–3306, Dec. 2000.

[38] K. Yao, "A representation theorem and its applications to spherically invariant random processes," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 5, pp. 600–608, Sep. 1973.

N. J.-B. Brunel was born in France in 1976. He received the M.S. degree in mathematics (probability) from the University Rennes 1, Rennes, France, in 2001 and the Ph.D. degree in statistics from the University Paris 6 (Pierre et Marie Curie), Paris, France, in 2005.

He was a Statistician at Rhodia in 2000. During his Ph.D. studies, he was supported by the "Délégation Générale pour l'Armement" (French Ministery of Defence). From 2006 to 2007, he was a post-Doctoral Fellow with EURANDOM, University of Eindhoven, The Netherlands and with IBISC, University of Evry, France. Since 2007, he has been an Associate Professor in statistics at Ecole Nationale Supérieure d'Informatique pour l'Industrie et l'Entreprise (ENSIIE), Evry, France. He is member of AMIS-Bio (a research group a machine learning and bioinformatics), which is a team of laboratory IBISC, Unversity of Evry, France. He is interested in parametric and nonparametric estimation, with a particular interest for dynamical models such as differential equations used in systems biology.

Jerome Lapuyade-Lahorgue was born in Cholet, France, in 1978. He received the Ph.D. degree in several extensions of hidden Markov chains with applications on radar processing from the Institut National des Télécommunications, Evry, France.

His Ph.D. was supported by the "Délégation Générale pour l'Armement "(French Ministery of Defence) with a consortium with the French Compagny "Thales Air Systems." Since December 2008, he has been pursuing a Postdoctoral Fellowship at the Datalogisk Institute, University of Copenhagen, Denmark, supported by the European Research Consortium for Informatics and Mathematics (ERCIM). His principal research interests concern Bayesian inference, hidden Markov model, information theory, Monte Carlo algorithms, and image and signal processing.

Wojciech Pieczynski received the Doctorat d'Etat degree in mathematical statistics from the Université Pierre et Marie Curie, Paris, France, in 1986.

He is currently a Professor at the Telecom ParisSud and Head of the Communications, Images, and Information Processing (CITI) Department. His research interests include mathematical statistics, stochastic processes, and statistical image processing.