Chapter 17 *(author unedited version for "Signal Processing for Remote Sensing", Second Edition, editor C. Chen, Taylor & Francis, to appear in 2012)*

Probability Density Function Estimation for Classification of High Resolution SAR Images [1]

Vladimir A. Krylov, Gabriele Moser, Sebastiano B. Serpico, and Josiane Zerubia

17.1 Introduction

In modern remote sensing, the use of synthetic aperture radar (SAR) represents an important source of information for Earth observation. Recent improvements have enabled modern satellite SAR missions, such as COSMO-SkyMed, TerraSAR-X, RADARSAT-2, to acquire high resolution (HR) data (up to metric resolution) with a very short revisit time (e.g. 12 hours for COSMO-SkyMed). In addition, SAR is robust with respect to lack of illumination and atmospheric conditions. Together, these factors explain the rapidly growing interest in SAR imagery for various applications, such as flood/fire monitoring, urban mapping and epidemiological surveillance. HR imagery allows to appreciate various ground materials resulting in highly mixed distributions. The resulting spatial heterogeneity is a critical problem in applications to image classification (estimation of class-conditional statistics) or filtering (estimation of local statistics, e.g., in moving-window approaches) [1]. Analysis and modeling of heterogenous HR SAR data pose a difficult statistical problem, which, to the best of our knowledge, has not been sufficiently addressed so far.

Accurate modeling of statistical information is a crucial problem in the context of SAR image processing and its applications. Specifically, an accurate probability density function (pdf) estimate can effectively improve the performance of SAR image denoising [1], classification [2] and target detection [3]. Over the years, a number of methods have been proposed for modeling SAR amplitude pdfs. Nonparametric methods, e.g., Parzen window estimator [4], and support vector machines [5], do not assume any specific analytical model for the unknown pdf, thus providing a higher flexibility, although usually involving manual specification of internal architecture parameters [4]. Parametric methods postulate a given mathematical model for each pdf and formulate the pdf estimation problem as a parameter estimation problem. Empirical pdf models, including lognormal [1], Weibull [1], Fisher [2] and, recently, the generalized Gamma distribution (GΓD) [6], have been reported to accurately model amplitude SAR images with different heterogenous surfaces. Several theoretical models, such as Rayleigh [1], Nakagami [1], generalized Gaussian Rayleigh (GGR) [7], symmetric-α-stable generalized Rayleigh (SαSGR) [8], $\mathcal{K}$ [9] ($\mathcal{K}$-root for amplitudes), $\mathcal{G}$ [10], have been derived from specific physical hypotheses for SAR images with different properties. However, several parametric families turned out to be effective only for specific land cover typologies [1], making the choice of a single optimal SAR amplitude parametric pdf model a hard task, especially in case of heterogenous imagery. To solve this problem, the dictionary-based stochastic expectation maximization (DSEM) approach [11] was designed and validated on medium resolution SAR imagery. It addressed the problem by adopting a finite mixture model [12] for the SAR amplitude pdf, i.e., by postulating the unknown amplitude pdf to be a linear combination of parametric components, each corresponding to a specific statistical population.

[1]This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. This material may not be reposted without the explicit permission of the copyright holder

In this chapter, we first address the general problem of modeling the statistics of single-channel SAR amplitude images and, specifically, HR SAR. Given the variety of approaches above, we extend and enhance the DSEM technique proposed in [11] for coarser-resolution SAR. We expect DSEM to be an appropriate tool for this modeling problem, since it is a flexible method, intrinsically modeling SAR statistics as resulting from mixing several populations, and it is not constrained to a specific choice of a given parametric model allowing to benefit from many of them (dictionary approach). Thus, we extend the earlier DSEM approach to HR satellite SAR imagery and enhance it by introducing a novel procedure for estimating the number of mixture components, which enables to appreciably reduce its computational complexity (as much as 5 times in some cases), resulting in an Enhanced DSEM algorithm (EDSEM).

Building on the proposed method for single channel SAR pdf estimation we proceed to multi-channel joint pdf estimation and classification. Contemporary satellite SAR missions are capable of registering polarimetric (PolSAR) imagery, which provides a more complete description of landcover scattering behavior than single-channel SAR data [1, 13]. The potential for improved classification accuracy with data in several polarizations, compared to single-channel data, explains the special interest to polarimetric SAR image classification. Furthermore, several current satellite SAR systems, e.g., TerraSAR-X, COSMO-SkyMed, RADARSAT-2, support, at least, dual-pol acquisition modes. In this chapter we investigate the quad polarization (quad-pol), dual polarization (dual-pol), as well as single polarization (single-pol) SAR imagery scenarios.

A wide variety of methods have been developed so far for the classification of polarimetric SAR data [13]. We indicate some recent methods classified based on the employed methodological approach: maximum likelihood [14–18], neural networks [19, 20], support vector machines [21], fuzzy methods [22, 23], stochastic complexity [24], spectral graph partitioning [25], wavelet texture models [26] and other approaches [27, 28]. In this chapter we develop a classification method based on a maximum likelihood approach. As such, this method explicitly specifies a pdf describing the statistics of polarimetric SAR data. Previously, several models have been proposed for this purpose: the classical Wishart distribution [14, 29], the K-distribution [15, 30] for textured areas, the K-Wishart distribution [17] designed to improve the distinguishability of non-Gaussian regions, the G-distribution [16, 31] for extremely heterogeneous areas, the Ali-Mikhail-Haq copula-based model [32] combined with the Sinclair matrix representation, and the KummerU distribution [18] for Fisher distributed texture. These models were developed for the multilook complex-valued SAR image statistics. In this chapter, we study the problem of PolSAR image classification using only the amplitude data and not the complex-valued data. This is an important data typology since several image products provided by novel high resolution satellite SAR systems are geocoded ellipsoid-corrected amplitude (intensity) images, and also because several earlier coarser resolution sensors (e.g., ERS) primarily used this modality.

The classification technique developed in this chapter combines the Markov random field (MRF) approach to Bayesian image classification with the finite mixture EDSEM amplitude pdf estimator. MRFs represent a general family of probabilistic image models that provide a convenient and consistent way to characterize context dependent data [33]. The resulting DSEM-MRF technique is a simple and efficient tool for single-channel SAR classification. In order to support multi-channel PolSAR data, copula theory is used for modeling the joint class-conditional distributions of the multiple channels, resulting in a Copula-DSEM-MRF approach (CoDSEM-MRF). The employed joint distribution modeling tool, copulas [34], is a rapidly developing statistical tool that was designed for constructing joint distributions from marginals with a wide variety of allowable dependence structures. For every class the choice of an optimal copula from a dictionary of copulas is performed by a dedicated criterion. The concept of copulas is relatively new in image processing, and has recently emerged in remote sensing methods [32, 35, 36]. We compare the results of the method proposed in this chapter with an earlier

developed copula-based classification technique [32] to demonstrate the higher adequacy of our model. The proposed CoDSEM-MRF HR PolSAR classification technique is based on three flexible statistical modeling concepts, i.e., copulas, finite mixtures and MRFs, and constitutes an efficient and robust approach with respect to possibly sophisticated statistics of the considered classes of interest.

The chapter is organized as follows. In Section 2 we introduce the EDSEM technique for the amplitude SAR pdf estimation. In Section 3 we proceed to multi-channel pdf construction by first introducing the concept of copulas and then derive a novel method for PolSAR joint pdf estimation. In section 4 we build on the result of the third section and introduce a supervised PolSAR classification approach. Section 5 presents experimental results on HR single channel SAR and PolSAR data to demonstrate the performance of the developed pdf estimation and classification techniques along with comparisons with relevant benchmark approaches. Section 6 closes the chapter with conclusions.

17.2 Single channel SAR amplitude pdf estimation based on mixture models

In this section we consider the single channel SAR imagery amplitude pdf estimation problem. To take into account possible heterogenous scenarios, when several distinct land-cover typologies are present in the same SAR image, a finite mixture model (FMM) [12] for the distribution of grey levels is assumed. An amplitude SAR image is modeled as a set $\mathcal{I} = \{r_1, \ldots, r_N\}$ of independent samples drawn from a mixture pdf with K components:

$$p_r(r) = \sum_{i=1}^{K} P_i p_i(r), \quad r \geqslant 0, \tag{17.1}$$

where $p_i(r)$ are parametric pdfs and $\{P_i\}$ are mixing proportions: $\sum_{i=1}^{K} P_i = 1$, with $0 \leqslant P_i \leqslant 1, i = 1, \ldots, K$. Each component $p_i(r)$ in (17.1) is modeled by resorting to a finite dictionary $\mathcal{D} = \{f_1, \ldots, f_8\}$ (see Table 17.1) of SAR-specific distinct parametric pdf families $f_j(r|\theta_j)$, parameterized by $\theta_j \in A_j, \quad j = 1, \ldots, 8$. Dealing with HR heterogenous SAR data, along with classical SAR pdf models, such as Lognormal, Weibull, Nakagami, $\mathcal{K}$-root, GGR and SαSGR, we include the Fisher model, which was demonstrated to perform well for heterogenous HR SAR imagery [37], and GΓD, highly flexible empirical model, including Lognormal, Weibull and Nakagami as special cases [6], into the considered dictionary $\mathcal{D}$.

As discussed in [11], considering the variety of estimation approaches for FMMs, an appropriate choice for this particular finite mixture estimation problem is the stochastic expectation maximization (SEM) scheme [12]. SEM was developed as a stochastic modification of the classical EM algorithm, involving stochastic sampling on every iteration, and demonstrating higher chances of avoiding local maxima of the likelihood function [12]. By definition, SEM is an iterative estimation procedure dealing with the problem of *data incompleteness*. In case of FMMs the complete data is represented by the set $\{(r_i, s_i), i = 1, \ldots, N\}$, where r_i are the observations (SAR amplitudes) and s_i - the missing labels: given an FMM with K components, $s_i \in \{\rho_1, \ldots, \rho_K\}$ denotes to which of the K components the i-th observation belongs.

Instead of adopting the maximum likelihood (ML) estimates, as the classical SEM scheme [12] suggests, in DSEM the Method of Log-Cumulants (MoLC) [2] for component parameter estimation is adopted, which has been demonstrated to be a feasible and effective estimation tool for all the pdfs in $\mathcal{D}$ [2, 6, 11]. MoLC has recently been proposed as a parametric pdf estimation technique suitable for distributions defined on $[0, +\infty)$, and has been widely applied in the context of SAR-specific parametric families for amplitude and intensity data modeling. MoLC adopts the Mellin transform by analogy to

the Laplace transform in moment generating function [2]. Given a non-negative random variable u, the second-kind characteristic function ϕ_u of u is defined as the Mellin transform [2] $\mathcal{M}$ of the pdf of u, i.e.:

$$\phi_u(s) = \mathcal{M}(p_u)(s) = \int_0^{+\infty} p_u(u) u^{s-1} du, \quad s \in \mathbb{C}.$$

The derivatives $\kappa_j = [\ln \phi_u]^{(j)}(1)$ are the j-th order log-cumulants, where $^{(j)}$ denotes for the j-th derivative, $j = 1, 2, \ldots$. If the Mellin transform converges for s lying in a neighborhood of 1, the following MoLC equations take place [2]:

$$\begin{cases} \kappa_1 = E\{\ln u\} \\ \kappa_j = E\{(\ln u - \kappa_1)^j\} \end{cases}, \quad j = 2, 3.$$

Analytically expressing κ_js, $j = 1, 2, 3$, as functions of unknown parameters and estimating these log-cumulants in terms of sample log-moments one derives a system of nonlinear equations. These equations have one solution for any observed values of log-cumulants for all the pdfs in $\mathcal{D}$ (see Table 17.1), except for, in some cases, GGR, $\mathcal{K}$-root, see [11], and GΓD due to a complicated parameter estimation procedure [38].

For the purpose of K estimation we adopt a procedure similar to the one suggested in [12], that consists of initializing SEM with $K_0 = K_{max}$, and then allowing components to be eliminated from the mixture during the iterative process, once their priors P_i become too small, thus decreasing K. This strategy provides efficient K^* estimates consistent with DSEM estimates [11], allowing however to significantly reduce the computational complexity, especially for high values of K_{max}.

Thus, each iteration of EDSEM consists of the following steps:

• **E-step:** compute, for each greylevel z and i-th component, the posterior probability estimates corresponding to the current pdf estimates, i.e. $z = 0, \ldots, Z-1$:

$$\tau_i^t(z) = \frac{P_i^t p_i^t(z)}{\sum_{j=1}^{K_t} P_j^t p_j^t(z)}, \quad i = 1, \ldots, K_t,$$

where $p_i^t(\cdot)$ is the ρ_i-conditional pdf estimate on the t-th step.

• **S-step:** sample the label $s^t(z)$ of each greylevel z according to the current estimated posterior probability distribution $\{\tau_i^t(z) : i = 1, \ldots, K_t\}$, $z = 0, \ldots, Z-1$.

• **MoLC-step:** for the i-th mixture component, compute the following histogram-based estimates of the mixture proportions and the first three log-cumulants:

$$P_i^{t+1} = \frac{\sum_{z \in Q_{it}} h(z)}{\sum_{z=0}^{Z-1} h(z)}, \quad \kappa_{1i}^t = \frac{\sum_{z \in Q_{it}} h(z) \ln z}{\sum_{z \in Q_{it}} h(z)}, \quad \kappa_{bi}^t = \frac{\sum_{z \in Q_{it}} h(z)(\ln z - \kappa_{1i}^t)^b}{\sum_{z \in Q_{it}} h(z)}, \quad i = 1, \ldots, K_t,$$

where $b = 2, 3$; $h(z)$ is the image histogram; $Q_{it} = \{z : s^t(z) = \rho_i\}$ is the set of grey levels assigned to the i-th component; then, solve the corresponding MoLC equations (see Table 17.1) for each parametric family $f_j(\cdot|\theta_j)$ ($\theta_j \in A_j$) in the dictionary, thus computing the resulting MoLC estimate θ_{ij}^t, $j = 1, \ldots, M$.

• **K-step:** $\forall i$, $i = 1, \ldots, K_t$: if $P_i^{t+1} < \gamma$, eliminate the i-th component and correspondingly update K_{t+1}. The choice of threshold γ does not appreciably affect EDSEM, provided it is small, e.g. 0.005.

• **Model Selection-step:** for each mixture component i, compute the log-likelihood of each estimated pdf $f_j(\cdot|\theta_{ij}^t)$ according to the data assigned to the i-th component:

$$L_{ij}^t = \sum_{z \in Q_{it}} h(z) \ln f_j(z|\theta_{ij}^t), \quad i = 1, \ldots, K_{t+1},$$

and define $p_i^{t+1}(\cdot)$ as the estimated pdf $f_j(\cdot|\theta_{ij}^t)$ yielding the highest value of $L_{ij}^t, \quad j = 1, \ldots, M$.

The sequence of estimates generated by SEM is a discrete-time random process $\{\Theta_t\}_{t=0}^{\infty}$ and converges to a unique stationary distribution, and the maximum likelihood estimate of the mixture parameters is asymptotically equivalent to the mathematical expectation of this stationary distribution. This behavior has been proved under suitable assumptions [12], which do not hold strictly for all the pdfs in $\mathcal{D}$. This means that, in general, the convergence of the SEM estimation procedure is not guaranteed when dealing with finite mixtures drawn from the considered dictionary $\mathcal{D}$. This owes to the fact, that in this chapter we suggest working with a complicated set of distributions, for most of which it is impossible to demonstrate the SEM convergence even if they were considered as the only types of distributions to be mixed. However, we recall that SEM, compared to the classical EM or other deterministic variants for FMM, was specifically designed to improve the exploratory properties of EM in case of multimodal likelihood function [12]. In our experience, we have very rarely observed the lack of convergence, i.e., if the sufficient number of SEM iterations was performed, the SEM estimates obtained after running the algorithm several times were consistently close even starting with the random initialization.

Table 17.1: Pdfs and MoLC equations for the parametric pdf families in $\mathcal{D}$. Here $\Gamma(\cdot)$ is the Gamma function [39], $K_\alpha(\cdot)$ the αth order modified Bessel function of the second kind [39], $J_0(\cdot)$ is the zero-th order Bessel function of the first kind [39], $\Psi(\cdot)$ the Digamma function [39], $\Psi(\nu,\cdot)$ the νth order polygamma function [39] and $G_\nu(\cdot)$ are the specific integral functions for GGR [7]

Family		Probability density function	MoLC equations
f_1	Lognormal	$f_1(r\|m,\sigma) = \frac{1}{\sigma r\sqrt{2\pi}} \exp[-\frac{(\ln r - m)^2}{2\sigma^2}],\ r > 0,$	$\kappa_1 = m, \quad \kappa_2 = \sigma^2.$
f_2	Weibull	$f_2(r\|\eta,\mu) = \frac{\eta}{\mu^\eta} r^{\eta-1} \exp[-\left(\frac{r}{\mu}\right)^\eta],\ r \geqslant 0,$	$\kappa_1 = \ln\mu + \eta^{-1}\Psi(1), \quad \kappa_2 = \eta^{-2}\Psi(1,1).$
f_3	Fisher	$f_3(r\|L,M,\mu) = \frac{\Gamma(L+M)}{\Gamma(L)\Gamma(M)} \frac{[ar]^L}{r[1+ar]^{L+M}},\ r > 0,$ with $a = L/(M\mu),$	$\kappa_1 = \ln\mu + (\Psi(L) - \ln L) - (\Psi(M) - \ln M),$ $\kappa_j = \Psi(j-1,L) + (-1)^j\Psi(j-1,M),\ j = 2,3.$
f_4	GΓD	$f_4(r\|\nu,\kappa,\sigma) = \frac{\nu}{\sigma\Gamma(\kappa)} \left(\frac{r}{\sigma}\right)^{\kappa\nu-1} \exp\left[-\left(\frac{r}{\sigma}\right)^\nu\right],\ r \geqslant 0,$	$\kappa_1 = \Psi(\kappa)/\nu + \ln\sigma, \quad \kappa_j = \Psi(j-1,\kappa)/\nu^j,\ j = 2,3.$
f_5	Nakagami	$f_5(r\|L,\lambda) = \frac{2}{\Gamma(L)} (\lambda L)^L r^{2L-1} \exp\left[-\lambda L r^2\right],\ r \geqslant 0,$	$2\kappa_1 = \Psi(L) - \ln\lambda L, \quad 4\kappa_2 = \Psi(1,L).$
f_6	$\mathcal{K}$-root	$f_6(r\|L,M,\mu) = \frac{4}{\Gamma(L)\Gamma(M)} [\lambda LM]^{(L+M)/2} r^{L+M-1}$ $\times K_{M-L}\left(2r\left[\lambda LM\right]^{1/2}\right),\ r \geqslant 0,\ L \geqslant M > 0,$	$2\kappa_1 = \Psi(L) + \Psi(M) - \ln LM\lambda,$ $2^j\kappa_j = \Psi(j-1,L) + \Psi(j-1,M),\ j = 2,3.$
f_7	GGR	$f_7(r\|\lambda,\gamma) = \frac{\gamma^2 r}{\lambda^2\Gamma^2(\lambda)} \int_0^{\pi/2} \exp\left[-(\gamma r)^{1/\lambda} s(\theta)\right] d\theta,\ r \geqslant 0,$ with $s(\theta) = \lvert\cos\theta\rvert^{1/\lambda} + \lvert\sin\theta\rvert^{1/\lambda},$	$\kappa_1 = \lambda\Psi(2\lambda) - \ln\gamma - \lambda G_1(\lambda)[G_0(\lambda)]^{-1},$ $\kappa_2 = \lambda^2[\Psi(1,2\lambda) + \frac{G_2(\lambda)}{G_0(\lambda)} - \left(\frac{G_1(\lambda)}{G_0(\lambda)}\right)^2].$
f_8	SαSGR	$f_8(r\|\alpha,\gamma) = r \int_0^{+\infty} \rho\ \exp[-\gamma\rho^\alpha] J_0(r\rho) d\rho,\ r \geqslant 0,$	$\alpha\kappa_1 = (\alpha-1)\Psi(1) + \ln\gamma 2^\alpha, \quad \kappa_2 = \alpha^{-2}\Psi(1,1).$

17.3 Polarimetric SAR amplitude joint pdf-estimation via copulas

In this section, we introduce a novel approach to PolSAR amplitude pdf-estimation based on the use of the EDSEM algorithm for single channel SAR pdf estimation and the statistical concept of copulas for constructing the joint multi-channel pdf function.

We recall, that a D-dimensional copula is a multivariate joint distribution defined on $[0,1]^D$ such that its every marginal distribution is uniform on $[0,1]$. We will define any copula by its cumulative distribution function (cdf) C. Some basic facts and definitions from the copula theory are summarized in Appendix A. The copula-approach to define a multivariate distribution is based on the idea that a simple transformation of marginal distributions can be made in such a way that each transformed marginal has a uniform distribution [34]. This approach develops the concept of multivariate cdf construction and, therefore, provides a more general and flexible method of joint pdf-estimation.

The developed in this chapter pdf-estimation approach consists of the following two steps:

EDSEM step. The marginal pdfs of the polarization channels are separately estimated by applying the EDSEM approach. For the d-th channel, $d = 1, \ldots, D$, the mixture EDSEM pdf estimator $p_d(x_d)$ and the corresponding cdf $F_d(x_d)$ are written in the following form:

$$p_d(x_d) = \sum_{i=1}^{K_d} P_{di} f_{di}(x_d), \qquad F_d(x_d) = \sum_{i=1}^{K_d} P_{di} F_{di}(x_d), \tag{17.2}$$

where x_d represents the amplitudes of the given polarization channel d and K_d is the number of components in the d-th mixture. F_{di} and p_{di} represent the i-th mixture component in the cdf and pdf domains, respectively, and P_{di} are the related mixture proportions; p_{di} is automatically drawn by EDSEM from a dictionary of SAR-specific parametric families, $i = 1, \ldots, K_d$.

In case of PolSAR pdf estimation, the calculation of cdfs is needed to merge the marginal EDSEM pdf estimates of polarization channels into joint pdfs via copulas (see Eq. (17.3)). Therefore, the list of pdfs in the dictionary $\mathcal{D}$ presented in Table 17.1 has been restricted, so that it includes the pdfs that either have an analytical closed-form expression (f_1, f_2), or a simple numerical approximation procedure (f_4, f_5) for the related cdf. For the rest of the pdf families there are no closed-form cdfs and numerical approximation is computationally intensive. Moreover, further experiments demonstrate that very accurate pdf estimates can be obtained with this reduced dictionary. Therefore, from now on we will consider the restricted dictionary $\overline{\mathcal{D}} = \{f_1, f_2, f_4, f_5\}$.

Now we address the reason why we suggest the use of EDSEM instead of adopting one of the single-family pdf models. First, EDSEM is an efficient and automatic tool for SAR amplitude pdf estimation, and it is capable of providing estimates of higher accuracy compared to single parametric pdf models [11, 40]. Specifically, EDSEM was experimentally designed to accurately estimate the statistics of a HR SAR imagery [40]. Second, the underlying mixture assumption in EDSEM enables accurate characterization of the inhomogeneous classes of interest, i.e., the classes that contain several different landcover subclasses. This is a very important property when dealing with HR imagery, since the corresponding image statistics are usually strongly mixed due to the high level of spatial detail appreciable at high resolution. However, even in the case of homogenous classes, EDSEM can be viewed as a tool for choosing the best single pdf model from the set of pdfs in the dictionary.

Copula step. The goal of this step is to merge the marginal pdfs, which correspond to polarization channels estimated on the EDSEM-step, into a joint multivariate pdf describing the joint amplitude distribution of a PolSAR image. As recalled in Section 17.1, a number of multivariate pdf models have been proposed earlier to deal with this problem. In this chapter we suggest a more flexible solution

to this problem, which allows to achieve higher pdf-approximation accuracy thanks to a wider panel of dependency structures considered. We model the joint pdf $p(\mathbf{x})$ via an appropriate copula from the marginal distributions estimated in Eq. (17.2). Thus, the joint pdf is constructed as:

$$p(\mathbf{x}) = p_1(x_1) \cdots p_D(x_D) \frac{\partial^D C^*}{\partial x_1 \cdots \partial x_D}(F_1(x_1), \ldots, F_D(x_D)). \tag{17.3}$$

where C^* is coming from an appropriate family of copulas.

In this chapter, we consider several families of one-parametric copulas. By taking advantage of the connection between copulas and the Kendall's tau ranking coefficient τ, we obtain closed-form equations to estimate the copula parameter θ for each copula from the empirically observed value $\hat{\tau}$.

In order to fully exploit the modeling potential of copulas, we consider a dictionary $\mathcal{D}_C$ of several one-parametric copulas $C_c(\mathbf{x}|\theta)$. In the bivariate case (corresponding to dual-pol SAR imagery) we consider 10 copulas: 6 Archimedean (Clayton, Gumbel, Frank, Ali-Mikhail-Haq, A12, A14) [34], a copula with a quadratic section (Farlie-Gumbel-Morgenstern) [34], 2 elliptical (Gaussian and Student-t[2] [41]) and a non-Archimedean copula with simultaneous presence of an absolutely continuous and a singular component (Marchal-Olkin) [34]. Here the names for A12 and A14 copulas originate from their positions in the list of Archimedean copulas in [34]. When $D \geqslant 3$ the same dictionary cannot be considered since the described parameter estimation strategy is not applicable to all the copula families. Therefore, in such case, we consider only the first three copula families in Table 17.2: Clayton, Gumbel and Frank. Such a choice of copulas is capable of modeling a sufficiently wide variety of dependence structures, and covers most copula applications [42]. The considered dictionary of copulas is summarized in Table 17.2 (further information can be found in [34, 41, 42]).

The choice of a specific copula C^* from the dictionary (see Eq. (17.3)) is performed as follows. First, given an estimator $\hat{\tau}$ of τ, we need to decide whether a specific copula is appropriate for modeling the dependence with such a level of Kendall's tau rank correlation. Indeed, some copulas are specific to marginals with a low level of dependence, others deal with strongly dependent marginals, and still others are capable of modeling all levels of dependency. In other words, the effective list of copulas is limited to those which are capable of accurately modeling the specific empirically estimated value $\hat{\tau}$ (see Table 17.2). Thus, we first discard the copulas for which the current sample $\hat{\tau}$ is outside the related τ-relevance interval; for the remaining copulas, we derive an estimate $\hat{\theta}$ of the related parameter by the above-mentioned Kendall's tau method. Next, from the list of remaining copulas we choose the copula with the highest p-value provided by a Pearson chi-square test-of-fitness (PCS) [43]. In general, PCS tests the null hypothesis that the frequency distribution of certain events observed in a sample is consistent with a particular theoretical distribution. The chi-square statistic is constructed as follows:

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}, \tag{17.4}$$

where O_i and E_i are the observed and the hypothetical frequencies, respectively, and n is the number of outcomes. PCS is one of the statistical tests whose results follow a chi-square distribution [43], i.e. $X^2 \sim \chi^2_{n-r-1}$, where r is the number of reductions of degrees of freedom (typically, the number of parameters for parametric cdfs). The reference to χ^2 distribution allows p-values for the considered null hypothesis to be calculated. In our case, the null hypothesis in PCS is that the sample frequencies

[2]Unlike all other considered copulas, the Student-t copula depends on two parameters: the linear correlation coefficient θ and the number of degrees of freedom ν. To avoid a cumbersome $\hat{\nu}$ ML estimation [41] we employ the following approach: we consider separately $\nu = 3k$, with k from 1 to 9 due to the fact that as ν grows large ($\nu \geqslant 30$) Student-t copula becomes indistinguishable from a Gaussian copula [41]. In other words, instead of a single two-parametric Student-t copula we consider hereunder nine one-parametric Student-t copulas with fixed values of ν.

$(F_1(x_1), \ldots, F_D(x_D))$ are consistent with the theoretical frequencies (probabilities) predicted by the parametric copula C_c $(c = 1, \ldots, 10)$. This is correct because if $\mathbf{x}$ is distributed with cdf $F(\cdot)$, and $\mathbf{x}_1, \cdots, \mathbf{x}_N$ are independent observations of $\mathbf{x}$, then $F(\mathbf{x}_i)$, $i = 1, \cdots, N$, are independent $[0, 1]$-uniformly distributed random variables [43].

Table 17.2: Considered dictionary $\mathcal{D}_C$ of one-parametric copulas: Clayton, Gumbel, Frank, Ali-Mikhail-Haq, A12, A14, Farlie-Gumbel-Morgenstern, Marchal-Olkin, Gaussian and Student-t, each defined by the function $C_c(\mathbf{u}|\theta)$, $\mathbf{u} \in \mathbb{R}^D$, along with corresponding dimensionalities, $\theta(\tau)$ dependencies and τ-intervals. Here $\phi^{-1}(t)$ and $t_\nu^{-1}(t)$ denote the quantile functions of a standard univariate normal and a standard univariate t_ν distributions respectively [41].

Copula	Dim	$C_c(\mathbf{u}\|\theta)$	$\theta(\tau)$ dependence	τ interval
Clayton	D	$\left(u_1^{-\theta} + \cdots + u_D^{-\theta} - D + 1\right)^{-1/\theta}$	$\theta = \frac{2\tau}{1-\tau}$	$\tau \in (0, 1]$
Gumbel	D	$\exp\left(-\left[(-\ln(u_1))^\theta + \cdots + (-\ln(u_D))^\theta\right]^{1/\theta}\right)$	$\theta = \frac{1}{1-\tau}$	$\tau \in [0, 1]$
Frank	D	$-\frac{1}{\theta}\ln\left(1 + \frac{(e^{-\theta u_1}-1)\cdots(e^{-\theta u_D}-1)}{(e^{-\theta}-1)^{D-1}}\right)$	$\tau = 1 - \frac{4}{\theta^2}\int_0^\theta \frac{t}{e^{-t}-1}dt$	$\tau \in [-1, 0) \cup (0, 1]$
Ali-Mikhail-Haq	2	$\frac{u_1 u_2}{1-\theta(1-u_1)(1-u_2)}$	$\tau = \frac{3\theta-2}{3\theta} - \frac{2}{3}\left(1-\frac{1}{\theta}\right)^2\ln(1-\theta)$	$\tau \in [-0.1817, 0.3333]$
A12	2	$\left(1 + \left[(u_1^{-1}-1)^\theta + (u_2^{-1}-1)^\theta\right]^{1/\theta}\right)^{-1}$	$\theta = \frac{2}{3-3\tau}$	$\tau \in [0.3334, 1]$
A14	2	$\left(1 + \left[(u_1^{-1/\theta}-1)^\theta + (u_2^{-1/\theta}-1)^\theta\right]^{1/\theta}\right)^{-\theta}$	$\theta = \frac{1+\tau}{2-2\tau}$	$\tau \in [0.3334, 1]$
Farlie-Gumbel-Morgenstern	2	$u_1 u_2\Big(1 + \theta(1-u_1)(1-u_2)\Big)$	$\theta = \frac{9}{2}\tau$	$\tau \in [-0.2222, 0.2222]$
Marchal-Olkin	2	$\min\left(u_1^{1-\theta}u_2, u_1 u_2^{1-\theta}\right)$	$\theta = \frac{2\tau}{\tau+1}$	$\tau \in [0, 1]$
Gaussian	2	$\int_{-\infty}^{\phi^{-1}(u_1)}\int_{-\infty}^{\phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\theta^2}}\exp\left(\frac{2\theta xy - x^2 - y^2}{2(1-\theta^2)}\right)\mathrm{dxdy}$	$\theta = \sin\left(\frac{\pi}{2}\tau\right)$	$\tau \in (-1, 1)$
Student-t	2	$\int_{-\infty}^{t_\nu^{-1}(u_1)}\int_{-\infty}^{t_\nu^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\theta^2}}\left(1 + \frac{x^2 - 2\theta xy + y^2}{\nu(1-\theta^2)}\right)^{-(\nu+2)/2}\mathrm{dxdy}$	$\theta = \sin\left(\frac{\pi}{2}\tau\right)$	$\tau \in (-1, 1)$

17.4 PolSAR image classification

In this section, we propose a novel Bayesian supervised classification approach for amplitude PolSAR images. As a likelihood term we propose the use of the joint pdf model developed in the previous sections based on the combined use of copulas and EDSEM for each thematic class. To introduce regularization and ensure robustness against speckle we adopt a Markov Random Field model in the form of the second order Potts model. To optimize the resulting energy we use a Modified Metropolis Dynamics approach. We will refer to the involved joint pdf model as CoDSEM, and to the related classification approach as CoDSEM-MRF. We notice that in case of single channel data there is no need to construct a joint pdf from the marginals, therefore, the classification can be obtained without copula modeling. We will refer to this simpler one-dimensional version as DSEM-MRF.

Once the CoDSEM pdfs for each class $m = 1, \dots, M$ are estimated on the training pixels, the following steps have to be performed in order to perform the classification.

MRF step. In order to take into consideration the contextual information disregarded by the pixel-wise Copula-EDSEM technique, and to gain robustness against the inherent noise-like phenomenon of SAR known as speckle, we adopt a contextual approach based on an MRF model. Following the classical definitions of MRFs (see, e.g., [44, 45]) on the two dimensional lattice S of N observations $\mathbf{y} = \{y_1, \dots, y_N\}$ and class labels $\mathbf{x} = \{x_1, \dots, x_N\}$, $x_i \in \{1, \dots, M\}$, we introduce an isotropic second-order neighborhood system C with cliques of size 2. The Hammersley-Clifford theorem [44] allows for the presentation of the joint probability distribution of an MRF as a Gibbs distribution:

$$P(\mathbf{x}) = Z^{-1} \exp(-U(\mathbf{x}|\beta)),$$

where $Z = \sum_{\mathbf{z}} \exp(-U(\mathbf{z}|\beta))$ is a normalizing constant, β is a positive parameter and $U(\mathbf{x}|\beta)$ is the MRF energy function associated with class labels. More specifically $U(\mathbf{x}|\beta)$ takes the following form:

$$U(\mathbf{x}|\beta) = \sum_{\{s,s'\} \in C} \left[-\beta \, \delta_{x_s = x_{s'}} \right],$$

where δ is the Kronecker delta function:

$$\delta_{x_s = x_{s'}} = \begin{cases} 1, & \text{if } x_s = x_{s'} \\ 0, & \text{otherwise.} \end{cases}$$

Image classification poses a problem of recovering the unobserved data, i.e., class labels. In the case of hidden MRFs, the unobserved data $\mathbf{x}$ are modeled by an MRF and the observed data $\mathbf{y}$, i.e., SAR amplitudes, are assumed to be conditionally independent given $\mathbf{x}$, i.e.:

$$p(\mathbf{y}_i | \mathbf{y}_{S \setminus \{i\}}, x_i) = p(\mathbf{y}_i | x_i), \qquad \forall i \in S,$$

where $\mathbf{y}_i$ is the D-vector of amplitudes $(y_1, \dots, y_D)$ at pixel i, $\mathbf{y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$, and $S \setminus \{i\}$ represents all the pixels of S except i. Involving then the Bayes formula, we get the following energy function for the full data $(\mathbf{y}, \mathbf{x})$:

$$U(\mathbf{x}|\mathbf{y}, \beta) = \sum_{i \in S} \Big(- \sum_{m=1}^{M} \ln p(\mathbf{y}_i | x_i = m) - \beta \sum_{s : \{i,s\} \in C} \delta_{x_i = x_s} \Big). \tag{17.5}$$

Here, the energy at each site i is constructed as a sum of two distinct contributions: probabilities of observing the amplitude vector $\mathbf{y}_i$ and the energy of label configuration in the neighborhood of i. In this formula, given the conditional pdfs $p(\cdot)$ are defined in Eq. (17.3). The energy function U in Eq. (17.5)

has a single parameter β that must be estimated. At this stage, all of the parameters involved in defining the pdfs $p(\cdot)$ are estimated, and, therefore, the resulting energy function (17.2)-(17.5) also has only one parameter. In order to estimate β, we employ a simulated annealing procedure [46,47] with a pseudo-likelihood function PL [44] of the following form:

$$\begin{cases} \ln PL(\mathbf{x}|\beta) = \ln \left[\prod\limits_{s \in S} P(x_s|\mathbf{x}_{S\setminus\{s\}}, \beta) \right] \\ P(x_s|\mathbf{x}_{S\setminus\{s\}}, \beta) = \frac{\exp(-U(x_s|\mathbf{x}_{S\setminus\{s\}}, \beta))}{\sum\limits_{z_s \in X_S} \exp(-U(z_s|\mathbf{x}_{S\setminus\{s\}}, \beta))} \end{cases},$$

for $X_s = \{\omega_1, \ldots, \omega_M\}$.

The proposed simulated annealing procedure generates a sequence $\{\beta_t\}$ of estimates, and employs the normal proposal distribution $N(\beta_t, 1)$ [48] together with an exponentially decreasing cooling schedule $T_t = 0.95 \cdot T_{t-1}$. Considering the convergence in distribution of the related Metropolis algorithm, the final estimate β^* was set by averaging the estimates on the last n iterations. The estimation of β was performed on an ML pre-classification of the image, which associated every pixel with the highest probability-density label according to the CoDSEM class models.

Energy minimization step. This step involves the search for a label configuration that minimizes the energy given by Eq. (17.5), that permits to find the optimal pixel labeling as defined by the Maximum a posteriori (MAP) principle [45]. For this optimization problem the iterative deterministic modified Metropolis dynamics (MMD) [49] algorithm is adopted. This is a compromise between the deterministic "iterated conditional modes" algorithm (ICM) [44], which is a fast local minimization technique, that is strongly dependent on the initial configuration, and the stochastic simulated annealing (SA) [46,47,50], which is a very computationally intensive global minimization approach. MMD is computationally feasible and provides reasonable results in real classification problems [49]. As well as SA, MMD requires a predefined cooling schedule and proceeds as follows:

1. sample a random initial label configuration $\mathbf{x}^0$, define an initial temperature T^0 and parameters: proportion $\alpha \in (0,1)$, number of labels visited per iteration $n_1 \in \mathbb{N}$, cooling schedule parameter $\tau \in (0,1)$, and the termination criterion $\gamma \in \mathbb{R}^+$; initialize iteration counter $k = 0$ and temperature $T_0 = T^0$;

2. set $i = 0$;

3. using uniform distribution pick up a label configuration $\boldsymbol{\eta}$ which differs exactly in one element from $\mathbf{x}^k$;

4. compute $\Delta = U(\boldsymbol{\eta}) - U(\mathbf{x}^k)$ and accept $\boldsymbol{\eta}$ according to the rule:

$$\mathbf{x}^{k+1} = \begin{cases} \boldsymbol{\eta}, & \text{if } \Delta \leqslant 0 \\ \boldsymbol{\eta}, & \text{if } \Delta > 0 \text{ and } \ln(\alpha) \leqslant -\frac{\Delta}{T_k} \\ \mathbf{x}^k, & \text{otherwise} \end{cases};$$

5. calculate $\Delta U_i = |U(\mathbf{x}^{k+1}) - U(\mathbf{x}^k)|$;

6. if $i < n_1$, set $i = i + 1$ and goto Step 3;

7. calculate $\Delta U = \sum_{i=1}^{n_1} \Delta U_i$;

8. if $\left|\Delta U / U(\mathbf{x}^{k+1})\right| > \gamma$ decrease the temperature $T_{k+1} = \tau T_k$, increase $k = k + 1$, and goto Step 2; stop otherwise.

During the early stages of this iterative procedure, the behavior of MMD is close to that of SA, and, on later stages, the MMD behaves as ICM. The proportion between the stages is controlled by the proportion parameter α. Therefore, MMD can be regarded as an ICM strategy with SA initialization and, as such, it provides better exploratory properties than ICM.

17.5 Experiments

In this section we present several experimental results with real space-borne SAR and PolSAR imagery that demonstrate the performance of the proposed single channel SAR pdf estimation approach EDSEM, multi-channel/PolSAR joint pdf estimation approach CoDSEM and the derived supervised classification method CoDSEM-MRF. All the experiments are completed with comparisons to the relevant benchmark methods to further explore the pros and cons of the proposed approaches.

17.5.1 Datasets for experiments

The experiments were performed on the following HR SAR datasets:

- Single-pol HH, single-look, 2.5 m ground resolution COSMO-SkyMed (CSK®), Stripmap image acquired over Piemonte, Italy (©ASI). We present experiments on a 700×500 subimage CSK1 (HH pol, see Fig. 17.1) and a 700×1000 subimage CSK2 (see Fig. 17.6).

- Dual-pol HH/VV, 2.66-look, 6.5 m ground resolution TerraSAR-X Stripmap, geocorrected image acquired over Sanchagang, China (©Astrium GEO-Information Services). We present experiments with the following subimages: 700×500 subimage TSX1 (VV pol, see Fig. 17.1), 700×500 subimage TSX2 (HH pol, see Fig. 17.1), 1000×1200 subimage TSX3 (see Fig. 17.2), and 750×750 subimage TSX4 (see Fig. 17.5).

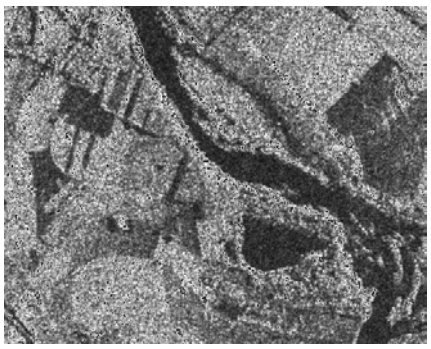
- Fine quad-pol HH/HV/VH/VV, single-look, 7.5 m resolution RADARSAT-2[3] image of Vancouver, Canada. The experiments are performed on a 500×500 subimage RS1 (HV pol, see Fig. 17.1), a 1000×1200 subimage RS2 (Fig. 17.7) and a 500×700 subimage RS3 (Fig. 17.8).

17.5.2 Single-channel pdf estimation experiments

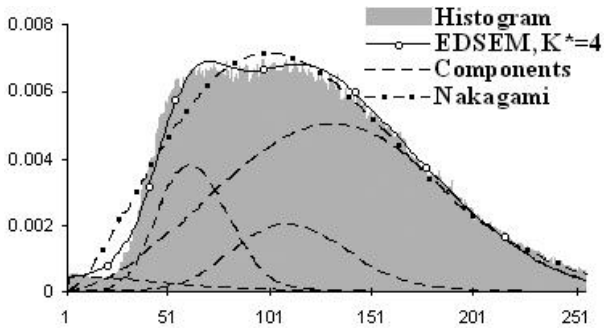
In this subsection we report the experiments of the EDSEM pdf estimation algorithm on several HR SAR datasets. We notice that EDSEM is also used as a structural component in the experiments of the next subsection.

The obtained EDSEM pdf estimates have been assessed quantitatively by computing Kolmogorov-Smirnov distances ($\mathcal{KS}$) between estimates and normalized image histograms [43]. The estimates were also assessed by the Kolmogorov-Smirnov goodness-of-fit test [51] and the corresponding p-values are reported in Table 17.3. Here $1 - p$ can be interpreted as a confidence level at which the corresponding fit hypothesis will be rejected. The following EDSEM parameters were used: initial number of mixture components $K_0 = 6$, number of iterations $T = 200$. Both values were set manually high enough to give sufficient burn-in for SEM and their further increase did not appreciably affect the results. For the sake of comparison, for all test images we provide the best fitting pdf (with the smallest $\mathcal{KS}$), with parameters estimated by MoLC, within the eight models in the dictionary $\mathcal{D}$ (see Table 17.3). For the considered images the accuracy improvement granted by EDSEM is significant ($\Delta\mathcal{KS}$ up to 0.05) and the employed goodness-of-fit test suggest the obtained estimates to be very accurate statistically.

[3]RADARSAT is an official mark of the Canadian Space Agency



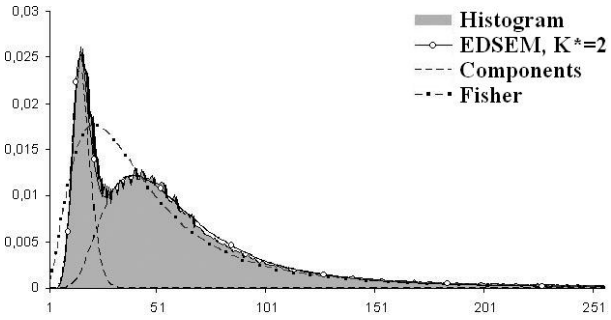
(a) CSK1 image (HH pol)



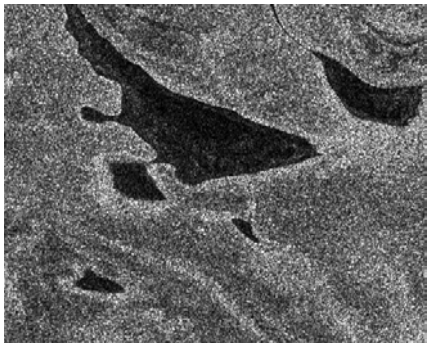
(b) CSK1 histograms



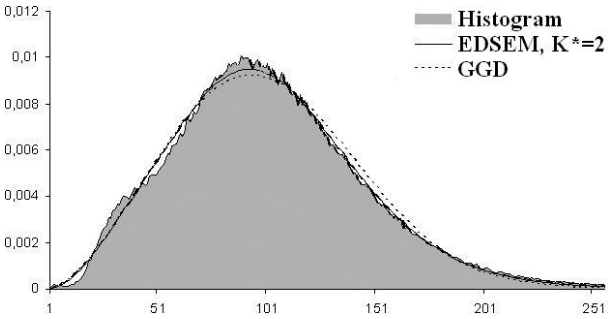
(c) RS1 image (HV pol)



(d) RS1 histograms



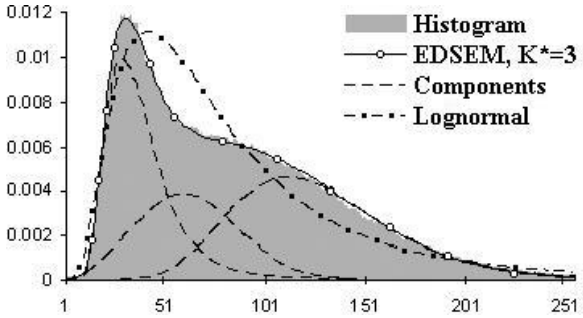
(e) TSX1 image (VV pol)



(f) TSX1 histograms



(g) TSX2 image (HH pol)



(h) TSX2 histograms

Figure 17.1: (a) CSK1 (©ASI), (c) RS1 (©MacDonald, Dettwiler and Associates Ltd.), (e) TSX1 and (g) TSX2 images (©Astrium) with corresponding plots of pdf estimates (b), (d), (f), (h). The plots contain: normalized image histogram, EDSEM pdf estimate with plots of K^* components in the estimated mixture, and the plot of the best fitting pdf model from $\mathcal{D}$.

Table 17.3: EDSEM results: K^* estimate, the estimated mixture (f_is as in Table 17.1), Kolmogorov-Smirnov distance $\mathcal{KS}$, p-value for the Kolmogorov-Smirnov test, the computation time t (in seconds), along with the best fitting model in $\mathcal{D}$ with $\mathcal{KS}$ and p-value

Image	EDSEM estimate					Best fit in $\mathcal{D}$		
	K^*	Mixture	$\mathcal{KS}$	p-value	t	Model	$\mathcal{KS}$	p-value
CSK1	4	f_5, f_4, f_1, f_5	0.007	1.0	192s	f_5	0.020	0.9869
RS1	2	f_6, f_3	0.010	1.0	115s	f_3	0.048	0.5341
TSX1	2	f_4, f_6	0.006	1.0	117s	f_4	0.024	0.9252
TSX2	3	f_4, f_3, f_2	0.004	1.0	148s	f_1	0.058	0.3552

Visual analysis of the EDSEM estimate plots (see Fig. 17.1) confirms an important improvement in the estimation accuracy. The computation times of the EDSEM reported in Table 17.3 were observed on an Intel Core 2 Duo 1.83GHz, 1Gb RAM, WinXP system.

We notice that, given a new model not contained in $\mathcal{D}$ and provided with corresponding MoLC-equations and solution, the process of adding such a model into the dictionary is straightforward. In the experiments of the next subsection we are employing a reduced dictionary $\overline{\mathcal{D}}$ of 4 pdfs and our experiments have demonstrated that this reduction still allows very accurate results to be obtained. Further still, the dictionary reduction, which was originally caused by the necessity to estimate cdfs, also enabled a significant computation acceleration (up to 50%).

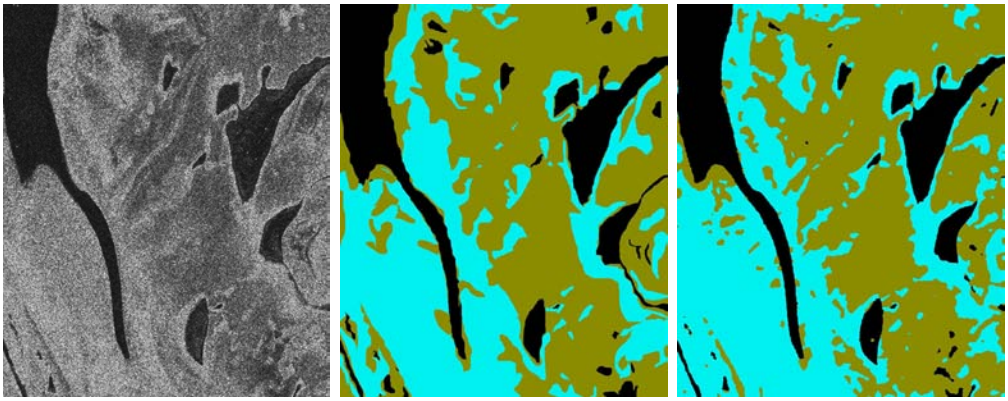
17.5.3 Joint pdf estimation and classification experiments

This subsection is devoted to experiments with PolSAR imagery. We have tested the CoDSEM method for multi-channel joint pdf estimation and the CoDSEM-MRF model for supervised PolSAR image classification. The experiments with TerraSAR-X datasets were performed within an epidemiological study and involved the classification of humid regions into $M = 3$ classes: "water", "wet" and "dry soil". Same classification was performed on COSMO-SkyMed datasets. RADARSAT-2 experiments investigated the urban area classification with $M = 3$ target classes: "water" for water surface and wet soil pixels, "urban" and "vegetation".

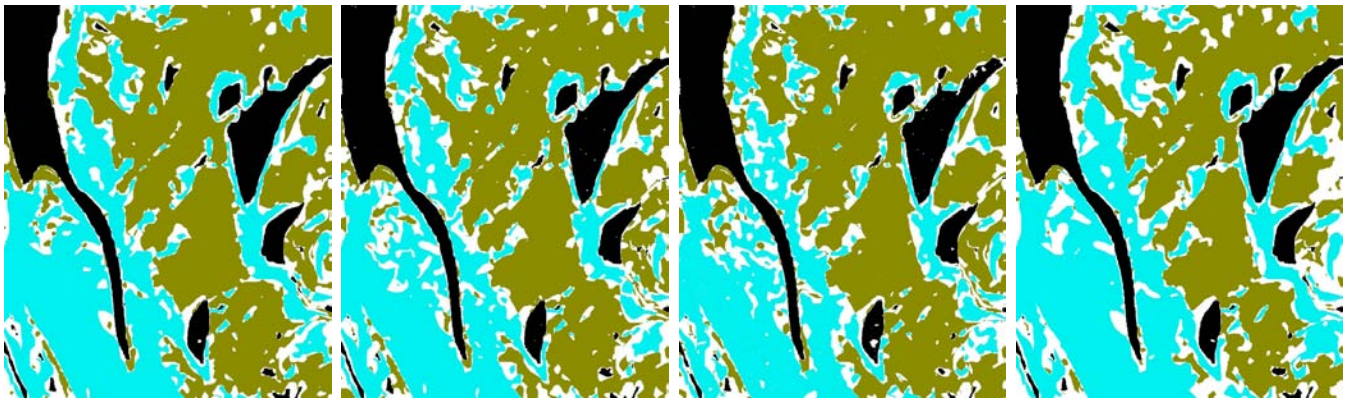
First we introduce the experimental settings that were employed. In the EDSEM-step, given the relative homogeneity of the target classes on the experiment datasets, an average value of the initial number of components $K_0 = 3$ was selected, thus, assuming mixtures with just a few components per class. In the Copula-step, the following settings in the PCS test were used. In the dual-pol case, the $[0, 1] \times [0, 1]$ square was divided into 25 equal squares $\mathcal{S}_i$ by 4 horizontal and 4 vertical lines parallel to the axes (i.e., $n = 25$). For every class $m = 1, \ldots, M$ and every cluster $\mathcal{S}_i$, $i = 1, \ldots, n$:

$$O_{im} = \sum_{(y_1,y_2)\in\omega_m} I_{\mathcal{S}_i} \left[F_1(y_1), F_2(y_2)\right],$$

where I is the indicator function and the summation is taken over all the training pixels (y_1, y_2) available for class m. For each copula $c = 1, \ldots, 10$ in $\mathcal{D}_C$, the value of E_{im} was calculated by integrating the related pdf over the square $\mathcal{S}_i$, thus retrieving the probability that a pair of $[0, 1]$-uniform random variables, whose joint distribution is defined by the c-th copula, belong to $\mathcal{S}_i$ $(i = 1, \ldots, n)$. In case of 4D quad-pol imagery, in order to pick the copula we have employed the strategy similar to the one employed for Kendall's tau estimation in Appendix A: we have used all 2D combinations of channels to calculate the values O_{im}. In the Energy minimization step, the revisit scheme on Step 3 of MMD was implemented raster-wise and the following parameter values were used : $T^0 = 5.0$, $\alpha = 0.3$, $\tau = 0.97$, $\gamma = 10^{-4}$ and n_1 was set equal to the size of the image in pixels as in [49]. Compared to ICM,



(a) TSX3 image (HH pol) (b) manual GT (c) CoDSEM-MRF map



(d) CoDSEM-MRF (e) 2DNG-MRF (f) K-NN-MRF (g) DSEM-MRF on HH

Figure 17.2: (a) HH channel of dual-pol TSX3 image (©Astrium), (b) manually created ground truth (GT), (c) CoDSEM-MRF classification map (water ■, wet soil ■, dry soil ■). Classification maps referenced to the GT: (d) CoDSEM-MRF, (e) 2DNG-MRF, (f) K-NN-MRF, and (g) DSEM-MRF on the HH pol image (correctly classified water ■, wet soil ■, dry soil ■, misclassification of all types □).

MMD generated significantly better results (above 5% of accuracy gain), starting at ML initialization (maximizing Eq. (17.3) at each pixel). SA obtained slightly better results (roughly a 1% accuracy gain) compared to MMD. However, SA generated a drastic increase in computational complexity (around 200 iterations for MMD and over 1500 for SA).

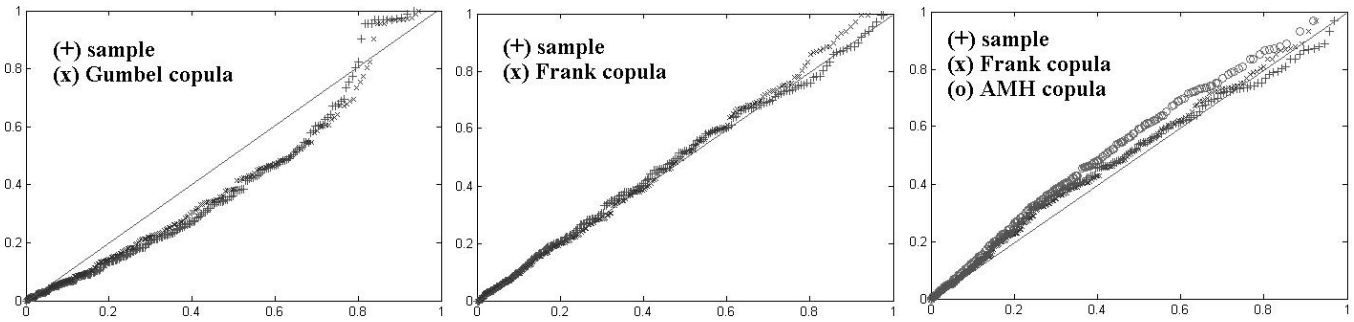
For every dataset, the CoDSEM-MRF algorithm was trained on a small 250×250 sub-image endowed with a manually annotated non-exhaustive ground truth (GT), that did not overlap with the test areas.

Table 17.4 reports the classification accuracies obtained on the test samples for each class ("class accuracy"), the resulting average accuracy (i.e., the arithmetic mean of the class accuracies) and overall accuracy (i.e., the percentage of correctly classified test samples, irrespective of their classes). The first test image TSX3 (see Fig. 17.2(a)) covers an area of a river delta in Sanchagang, China. For this experiment, an exhaustive ground truth was manually created, and the classification results were referenced to it in order to calculate the accuracies and demonstrate the results visually. The automatically selected copulas were: Gumbel for the "water", and Frank for the "wet soil" and "dry soil" classes. The MRF parameter estimation provided $\beta^* = 1.408$ and the overall classification accuracy achieved by CoDSEM-MRF was 84.5%. The resulting CoDSEM-MRF classification map is shown in Fig. 17.2(c) and referenced to the GT in Fig. 17.2(d).

In order to appreciate the classification gain from dual-pol data (HH/VV), compared to single-channel data, we performed an experiment of DSEM-MRF classification on the HH channel, which reported an overall accuracy of 80.6%. On average, the overall accuracy gain from adding the second

Table 17.4: Classification accuracies on the considered test images.

Image	Method	Water	Wet soil	Dry soil	Average	Overall
TSX3	CoDSEM-MRF	90.02%	82.56%	84.80%	85.79%	**84.55%**
	2DNG-MRF	89.33%	82.12%	76.51%	82.78%	**81.49%**
	K-NN-MRF	87.15%	87.81%	71.45%	82.14%	**81.95%**
	DSEM-MRF on HH pol	90.00%	69.93%	91.28%	83.74%	**80.61%**
TSX4	CoDSEM-MRF, $\beta^* = 1.324$	92.48%	94.59%	85.16%	90.49%	**92.41%**
	CoDSEM-MRF, $\beta = 0.5$	91.52%	89.31%	77.48%	86.11%	**87.98%**
	2DNG-MRF	92.59%	89.33%	86.01%	89.31%	**89.74%**
	K-NN-MRF	90.21%	98.56%	78.91%	89.23%	**92.61%**
CSK2	DSEM-MRF	95.57%	92.22%	92.45%	93.41%	**93.14%**
	$\mathcal{K}$-root-MRF	89.64%	87.78%	96.83%	91.53%	**90.94%**
		Water	Vegetation	Urban		
RS2	CoDSEM-MRF	98.04%	90.33%	71.49%	86.61%	**88.69%**
	K-NN-MRF	98.85%	89.02%	62.89%	83.59%	**86.42%**
RS3	CoDSEM-MRF	97.30%	95.12%	86.12%	92.85%	**94.18%**
	K-NN-MRF	98.10%	96.06%	81.33%	91.83%	**93.51%**



(a) Water, $\hat{\tau} = 0.359$ (b) Wet soil, $\hat{\tau} = 0.393$ (c) Dry soil, $\hat{\tau} = 0.265$

Figure 17.3: K-plots display the graphical goodness-of-fit for copula models on the learning stage for TSX3 image. For the considered three classes the sample HH-VV dependences (+) are fitted by automatically selected copulas (×). The K-plot for the Ali-Mikhail-Haq (AMH) copula (o) is also presented for the dry soil class (c). The diagonal (–) represents the independency scenario.

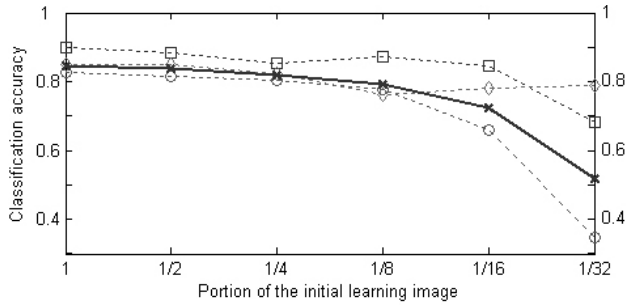


Figure 17.4: The sensitivity of CoDSEM-MRF classification accuracy to the size of the learning image. The class accuracies (water □, wet soil ∘, dry soil ♢) and overall accuracies (solid line **x**) are reported for TSX3 image. The initial learning image is 250 × 250 pixels.

polarization channel on the considered datasets was 3 to 7%.

For the sake of comparison, we also provide classification maps obtained by the benchmark 2D Nakagami-Gamma (2DNG) model [29] (Fig. 17.2(e)) and K-nearest neighbors (K-NN) method [45] (Fig. 17.2(f)). In order to perform a fair comparison among contextual methods, the above models were combined with the contextual MRF approach. The same MRF parameter estimation and energy minimization procedures as in CoDSEM-MRF were employed in these benchmark experiments as well. For the 2DNG model, the equivalent number of looks was set to $L = 2.66$, and for the K-NN model, $K^* = 40$ was estimated by cross-validation [45]. The overall accuracies of these two benchmark approaches were 81.5% for 2DNG-MRF and 81.9% for K-NN-MRF on the dual-pol HH/VV TSX3 image. These accuracies were lower than the overall classification accuracy achieved by CoDSEM-MRF. K-NN-MRF reported higher accuracy for the "wet soil" class, however this was achieved at the expense of accuracy for the "dry soil."

To demonstrate the accuracy of the Copula-DSEM joint pdf model directly we present K-plots of the sample HH-VV dependence and the dependence estimated by copulas (see Fig. 17.3) on the learning image for TSX3. A K-plot (from Kendall-plot) [52] is a rank based graphical tool developed for visualization of dependence structure between two random variables. In our case, we build a sample K-plot as well as K-plots corresponding to copula models. One can see a good agreement between observations and the automatically selected models.

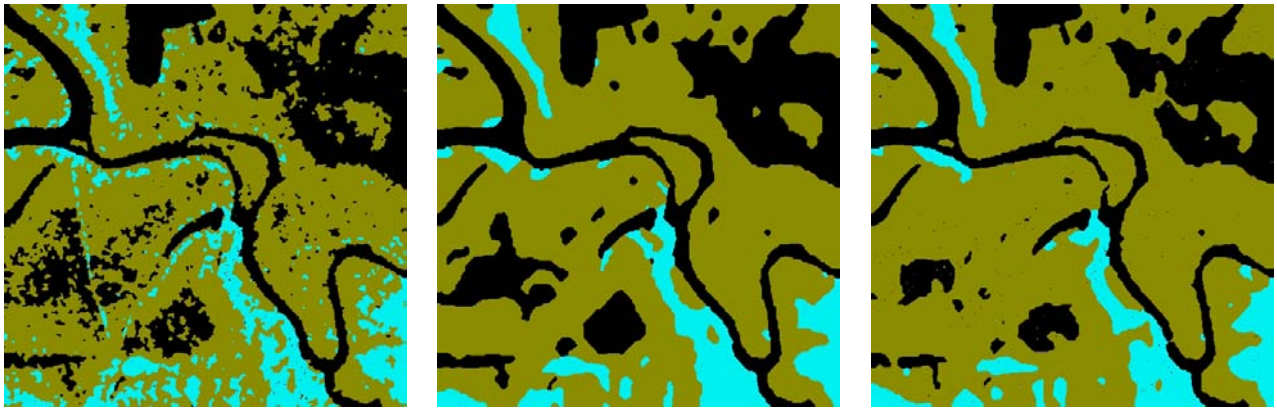
We now briefly compare the developed CoDSEM-MRF approach with the previously proposed copula-based method [32]. Firstly, the approach [32] is based on the use of the Ali-Mikhail-Haq (AMH) copula [34], which can model dependencies corresponding to Kendall's correlation coefficient $\tau \in [-0.1817, 0.3333]$. On the employed TerraSAR-X dataset we have observed empirical values $\hat{\tau}_{\text{water}} = 0.359, \hat{\tau}_{\text{wet}} = 0.393, \hat{\tau}_{\text{dry}} = 0.265$. Therefore, the AMH copula can be used only for the dry soil class and its K-plot is presented in Fig. 17.3(c). However, even in this case, the goodness-of-fit provided by the automatically selected Frank copula is better. Thus, the use of dictionary-based copula selection approach is more accurate and constitutes a more flexible model with a wider range of applicability. Secondly, the use of the EDSEM finite mixture estimation approach for marginal pdf estimation provides higher accuracy estimates than singular pdf models, whose use was suggested in [32], especially for HR SAR images.

It is well known that the acquisition of the training sets required for supervised classification is a very costly procedure. Therefore, the classification models are designed to be as robust as possible with respect to small learning sets. To evaluate this characteristic of the proposed algorithm, we present an experimental study (see Fig. 17.4) of the classification accuracy as a function of the size of the employed training set. We start with the learning image of 250×250 pixels, and then repeatedly reduce its size by a half at each iteration, i. e. by discarding 50% randomly selected pixels the learning set of each thematic class, until 1/32 of the initial learning image is kept. At each iteration we evaluate the class and overall accuracies of classification on the TSX3 image. To make the accuracy estimates more consistent, the whole process has been repeated three times and the averaged results are presented in Fig. 17.4. We observe that the algorithm behaves robustly till 1/8 (90×90 pixels) of the initial learning image is kept. When very strong subsampling rates ($\geqslant 16$) are applied, the wet area classification rate drops significantly. This is due to the fact that the histogram of wet soil lies "between" those of water and dry soil. Thus, when the training set becomes unrepresentatively small, a lot of wet soil pixels are misclassified as dry soil. This result suggests the capability of the developed algorithm to perform fairly good and consistently on small training sets, i.e., up to 100×100 training images for the considered dataset.

The most time consuming stages of the algorithm are the MRF parameter estimation and energy minimization steps. Given the fairly low value $K_0 = 3$, the EDSEM method was very fast. The



(a) TSX4 image (HH pol) (b) nonexhaustive GT map (c) 2DNG-MRF map



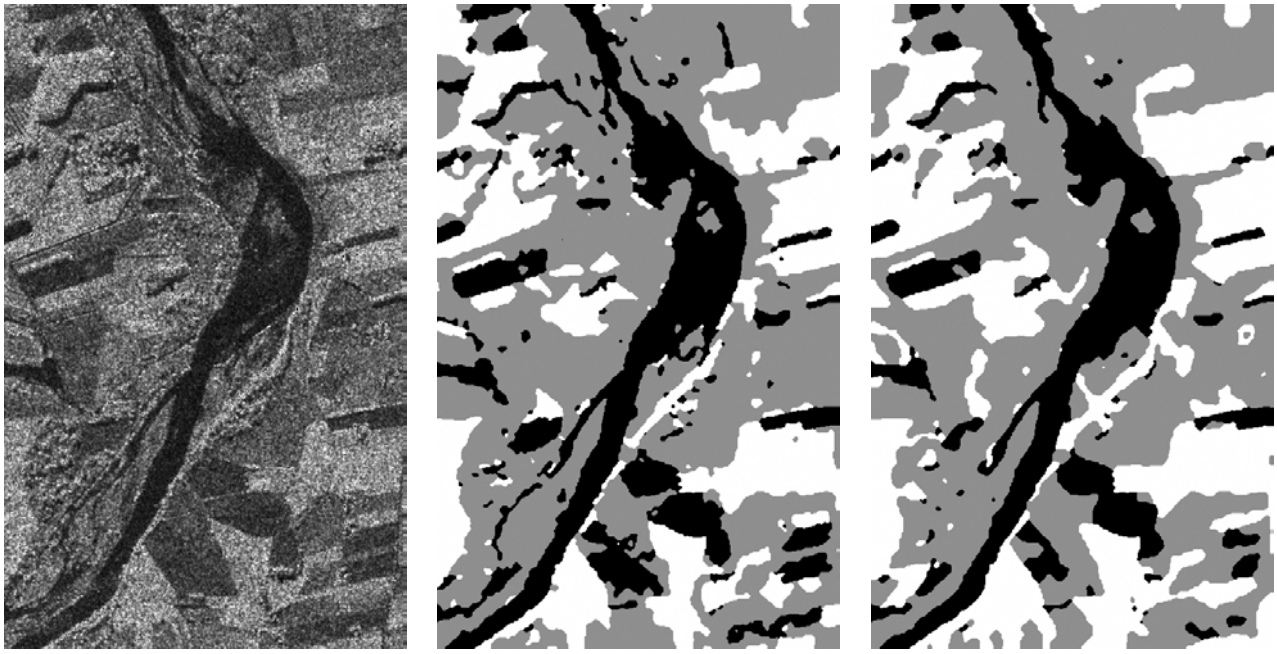
(d) CoDSEM-MRF, $\beta = 0.5$ (e) CoDSEM-MRF, $\beta^* = 1.324$ (f) K-NN-MRF map

Figure 17.5: (a) HH channel of the dual-pol TSX4 image (©Astrium), (b) nonexhaustive ground truth (GT) map (water ■, wet soil ■, dry soil ■, outside GT □). Classification maps: (c) 2DNG-MRF, CoDSEM-MRF with (d) manually set $\beta = 0.5$ and (e) automatically estimated $\beta^* = 1.324$, (f) K-NN-MRF (same colors).

experiments were conducted on a Core 2 Duo 1.83GHz, 1Gb RAM, WinXP system. With the number of iterations equal to 200, the β-estimation on a roughly 1000×1000 image took around 80 seconds. The average of 200 iterations required for convergence with MMD (Energy minimization step) took about 100 seconds. Thus about 200 seconds were required for the complete classification process of a 1000×1000 dual-pol image with three classes.

The second test image TSX4 (see Fig. 17.5) originates from the same TerraSAR-X dataset. Therefore, the same learning image has been employed, and, consequently, the same set of copulas were selected. The overall classification accuracy obtained by CoDSEM-MRF was 92.4% with $\beta^* = 1.324$. This accuracy is higher than that reported for the TSX3 image. However, this increase is mostly due to the use of a non-exhaustive test map in this experiment. This map did not contain many class transition regions where misclassified pixels may be visually spotted. Once again, we provide comparisons with the benchmark 2DNG-MRF (Fig. 17.5(c)) and K-NN-MRF (Fig. 17.5(f)) classification approaches. Consistently with our earlier observations, here CoDSEM-MRF outperformed appreciably 2DNG-MRF (89.7% of overall accuracy). K-NN-MRF with 92.6%, on the other hand, reported about the same level of overall accuracy as CoDSEM-MRF. Here we come to the same conclusion as with the TSX3 image: K-NN-MRF performs better on "wet soil" and far worse on "dry soil". We notice also that the average accuracy of K-NN-MRF is inferior to that of CoDSEM-MRF on both TSX3 and TSX4.

A slight oversmoothing effect can be noticed in the results obtained by CoDSEM-MRF on TSX4, see Fig. 17.5(e). Therefore, we also show the classification map obtained by manually setting a smaller



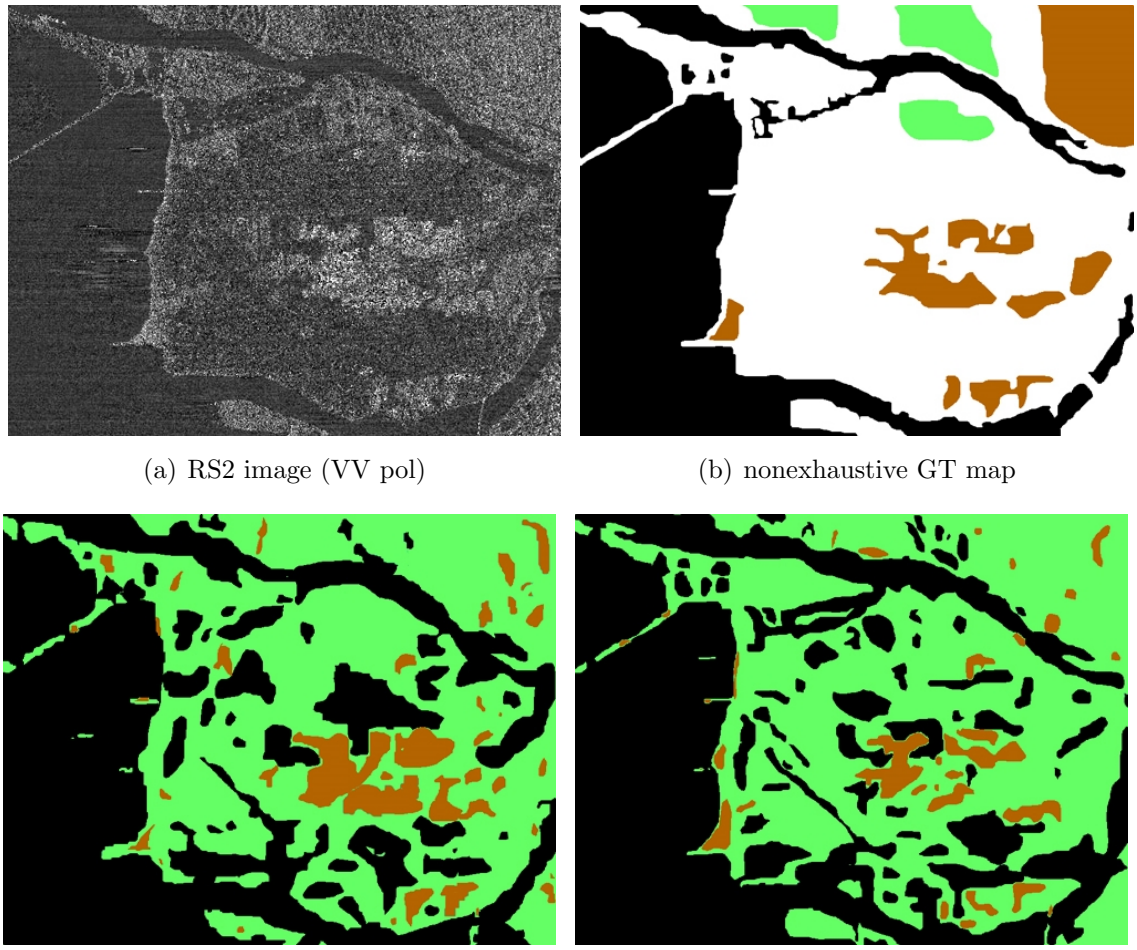
(a) CSK2 image (b) DSEM-MRF map (c) $\mathcal{K}$-root-MRF map

Figure 17.6: (a) HH polarization CSK2 image (©ASI), and classification maps: (b) DSEM-MRF and (c) $\mathcal{K}$-root-MRF (water ■, wet soil ■, dry soil □).

value of the MRF parameter $\beta = 0.5$ (Fig. 17.5(d)). One can see that the the spatial details are more precise at the expense of a noisier segmentation. The classification accuracies achieved with $\beta = 0.5$ (see Table 17.4) are inferior to those obtained in the $\beta^* = 1.324$ case. This suggests that, at least for this dataset, stronger regularization is preferable.

We now move to the experiments with COSMO-SkyMed imagery. Here the proposed classification approach was tested on a single-pol CSK2 image (see Fig. 17.6(a)). A non-exhaustive GT was employed. The overall classification accuracy of DSEM-MRF (Fig. 17.6(b)) was 93.1% with $\beta^* = 1.566$. We compare this approach with the $\mathcal{K}$-root-MRF contextual classification approach that is based on a $\mathcal{K}$-root model [9] for each class-conditional statistics. The $\mathcal{K}$ distribution is a well-known model for a possibly textured SAR multilook single-channel intensity and $\mathcal{K}$-root is the corresponding amplitude parametric pdf. The overall classification accuracy reported in this experiment (Fig. 17.6(c)) was equal to 90.9%. The improved performance of DSEM-MRF is due to the more accurate pdf estimates generated by DSEM than by the $\mathcal{K}$-root model. Notice here, that in case of single-channel SAR we do not use the K-NN-MRF classifier, as in case of one dimensional discrete feature space (i.e. with amplitude values $1, \ldots, 256$) and target classes with close histograms the use of K-NN classifier becomes impossible. More specifically, if we consider the distance zero or one in feature space, we already have hundreds of neighbors of which we then need to somehow select K^* "voting" neighbors.

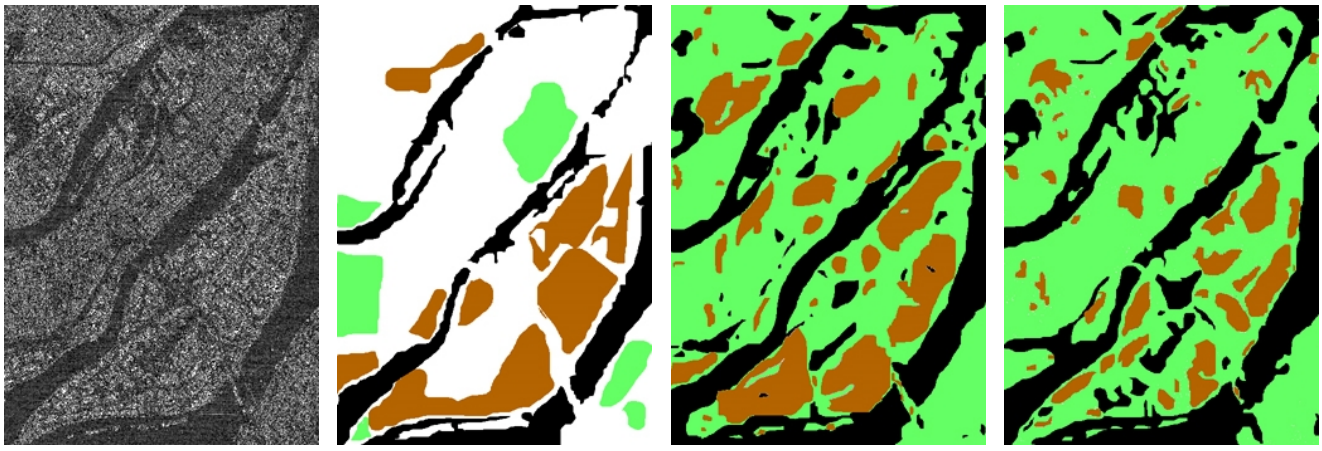
Finally, we consider the quad-pol RADARSAT-2 imagery (see Fig. 17.7 and Fig. 17.8) classification into the following cases: "water", "vegetation" and "urban". We employ a non-exhaustive ground truth map and compare the result with the K-NN-MRF classification, with $K^* = 35$ estimated by cross-validation [45] and $\beta = 1.419$. We remind again that only the three copula families from $\mathcal{D}_C$ were considered, and the following were automatically selected: Gumbel for "water", and Frank for "vegetation" and "urban" classes. We stress here that the comparison with the Nakagami-Gamma model is not feasible, because this model has not been defined for dimensionality higher than two [29]. In these experiments we have observed about the same level of classification accuracy for both CoDSEM-MRF and K-NN-MRF on "water" and "vegetation" classes, and an appreciable increase of accuracy for the "urban" class in case of CoDSEM-MRF classification. Therefore, the higher overall and average



(a) RS2 image (VV pol) (b) nonexhaustive GT map

(c) CoDSEM-MRF map (d) K-NN-MRF map

Figure 17.7: (a) VV channel of the quad-pol RS2 image (RADARSAT-2 Data and Products ©MacDonald, Dettwiler and Associates Ltd., 2008 - All Rights Reserved), (b) nonexhaustive ground truth (GT) map (water ■, urban ■, vegetation ■, outside GT □). Classification maps: (c) CoDSEM-MRF, (d) K-NN-MRF (same colors).



(a) RS3 image (HV pol) (b) nonexhaustive GT map (c) CoDSEM-MRF map (d) K-NN-MRF map

Figure 17.8: (a) HV channel of the quad-pol RS3 image (RADARSAT-2 Data and Products ©MacDonald, Dettwiler and Associates Ltd., 2008 - All Rights Reserved), (b) nonexhaustive ground truth (GT) map (water ■, urban ■, vegetation ■, outside GT □). Classification maps: (c) CoDSEM-MRF, (d) K-NN-MRF (same colors).

accuracies demonstrated by the CoDSEM-MRF method in these quad-pol RADARSAT-2 experiments suggest the developed algorithm to be a competitive and efficient supervised classification approach. Moreover, from the methodological point of view, CoDSEM might be preferable to K-NN as it provides an explicit description of a statistical model for the data, whereas the latter operates as a "black box".

17.6 Conclusions

In this chapter we have considered the issues of amplitude PolSAR pdf estimation and supervised classification in the framework of this high resolution satellite imagery. We have proposed a general flexible pdf-estimation method for single channel SAR images and specifically validated it with heterogeneous HR SAR data, which represent a very up-to-date and relevant case of SAR imagery. The developed model is based on the dictionary-based stochastic expectation maximization (DSEM) approach recently developed in [11] for medium resolution SAR. The developed Enhanced DSEM (EDSEM) extended DSEM to the novel type of imagery (heterogenous HR), enhanced by a novel efficient procedure for estimating the number of mixture components, reported very accurate and computationally fast estimation results in experiments with HR SAR images acquired by the aforementioned modern satellite systems. We stress here that the problem of modeling the statistics of HR satellite SAR amplitude images has not been satisfactorily addressed so far, and the proposed EDSEM technique looks very promising for this type of imagery, since it is based on a finite mixture approach, and intrinsically takes into account heterogeneity, which is an inherent HR image property. The obtained results suggest EDSEM to be an attractive approach for various application problems, e.g., it can be efficiently used for SAR image segmentation (to discriminate the resulting mixture components) [11] and to class-conditional pdf modeling in supervised HR SAR classification. The extension of the proposed method to intensity data (instead of amplitudes) is straightforward and can be performed by replacing the amplitude pdfs in the dictionary with the corresponding intensity pdfs.

We have introduced a novel model for the joint pdf of the amplitude PolSAR imagery. Based on flexible statistical tools, such as, copulas and EDSEM estimation technique, this approach proved to be accurate, easy to implement and fast from the parameter-estimation point of view. It enables a new level of flexibility in modeling the joint distribution from marginal single-channel distributions, which results in higher classification accuracies. The experiments with several datasets of HR SAR imagery have demonstrated the high applicability and accuracy of this model.

Based on the proposed joint pdf model we have developed a supervised classification algorithm for single channel and PolSAR satellite imagery. Specifically, it combines the Markov random field approach to Bayesian image classification, finite mixture technique for probability density function estimation, and copulas for multivariate pdf modeling. These three statistical concepts ensure high flexibility and applicability of the developed method to HR PolSAR image classification. Structurally, the proposed classification algorithm can be described as supervised and semiautomatic (a few EDSEM and MMD parameters have to be specified). The accuracy of the proposed algorithm was validated on classification into 3 target classes on several HR satellite SAR images: a single-pol COSMO-SkyMed image, a dual-pol TerraSAR-X image and a quad-pol RADARSAT-2 image. The experiments demonstrated a high level of accuracy on the experimental datasets and outperformed several parametric and nonparametric contextual benchmark algorithms.

Finally, as far as the future extensions of this work are concerned, a promising direction is to explore the use of more efficient optimization approaches. One of the best candidates to replace MMD is an appropriate graph cut based approach [53]. Such approaches yield very good approximations in the MAP segmentation problem and are known to be fast. Another direction of development lies in the specialization of this model to urban area classification, which is an important and relevant application of

SAR classification. To this end, the contextual MRF model would also need to incorporate geometrical information.

Appendix A. Copula theory

A *D-dimensional copula* is a function $C : [0,1]^D \to [0,1]$, which satisfies:

1. $C(\mathbf{x}) = 0$ for any $\mathbf{x}$: $\exists i \in \{1, \ldots, D\}$ so that $x_i = 0$.
2. $C(\mathbf{x}) = x_d$ for any $\mathbf{x}$: $x_i = 1$, for all $i \neq d$.
3. the *D-increasing* condition: for any $0 \leqslant x_{i,1} \leqslant x_{i,2} \leqslant 1$, where $i = 1, \ldots, D$,

$$V_C \equiv \sum_{i_1=1}^{2} \cdots \sum_{i_D=1}^{2} (-1)^{(i_1+\cdots+i_D)} C(x_{1,i_1}, \ldots, x_{D,i_D}) \geqslant 0.$$

The important property of copulas is given by the *Sklar's theorem* [34], which states the existence of a copula C, that models the joint distribution H of arbitrary random variables $X_1, \ldots, X_D$ with cdfs $F_1, \ldots, F_D$:

$$H(\mathbf{x}) = C(F_1(x_1), \ldots, F_D(x_D)), \forall \mathbf{x} \in \mathbb{R}^D. \tag{17.6}$$

Given absolutely continuous random variables with pdfs $f_1(x), \ldots, f_D(x)$ with corresponding cdfs $F_1(x), \ldots, F_D(x)$, the pdf of the joint pdf $h(\mathbf{x}), x \in \mathbb{R}^D$ corresponding to (17.6) is given by:

$$h(\mathbf{x}) = f_1(x_1) \cdots f_D(x_D) \frac{\partial^D C}{\partial x_1 \cdots \partial x_D}(F_1(x_1), \ldots, F_D(x_D)). \tag{17.7}$$

where $\frac{\partial^D C}{\partial x_1 \cdots \partial x_D}(F_1(x_1), \ldots, F_D(x_D))$ is the pdf corresponding to copula $C(\mathbf{x})$, provided this derivatie exists in $[0,1]^D$.

An important family of copulas are Archimedean copulas, which have a simple analytical form and yet provide a wide variety of modeled dependence structures. An *Archimedean* copula is a copula C, defined as:

$$C(\mathbf{x}) = \phi^{-1}(\phi(x_1) + \cdots + \phi(x_D)), \tag{17.8}$$

where the *generator function* $\phi(u)$ is a function satisfying the following properties:

1. $\phi(u)$ is continuous on $[0,1]$;
2. $\phi(u)$ is decreasing, $\phi(1) = 0$;
3. $\phi(u)$ is convex,

and $\phi^{-1}(u)$ denotes the inverse of $\phi(u)$.

A common way to perform copula parameter estimation is by using its connection with Kendall's tau, which is a ranking correlation coefficient [34]. *Kendall's tau* is a concordance-discordance measure between two independent realizations (X, Y) and $(\hat{X}, \hat{Y})$ from the same cdf $H(x, y)$ defined as:

$$\tau = \text{Prob}\{(X - \hat{X})(Y - \hat{Y}) > 0\} \; - \text{Prob}\{(X - \hat{X})(Y - \hat{Y}) < 0\}.$$

Given realizations (x_l, y_l), $l = 1, \ldots, N$, the empirical estimator of Kendall's tau is given by:

$$\hat{\tau} = \frac{4}{N(N-1)} \sum_{i \neq j} I[x_i \leqslant x_j] I[y_i \leqslant y_j] - 1, \tag{17.9}$$

where $I[\cdot]$ denotes the indicator function.

In case of $D = 2$, i.e. *bivariate* copulas, by integrating in the definition of τ over the distribution of $(\hat{X}, \hat{Y})$, we get the general relationship between Kendall's τ and the bivariate copula C associated with $H(x_1, x_2)$, expressed by the following Lebesgue-Stieltjes integral:

$$\tau + 1 = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v), \tag{17.10}$$

In the specific case of bivariate Archimedean copulas, the relationship is expressed in terms of generator function $\phi(t)$:

$$\tau = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt. \tag{17.11}$$

In the general multivariate case ($D \geqslant 3$), (17.10) takes the same form, where $C(u, v)$ is a bivariate copula of the same type[4] and $\tau = \overline{\tau}$ is the average of $D(D-1)/2$ consistent Kendall's τ estimates (17.9) corresponding to all combinations of bivariate marginals (X_{d_1}, X_{d_2}), $0 \leqslant d_1 < d_2 \leqslant D$, see [54]. Equation (17.11) holds with the same modification $\tau = \overline{\tau}$.

The use of Kendall's tau based parameter estimation, as compared to other copula parameter estimation strategies, is motivated by a relatively small number of samples N available in our experiments [54].

Acknowledgments

This research was a collaborative effort between the Institut National de Recherche en Informatique et en Automatique (INRIA) Sophia Antipolis - Méditerranée Center, France, and the Dept. of Biophysical and Electronic Engineering (DIBE) of the University of Genoa, Italy. The work was carried out with the partial financial support of the INRIA, France, and the French Space Agency (CNES). This support is gratefully acknowledged. The authors would like to thank the Italian Space Agency for providing the COSMO-SkyMed (CSK®) image of Piemonte (COSMO-SkyMed Product - ©ASI - Agenzia Spaziale Italiana - 2008. All Rights Reserved). The TerraSAR-X image of Sanchagang (©Astrium GEO-Information Services, 2008) was taken at *http://www.infoterra.de/*, and the RADARSAT-2 image of Vancouver (©MacDonald, Dettwiler and Associates Ltd., 2008) – at *http://www.radarsat2.info/*.

[4]Recall that by definition (17.8) an Archimedean copula is defined solely by its generator function and, as such, can be defined for any dimensionality factor D.

Bibliography

[1] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, 2nd ed. NC, USA: SciTech, Raleigh, 2004.

[2] C. Tison, J.-M. Nicolas, F. Tupin, and H. Maitre, "A new statistical model for Markovian classification of urban areas in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2046–2057, 2004.

[3] A. Banerjee, P. Burlina, and R. Chellappa, "Adaptive target detection in foliage-penetrating SAR images using alpha-stable models," *IEEE Trans. Image Process.*, vol. 8, no. 12, pp. 1823–1831, 1999.

[4] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification.* New York: Wiley Interscience, 2001.

[5] P. Mantero, G. Moser, and S. B. Serpico, "Partially supervised classification of remote sensing images using SVM-based probability density estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 559–570, 2005.

[6] H.-C. Li, W. Hong, Y.-R. Wu, and P.-Z. Fan, "An efficient and flexible statistical model based on generalized gamma distribution for amplitude SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2711–2722, 2010.

[7] G. Moser, J. Zerubia, and S. B. Serpico, "SAR amplitude probability density function estimation based on a generalized Gaussian model," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1429–1442, 2006.

[8] E. E. Kuruoglu and J. Zerubia, "Modelling SAR images with a generalization of the Rayleigh distribution," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 527–533, 2004.

[9] E. Jakeman and P. N. Pusey, "A model for non-Rayleigh sea echo," *IEEE Trans. Antennas Propagat.*, vol. 24, pp. 806–814, 1976.

[10] A. C. Frery, H.-J. Muller, C. C. F. Yanasse, and S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, 1997.

[11] G. Moser, S. Serpico, and J. Zerubia, "Dictionary-based Stochastic Expectation Maximization for SAR amplitude probability density function estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 1, pp. 188–199, 2006.

[12] G. Celeux, D. Chauveau, and J. Diebolt, "Stochastic versions of the EM algorithm: an experimental study in the mixture case," *J. Statist. Comp. Sim.*, vol. 55, no. 4, pp. 287–314, 1996.

[13] J.-S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications.* New York: CRC Press, 2009.

[14] J.-S. Lee, M. R. Grunes, and R. Kwok, "Classification of multilook polarimetric SAR imagery based on complex Wishart distribution," *Int. J. Remote Sensing*, vol. 15, no. 11, pp. 2299–2311, 1994.

[15] J.-M. Beaulieu and R. Touzi, "Segmentation of textured polarimetric SAR scenes by likelihood approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2063–2072, 2004.

[16] A. C. Frery, C. C. Freitas, and A. H. Correia, "Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3098–3109, 2007.

[17] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Classification with a non-Gaussian model for PolSAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2999–3009, 2008.

[18] L. Bombrun, G. Vasile, M. Gay, and F. Totir, "Hierarchical segmentation of polarimetric SAR images using heterogeneous clutter models," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 726–737, 2011.

[19] Y. Ito and S. Omatu, "Polarimetric SAR data classification using competitive neural networks," *Int. J. Remote Sensing*, vol. 19, no. 14, pp. 2665–2684, 1998.

[20] K. S. Chen, W. P. Huang, D. H. Tsay, and F. Amar, "Classification of multifrequency polarimetric SAR imagery using a dynamic learning neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 3, pp. 814–820, 1996.

[21] C. Lardeux, P.-L. Frison, C. Tison, J.-C. Souyris, B. Stoll, B. Fruneau, and J.-P. Rudant, "Support vector machine for multifrequency SAR polarimetric data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 4143–4152, 2009.

[22] C. T. Chen, K. S. Chen, and J.-S. Lee, "The use of fully polarimetric information for the fuzzy neural classification of SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 2089–2100, 2003.

[23] P. R. Kersten, J.-S. Lee, and T. L. Ainsworth, "Unsupervised classification of polarimetric synthetic aperture radar images using fuzzy clustering and EM clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 519–527, 2005.

[24] J. Morio, F. Goudail, X. Dupuis, P. Dubois-Fernandez, and P. Refregier, "Polarimetric and interferometric SAR image partition into statistically homogeneous regions based on the minimization of the stochastic complexity," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3599–3609, 2007.

[25] K. Ersahin, I. Cumming, and R. Ward, "Segmentation and classification of polarimetric SAR data using spectral graph partitioning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 164–174, 2010.

[26] G. D. De Grandi, J. S. Lee, and D. L. Schuler, "Target detection and texture segmentation in polarimetric SAR images using a wavelet frame: Theoretical aspects," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3437–3453, 2007.

[27] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, 1997.

[28] J.-S. Lee, M. R. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 722–731, 2004.

[29] J.-S. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller, "Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1017–1028, 1994.

[30] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak, "K-distribution and polarimetric terrain radar clutter," *J. Electromagn. Waves Applicat.*, vol. 3, pp. 747–768, 1989.

[31] C. C. Freitas, A. C. Frery, and A. H. Correia, "The polarimetric G distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, 2005.

[32] G. Mercier and P.-L. Frison, "Statistical characterization of the Sinclair matrix: Application to polarimetric image segmentation," in *Proceedings of IGARSS*, Cape Town, South Africa, 2009, pp. III–717–III–720.

[33] Z. Kato, J. Zerubia, and M. Berthod, "Unsupervised parallel image classification using Markovian models," *Pattern Recognit.*, vol. 32, no. 4, pp. 591–604, 1999.

[34] R. B. Nelsen, *An Introduction to Copulas*, 2nd ed. New-York: Springer, 2006.

[35] G. Mercier, S. Derrode, W. Pieczynski, J.-M. Nicolas, A. Joannic-Chardin, and J. Inglada, "Copula-based stochastic kernels for abrupt change detection," in *Proceedings of IGARSS*, Denver, USA, 2006, pp. 204–207.

[36] G. Mercier, G. Moser, and S. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428–1441, 2008.

[37] F. Galland, J.-M. Nicolas, H. Sportouche, M. Roche, F. Tupin, and P. Refregier, "Unsupervised synthetic aperture radar image segmentation using Fisher distributions," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2966–2972, 2009.

[38] K. Song, "Globally convergent algorithms for estimating generalized gamma distributions in fast signal and image processing," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1233–1250, 2008.

[39] M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions*. New York: Dover, 1964.

[40] V. Krylov, G. Moser, S. Serpico, and J. Zerubia, "Dictionary-based probability density function estimation for high-resolution SAR data," in *Proceedings of SPIE*, vol. 7246, San Jose, USA, 2009, p. 72460S.

[41] G. Frahm, M. Junker, and A. Szimayer, "Elliptical copulas: Applicability and limitations," *Stat. Prob. Lett.*, vol. 63, pp. 275–286, 2003.

[42] D. Huard, G. Évin, and A.-C. Favre, "Bayesian copula selection," *Comput. Statist. Data Anal.*, vol. 51, no. 2, pp. 809–822, 2006.

[43] E. Lehmann and J. Romano, *Testing statistical hypotheses*, 3rd ed. New York: Springer, 2005.

[44] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Stat. Soc. B*, vol. 48, pp. 259–302, 1986.

[45] C. M. Bishop, *Pattern Recognition and Machine Learning*. New-York: Springer, 2006.

[46] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 6, pp. 721–741, 1984.

[47] P. van Laarhoven and E. Aarts, *Simulated annealing: theory and applications*. Dordrecht: Kluwer, 1992.

[48] W. Hastings, "Monte Carlo sampling method using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.

[49] M. Berthod, Z. Kato, S. Yu, and J. Zerubia, "Bayesian image classification using Markov random fields," *Image and Vision Computing*, vol. 14, pp. 285–295, 1996.

[50] W. Michiels, E. Aarts, and J. Korst, *Theoretical Aspects of Local Search*. Berlin: Springer, 2007.

[51] M. Stephens, "Test of fit for the logistic distribution based on the empirical distribution function," *Biometrika*, vol. 66, pp. 591–595, 1979.

[52] C. Genest and J.-C. Boies, "Detecting dependence with Kendall plots," *Amer. Statistician*, vol. 57, no. 4, pp. 275–284, 2003.

[53] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222 –1239, 2001.

[54] I. Kojadinovic and J. Yan, "Comparison of three semiparametric methods for estimating dependence parameters in copula models," *Insurance: Mathematics and Economics*, vol. 47, pp. 52–63, 2010.