Modeling the statistics of high resolution SAR images

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

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Modeling the statistics of high resolution SAR images

Vladimir Krylov — Gabriele Moser — Sebastiano B. Serpico — Josiane Zerubia

N° 6722

November 2008

Thème COG

ISSN 0249-6399 ISRN INRIA/RR--6722--FR+ENG

Modeling the statistics of high resolution SAR images

Vladimir Krylov* , Gabriele Moser† , Sebastiano B. Serpico‡ , Josiane Zerubia§

Thème COG — Systèmes cognitifs
Projet Ariana

Rapport de recherche n° 6722 — November 2008 — 43 pages

Abstract: In the context of synthetic aperture radar (SAR) image processing a crucial problem is represented by the need to develop accurate models for the statistics of the pixel intensities. In the current research report, we address the problem of parametric probability density function (pdf) estimation for modeling the amplitude distribution for high resolution SAR images. Hitherto, several theoretical and heuristic models for the pdfs of SAR data have been proposed in the literature, most of them being highly effective for some particular land-cover typologies. Thus, given some SAR image with no prior information about the typology, the choice of a single optimal SAR parametric pdf becomes a hard task. In this report, we develop an estimation algorithm addressing the problem of pdf selection by adopting a finite mixture model (FMM) for the amplitude pdf, by mixing components belonging to a given dictionary of SAR-specific pdfs. The proposed method automatically integrates the procedures of selection of the optimal model for each component, of parameter estimation, along with the optimization of the number of components, by combining the Stochastic Expectation Maximization (SEM) iterative methodology and the recently proposed "method-of-log-cumulants" (MoLC) for parametric pdf estimation for non-negative random variables. Experimental results on several real COSMO-SkyMed and RAMSES sensor images are presented, showing the capabilities of the proposed method to accurately model the statistics of SAR amplitude data.

Key-words: synthetic aperture radar (SAR) image, probability density function (pdf), parametric estimation, finite mixture models, stochastic expectation maximization (SEM).

* EPI Ariana, UR INRIA Sophia Antipolis Méditeranée, 2004, Route des Lucioles, B.P.93, FR-06902, Sophia Antipolis Cedex (France); Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, 119991 Leninskie Gory, Moscow (Russia), e-mail: vl_krylov@mail.ru.

† Dept. of Biophysical and Electronic Engineering (DIBE), University of Genoa, Via Opera Pia 11a, I-16145, Genoa (Italy), e-mail: gemini@dibe.unige.it.

‡ Dept. of Biophysical and Electronic Engineering (DIBE), University of Genoa, Via Opera Pia 11a, I-16145, Genoa (Italy), e-mail: vulcano@dibe.unige.it.

§ EPI Ariana, UR INRIA Sophia Antipolis Méditeranée, 2004, Route des Lucioles, B.P.93, FR-06902, Sophia Antipolis Cedex (France), e-mail: Josiane.Zerubia@sophia.inria.fr.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Modélisation des statistiques des images radar (RSO) haute résolution

Résumé : En télédétection, un problème vital est le besoin de développer des modèles précis pour représenter les statistiques des intensités des images. Dans ce rapport de recherche, nous traitons le problème d'estimation de la densité de probabilité pour la modélisation d'amplitude d'une image haute résolution de type Radar à Synthèse d'Ouverture (RSO). Précédemment, plusieurs modèles théoriques et heuristiques ont été ultilisés pour représenter l'amplitude d'un signal du type RSO et ils ont montré leur efficacité pour certains types d'occupation du sol, rendant ainsi difficile le choix d'un seul modèle de densité de probabilité paramétrique.

Dans ce rapport de recherche, nous introduisons un algorithme d'estimation fondé sur un modèle de mélange fini de densités de probabilité d'amplitude dont les composantes appartiennent à un dictionnaire spécifique. La mèthode proposée intègre, de façon automatique: les procédures de sélection d'un modèle optimal pour chaque composante, l'estimation des paramètres, l'optimisation du nombre de composantes. Pour ce faire, nous utilisons simultanément l'algorithme EM stochastique et la méthode des log-cumulants en vue de l'estimation de la densité de probabilité paramétrique. Des résultats expérimentaux sur plusieurs images RSO réelles (issues des capteurs COSMO-SkyMed et RAMSES) sont présentés montrant que la méthode proposée est suffisamment précise pour modéliser des statistiques d'images d'amplitude RSO.

Mots-clés : image radar à synthèse d'ouverture (RSO), densité de probabilité, estimation paramétrique, modèles de mélange fini, EM stochastique (SEM).

Contents

1 Introduction

Synthetic aperture radar (SAR) is an active imagery system that can be operational regardless of the weather conditions and time of the day. SAR images are becoming widely used nowadays in various applications, e.g. in flood/fire monitoring, agriculture assessment, urban mapping. Modern SAR systems are often capable of providing very high resolution images (above 50 cm of land resolution). In the context of remotely sensed data analysis, a crucial problem is represented by the need to develop accurate models for the statistics of the pixel intensities. Focusing on Synthetic Aperture Radar (SAR) [10][11][40][47] data, this modeling process turns out to be a crucial task, for instance, for classification [15] or for denoising [40].

In this research report, we address the problem of probability distribution function (pdf) estimation in the context of SAR amplitude data analysis. Specifically, several different theoretic and heuristic models for the pdfs of SAR data have been proposed in the literature, and have proved to be effective for specific land-cover typologies. We further present the overview of the existing models.

Due to the variety of parametric pdf families each of them being efficient for some specific type of landcover the choice of a single optimal SAR amplitude pdf becomes a hard task. A remotely sensed image, in general, can depict a varied scene, jointly presenting several distinct land cover typologies. In addition, we aim to work with high resolution data that is likely to present additional complications in form of more complex histograms.

In this report, an innovative SAR amplitude parametric estimation algorithm is proposed, which addresses these problems by adopting a finite mixture model (FMM) [45][46] for the amplitude pdf, i.e., by postulating the unknown amplitude pdf to be a linear combination of parametric components, each one corresponding to a specific land cover type [13][44]. In order to take explicitly into account the possible differences in the statistics of the mixture components, we avoid choosing *a priori* a specific parametric family for each component, but we allow each of them to belong to a given dictionary of SAR-specific pdfs. The method extends to high resolution SAR an effective technique developed for lower resolution [34].

Specifically, the proposed algorithm automatically integrates the procedures of selection of the optimal model for each component and of parameter estimation, by combining the stochastic expectation Maximization (SEM) algorithm [4][8][30] and the *method-of-log-cumulants* (MoLC) [38][39]. The former is a stochastic iterative parametric estimation methodology, dealing with problems of data incompleteness and developed as an improvement of the standard expectation-maximization (EM) algorithm [14][45], in order to increase its capability to compute maximum likelihood (ML) estimates [53]. MoLC [38] is a recently proposed estimation approach originating from the adoption of the Mellin transform [49] (instead of the usual Fourier transform) in the computation of characteristic functions, and from the corresponding generalization of the concepts of moment and of cumulant [43]. We adopt this method both for its good estimation properties [37][38][51] and because it turns out to be feasible and fast for all the parametric families in the dictionary [35][37]. On the contrary, the well-known ML and it method-of-moments (MoM) estimation strategies [25][38] involve numerical difficulties for several of parametric families from the dictionary [35][38][40]. Ad-

ditionally, the developed method automatically performs an optimal choice of the number K of mixture components, by allowing K to decrease during the iterative process in order to find the best estimate, starting from initial upper bound value.

The proposed parametric approach is validated using an image acquired by the RAMSES airborne sensor and several real COSMO-SkyMed images. The experimental results show the developed algorithm to accurately model the amplitude distribution of all the considered images, both from a qualitative viewpoint (i.e., visual comparison between the data histogram and the estimated pdf) and from a quantitative viewpoint (i.e., correlation coefficient, Kolmogorov-Smirnov distance between the data histogram and the estimated pdf), thus showing its effectiveness and flexibility.

The research report is organized as follows. In Section 2 we present the state of the art overview of the existing SAR-specific pdf models. In Section 3, the proposed FMM-based estimation scheme is presented and the SEM and the MoLC methods are described. Section 4 reports the results of the application of the proposed approach to the statistical modeling of the grey-level of several real SAR images, showing the method to fit the amplitude distribution more efficient than previously proposed parametric models for SAR amplitude data. Finally, conclusions are drawn in Section 5.

2 Overview of the SAR-specific pdf models

In this section, we present an overview of the parametric models proposed for modeling SAR statistics. The list of models would not be exhaustive: we will only present the main statistical families; some other models, based on or with similar properties to more basic ones have been deduced. For the application purposes most of the models will be endowed with the method-of-log-cumulants (MoLC) equations [52], a recently proposed scheme for parametric estimation.

Looking from the methodological point of view, for the task of estimation one can apply either parametric or nonparametric strategies. A parametric approach postulates a predefined mathematical model in the form of the probability density function (pdf) and presents the estimation problem as a parameter estimation problem. Several classical statistical schemes are usually applied to deal with parameter estimation; these are, e.g., the *maximum-likelihood methodology* (ML) [40] and the *method of moments* (MoM) [40]. Nonparametric pdf estimation approaches on the other hand do not assume any specific form for the unknown pdf, thus providing a higher flexibility, although usually involving an additional fine tuning to be done manually by the user [15]. In particular, several nonparametric kernel-based estimation and regression schemes have been described in the literature, that have been proven to be effective estimation tools, such as, for instance, standard Parzen window estimators [15], artificial neural networks [5] [7] and support vector machines [29] [54].

This section will be organized as follows: first we describe the construction of the most basic theoretical model, then we present several heuristic models and, finally, we go to the class of theoretical, or better to say theoretically based, SAR statistical models.

A standard model of the complex signal statistics for the case of singlelook SAR is derived as follows: the signal is backscattered by a given ground area, being illuminated by a singlelook SAR sensor; we also assume that the number of scatterers is large, the scatterers are independent and small (compared to the ground area), the scattering instantaneous phases are independent of the amplitudes and uniformly distributed in $[0, 2\pi]$, and there is no single scatterer dominating the scene [40] [25]. We denote by z the complex signal received by the SAR sensor from the ground area corresponding to a given pixel, so that:

$$z = x + iy = \sqrt{v}\exp(i\theta),$$

where x, y, v and θ are the real part, the imaginary part, the *intensity*, and the *phase* of the complex signal, respectively. The model assumes the presence of a finite set of n independent scattering entities in the observed area, thus interpreting z as the result of the interference of the corresponding contributions. This interference phenomenon motivates the usual noise-like granular aspect of SAR images, known as *speckle* [40]. In particular, assuming the number of scatterers to be large, according to the central limit theorem, the real and imaginary parts of the backscattered signal are assumed to be jointly Gaussian. As a matter of fact they turn out to be independent, zero-mean Gaussian random variables with equal variances, thus yielding an exponential distribution for the signal intensity and

a **Rayleigh** distribution for the signal amplitude (we call $r = \sqrt{v}$ the *amplitude*) [40]:

$$p(v) = \lambda \exp(-\lambda v)$$

$$p(r) = 2\lambda r \exp(-\lambda r^2),$$

where λ is a distribution parameter to be estimated according to the image data by using, for example, ML or MoM. Here $r, v \geqslant 0$, their pdfs are zero on $(-\infty, 0)$ and from now on we shall explicitly define their pdfs only on $[0, +\infty)$. However, real SAR amplitude data often present significantly non-Rayleigh empirical distributions by, for instance, exhibiting heavier distribution tails and, thus, requiring a more accurate pdf characterization.

2.1 Heuristic models

Several heuristic models have been proposed in order to overcome the descriptive weakness of Rayleigh distribution in some cases. The log normal and Weibull, the two-parametric distributions, have shown good results in particular cases. For instance, the amplitude of low-resolution sea clutter has been fitted to the one-parameter Rayleigh distribution, corresponding to pure speckle, while the log normal has been applied at a higher resolution. A wide range of ocean measurements at different resolutions were shown to be consistent with the Weibull distribution. Land clutter was found to be Rayleigh distributed over uniform region, but log normal over built-up areas. The Weibull distribution has also been applied extensively to land and sea-ice clutter [40].

The **log normal** distribution is given by:

$$p(x) = \frac{1}{x\sqrt{2\pi V}} \exp\left[-\frac{(\ln x - \beta)^2}{2V}\right],$$

where β and V are the mean and the variance of $\ln x$ respectively; here x can stand for either v or r. Its MoLC equations are written as follows:

$$\begin{cases} \kappa_1 = \beta \\ \kappa_2 = V, \end{cases}$$

where κ_1 and κ_2 are the first and the second log-cumalants respectively as defined in [52]. The log normal distribution predicts zero probability of the observable having a zero value, thus being far from the ideal representation of a single-look intensity speckle.

The **Weibull** distribution is given by:

$$p(x) = \frac{cx^{c-1}}{b^c} \exp\left[-\left(\frac{x}{b}\right)^c\right],$$

where b is a scaling parameter and c controls the shape. Its MoLC equations are:

$$\begin{cases} \kappa_1 = \ln b + \Psi(1)c^{-1} \\ \kappa_2 = \Psi(1,1)c^{-2}. \end{cases}$$

Here $\Psi(\cdot)$ is the digamma function [6] and $\Psi(\nu,\cdot)$ is the νth order polygamma function [6]. The Weibull distribution is identical to the Rayleigh pdf if $c=2$ and becomes negative exponential when $c=1$. Thus, it can describe single-look speckle precisely for both amplitude and intensity. Unfortunately, it lacks the ability to represent multilook speckle correctly.

A further empirical model, the **Fisher** distribution [52], has shown very good application results, thanks to its capability to model a big variety of distribution tails. The distribution is given by:

$$p(x)=\frac{\Gamma(L+M)}{\Gamma(L)\Gamma(M)}\,\frac{L}{M\mu}\,\frac{\left(\frac{Lx}{M\mu}\right)^{L-1}}{\left(1+\frac{Lx}{M\mu}\right)^{L+M}},$$

with $L,M>0$ - the *degrees of freedom* parameters. Up to this moment, Fisher distribution has not been theoretically proven to have relation to the physics of wave scattering. But it has been shown that it has a natural relationship to multiplicative noise by the bias of Mellin transform and second-kind statistics [52].

The MoLC equations for the Fisher pdf are given by:

$$\begin{cases}\kappa_1=\ln\mu+(\Psi(L)-\ln L)-(\Psi(M)-\ln M)\\ \kappa_2=\Psi(1,L)+\Psi(1,M)\\ \kappa_3=\Psi(2,L)-\Psi(2,M).\end{cases}$$

In a recent work [28] the so-called **Generalized Gamma** distribution (GΓD) was proposed as an empirical statistical model for SAR images. The GΓD is defined by:

$$p(x)=\frac{\nu}{\sigma\Gamma(\kappa)}\,\left(\frac{x}{\sigma}\right)^{\kappa\nu-1}\exp\left\{-\left(\frac{x}{\sigma}\right)^{\nu}\right\},$$

where ν, κ and σ are the positive real values corresponding to power, shape and scale parameters respectively, and $\Gamma(\cdot)$ denotes the Gamma function. The GΓD contains a large variety of distributions, including the Rayleigh ($\nu=2$, $\kappa=1$), exponential ($\nu=1$, $\kappa=1$), Nakagami ($\nu=2$), Gamma ($\nu=1$), log-normal ($\kappa\to\infty$) and Weibull ($\kappa=1$).

The MoLC equations for the GΓD are as follows:

$$\begin{cases}\kappa_1=\Psi(\kappa)/\nu+\ln\sigma\\ \kappa_2=\Psi(1,\kappa)/\nu^2\\ \kappa_3=\Psi(2,\kappa)/\nu^3.\end{cases}$$

As a generalization of some of the mentioned distributions, GΓD demonstrated a better fitting of data as compared to Weibull, Nakagami and K-distribution (see below) on a number of SAR images [28].

2.2 Theoretical models

Several theoretical models have also been proposed in order to enhance the descriptive quality or to generalize the most basic Rayleigh model to different typologies of radar data.

The first one, the **Nakagami-Gamma** distribution has been proposed to model the SAR statistics in the presence of a single strong reflector in a homogeneous clutter [52]. The Gamma distribution has been introduced as a model for a multilook SAR intensity pdf:

$$p(v) = \frac{(\lambda L)^L}{\Gamma(L)}\, v^{L-1} \exp(-\lambda L v).$$

This model generalizes the exponential distribution of intensities by averaging L single-look exponential distributions [40]. The corresponding amplitude takes the form of the Nakagami distribution [52]:

$$p(r) = \frac{2(\lambda L)^L}{\Gamma(L)}\, r^{2L-1} \exp(-\lambda L r^2).$$

The MoLC equations for the Nakagami distribution are:

$$\begin{cases} 2\kappa_1 = -\ln\lambda + \Psi(L) - \ln L \\ 4\kappa_2 = \Psi(1, L). \end{cases}$$

The Nakagami-Gamma model has also been extended to the case of multilook polarimetric and interferometric data [21] [22] [27] and to finite mixtures of Gamma components [39]; its application has been generalized by letting the integer number L of looks be a real positive parameter (interpreted as an *equivalent number of looks* (ENL)) to be estimated together with λ according to the image data.

In [25], a generalized version of the central limit theorem is applied in order to extend the standard scattering model by allowing the real and imaginary parts of the backscattered signal to be jointly symmetrically α-stable ($S\alpha S$) distributed random variables [26], thus resulting in the following generalized **heavy-tailed Rayleigh** model for the amplitude pdf:

$$p(r) = r \int_0^{+\infty} \rho \exp(-\gamma \rho^\alpha) J_0(r\rho) d\rho,$$

where α and γ are positive parameters and J_0 is the zeroth order Bessel function of the first kind [6]. The MoLC equations for the $S\alpha S$ model are:

$$\begin{cases} \alpha\kappa_1 = \Psi(1)(\alpha - 1) + \alpha \ln 2 + \ln\gamma \\ \kappa_2 = \Psi(1, 1)\alpha^{-2}. \end{cases}$$

A different approach to SAR scattering modeling is proposed in [20] by assuming the number n of scatterers to be in itself a random variable and the population of scatterers to be controlled by a *birth-death migration process*. In this case, a **K distribution** is obtained for the signal intensity [20] [41]:

$$p(v) = \frac{2(\lambda L M)^{(L+M)/2}}{\Gamma(L)\Gamma(M)}\, v^{(L+M-2)/2} K_{M-L}\left[2(\lambda L M)^{1/2}\right],$$

where λ, L and M are positive distribution parameters, and $K_\tau(\cdot), \tau > 0$ is the τth order modified Bessel function of the second kind [6]. The corresponding amplitude distribution (we refer to it as *K-root* distribution) is given by:

$$p(r) = \frac{4(\lambda LM)^{(L+M)/2}}{\Gamma(L)\Gamma(M)} r^{L+M-1} K_{M-L}\left[2r(\lambda LM)^{1/2}\right].$$

The MoLC equations for K-root are given by:

$$\begin{cases} 2\kappa_1 = -\ln\lambda + \Psi(L) - \ln L + \Psi(M) - \ln M \\ 4\kappa_2 = \Psi(1, L) + \Psi(1, M) \\ 8\kappa_3 = \Psi(2, L) + \Psi(2, M). \end{cases}$$

The same pdf can be obtained by assuming a multiplicative noise model for the SAR intensity by expressing v as the product of two Gamma-distributed components representing a signal and a noise contribution, respectively [40]. Further extensions of the backscattering modeling approach assuming n as a random variable are proposed in [13] [42] [57] [17]. MoM turns out to be feasible for the parameter estimation task concerning a K-distributed random variable, whereas no closed form is currently available for ML parameter estimation, thus requiring intensive numerical computations or analytical approximations of the pdf itself.

Inverse Gaussian (IG) distributions have also been employed to model the amplitude statistics [17] [36] [16]. A multiplicative model is adopted in [17], assuming a Nakagami distribution for the speckle noise component and a generalized inverse Gaussian (GIG) law for the signal component, thus resulting in the following distribution (named **G distribution**) for the amplitude return:

$$p(r) = \frac{2L^L \left(\frac{\lambda}{\gamma}\right)^{\alpha/2}}{\Gamma(L)K_\alpha(2\sqrt{\lambda\gamma})} r^{2L-1} \left(\frac{\gamma + Lr^2}{\lambda}\right)^{(\alpha-L)/2} K_{\alpha-L}\left[2\sqrt{\lambda(\gamma + Lr^2)}\right],$$

where L is the ENL of the Nakagami speckle distribution, and α, λ, and γ are the parameters characterizing the GIG signal distribution. Such a general model includes, as particular cases, the *square root of K* described above and the $\mathbf{G}^0$ distribution, which is derived in [17] with a moment-based parameter estimation algorithm and is successfully applied to the statistics of SAR amplitude data over extremely heterogeneous areas. In [52], the $\mathbf{G}^0$ distribution is proved to be equivalent to a Fisher pdf and is applied to the characterization of the statistics of high-resolution SAR imagery over urban areas. A further particular case of the $\mathbf{G}$ model (named the *harmonic branch* $\mathbf{G}^h$) is proposed in [36] and endowed with a moment-based estimation approach to images of urban areas and mixed terrain. In [16], a normal IG distribution is proposed for the real and imaginary parts of the backscattered complex signal, thus resulting in an amplitude pdf formulated as a combination of an IG pdf and a Rice pdf:

$$p(r) = \frac{\sqrt{\frac{2}{\pi}}\alpha^{3/2}\delta \exp\left(\delta\sqrt{\alpha^2 - \beta^2}\right) r}{(\delta^2 + r^2)^{3/4}} K_{3/2}\left(\alpha\sqrt{\delta^2 + r^2}\right) I_0(\beta r),$$

where α, β and δ are the distribution parameters, and $I_o(\cdot)$ is the zeroth order modified Bessel function of the first kind. For this *Rician inverse Gaussian* (RiIG) pdf, a case-specific iterative parameter estimation algorithm is developed in [16].

In the recent **Generalized Gaussian Rayleigh** Model, the theoretical circular Gaussian speckle model is generalized, on a phenomenological basis, by assuming the real and imaginary parts of the backscattered signal to be independent zero-mean generalized Gaussian (GG) [35] random variables and by deriving the resulting amplitude pdf analytically. The resulting model is given by:

$$p(r) = \frac{\gamma^2 c^2 r}{\Gamma^2\left(\frac{1}{c}\right)} \int_0^{\pi/2} \exp[-(\gamma r)^c(|\cos\theta|^c + |\sin\theta|^c)]d\theta \ ,$$

where $c, \gamma > 0$ are the distribution parameters. The MoLC equations for this model are:

$$\begin{cases} \kappa_1 = \lambda\Psi(2\lambda) - \ln\gamma - \lambda G_1(\lambda)[G_0(\lambda)]^{-1} \\ \kappa_2 = \lambda^2\Psi(1,2\lambda) + \lambda^2 G_2(\lambda)[G_0(\lambda)]^{-1} - \lambda^2[G_1(\lambda)]^2[G_0(\lambda)]^{-2}, \end{cases}$$

where $\lambda = \frac{1}{c}$ and $G_\nu(\cdot)$ is the integral function introduced in [41] for GGR parametric estimation. The application has shown accurate results of the GG model for images acquired by different SAR sensors at medium resolution with various types of scenes. The applicability of this model was also shown to be complementary to those of the K-root. Some results were also achieved in the application to the multilook images, although such images were not explicitly addressed in the methodology of this model [35].

We stress here that for all the models above provided with MoLC-equations except for GGR and K-root the equations have the unique solution for whatever the sample estimates of log-moments are. For GGR and K-root some combinations of sample estimates may lead to no solutions of MoLC-equations ([20], [35]).

3 A dictionary-based finite mixture model for SAR data pdf estimation

3.1 Finite mixture models for SAR amplitude data

In order to take explicitly into account the possible presence in a given SAR amplitude [40] image $\mathcal{I}$ of several distinct land-cover typologies, yielding different contributions to the statistics of the pixel intensity, we assume a finite mixture model (FMM) [45][46] for the grey level pdf. Specifically, we model the SAR image as a set $\mathcal{I} = \{r_1, r_2, \ldots, r_N\}$ of independent and identically distributed (i.i.d.) samples drawn according to the following pdf[1]:

$$p_r(r|\theta) = \sum_{i=1}^{K} P_i p_i(r|\theta_i), \quad r \geq 0, \tag{1}$$

where $p_i(\cdot|\cdot) : [0, +\infty) \times \Theta_i \rightarrow [0, +\infty)$ is a probability density function depending on a vector θ_i of parameters, taking values in a set $\Theta_i \subset \mathbb{R}^{\ell_i}$ $(i = 1, 2, \ldots, K)$, $\{P_1, P_2, \ldots, P_K\}$ is a set of mixing proportions such that:

$$\sum_{i=1}^{K} P_i = 1, \qquad 0 \leq P_i \leq 1, \quad i = 1, 2, \ldots, K, \tag{2}$$

and θ is a vector collecting all the parameters of the distribution, i.e.:

$$\theta = (\theta_1, \theta_2, \ldots, \theta_K; P_1, P_2, \ldots, P_K). \tag{3}$$

Denoting as Θ the set of possible parameter vectors, i.e.:

$$\Theta = \Theta_1 \times \Theta_2 \times \ldots \times \Theta_K \times \left\{ P \in [0,1]^K : \sum_{i=1}^{K} P_i = 1 \right\}, \tag{4}$$

the problem of FMM parametric estimation, i.e., the computation of a parameter vector $\theta^* \in \Theta$ optimally representing the observed image data $\mathcal{I}$, has been addressed according to several different approaches [45][46]. Specifically, the i.i.d. assumption allows obtaining the following expression for the log-likelihood function [53] of the image data $\mathcal{I}$ [45][46]:

$$L_{\mathcal{I}}(\theta) = \sum_{k=1}^{N} \ln p_r(r_k|\theta) = \sum_{k=1}^{N} \ln \left[\sum_{i=1}^{K} P_i p_i(r_k|\theta_i) \right], \quad \theta \in \Theta. \tag{5}$$

The computation of Maximum Likelihood (ML) estimates involves the maximization of this function, but the solution of this maximization problem is not feasible analytically [46] and

[1] This approach is widely accepted in the context of estimation theory [15][18][53] and operatively corresponds to discard in the estimation process the contextual information associated to the correlation between neighboring pixels in the image, thus exploiting only the greylevel information.

also involves several numerical difficulties, due, for instance, to the common phenomenon of presence of several local maxima [45]. In order to get round this problems in the computation of ML estimates, the use of the Expectation-Maximization (EM) algorithm has been proposed [14][32][45], which formalizes the problem of the estimation of the parameters of a FMM as an incomplete data problem and introduces a sequence $\{\theta^t\}_{t=0}^{\infty}$ of parameter estimates, iteratively maximizing a Pseudo-Likelihood function, i.e. [14][45]:

$$\theta^{t+1} = \arg\max_{\theta\in\Theta} Q_{\mathcal{I}}(\theta|\theta^t) \tag{6}$$

where

$$Q_{\mathcal{I}}(\theta|\theta^t) = \sum_{k=1}^{N}\sum_{i=1}^{K} \tau_{ik}^t[\ln P_i + \ln p_i(r_k|\theta_i)], \quad \tau_{ik}^t = \frac{P_i^t p_i(r_k|\theta_i^t)}{p_r(r_k|\theta^t)}, \quad \theta\in\Theta. \tag{7}$$

In the context of mixture densities EM has been proved to converge to a stationary point of the log-likelihood function $L_{\mathcal{I}}(\cdot)$ [14][45][56], although not converging, in general, to a global maximum point, and exhibiting sometimes long convergence times [8][46]. In addition, the maximization problem (6), although analytically tractable in several applications [32] (e.g.: mixtures of exponential families [45]), does not always yield a closed form solution, and moreover the convergence point can belong to the boundary of the parameter space Θ, thus possibly involving analytical singularities [46]. Several variants of EM have been introduced in order to address these difficulties. In [19], [24], and [50] modified versions of EM, as well as regularized covariance estimators, are proposed in order to optimize the robustness of the estimation process in the context of Gaussian Mixtures Models (GMMs) for hyperspectral data classification. A simplified version of EM, named Classification EM (CEM) [4], has been developed, which converges in a finite number of iterations, but yields, in general, biased parameter estimates. A sequential version of EM, namely, the "Component-wise EM for mixtures" (CEM^2), is proposed in [9], which aims at a reduction in the computation time and also at avoiding analytical singularities [46]. The Stochastic EM (SEM) [8] avoids the computation of the Pseudo-Likelihood function $Q_{\mathcal{I}}(\cdot|\cdot)$ and the related analytical maximization issues, by integrating a stochastic sampling procedure in the estimation process. Hence, the sequence of parameter estimates generated by SEM is a discrete time random process, which does not converge pointwise, but has been proved to be an ergodic and homogeneous Markov chain, converging to a unique stationary distribution, which is expected to be concentrated around the global maxima of the log-likelihood function [8]. Simulated Annealing EM (SAEM) [8] is a combination of EM and SEM allowing SEM to converge also almost surely (a.s.) [43][55], although at a local maximum and requiring the preliminary definition of a suitable annealing schedule. Monte Carlo EM (MCEM) and Simulated Annealing Monte Carlo EM (SA-MCEM) are further stochastic variants of EM substituting a Monte Carlo sampling procedure to the computation of the $Q_{\mathcal{I}}(\cdot|\cdot)$ function and a.s. converging, under mild assumptions, to a local maximum [8]. In [12] the Iterated Conditional Expectation (ICE) method is used, as an alternative to EM and SEM, in the FMM estimation for image segmentation and clustering purposes and

in [46] the Minimum Message Length (MML) principle is applied in order to formulate a FMM unsupervised parametric estimator, aiming at avoiding convergence to the boundary of the parameter space, as well as to improve, with respect to EM, robustness with respect to initialization, but still involving the maximization problem (see Eq. 6).

We will adopt here the SEM algorithm, both due to its capability to avoid local maxima of the log-likelihood function and thanks to its independence from the analytical maximization process in Eq. (6). In fact, the adoption of several of the usual SAR amplitude or intensity parametric models (such as Weibull or K) for the mixture components yields no closed form solution for this optimization problem, thus strongly complicating the application of most above-mentioned estimators.

3.2 SEM for FMM parametric density estimation

A general parametric density estimation problem assumes the availability of an observation random vector $x \in X \subset \mathbb{R}^n$, whose density function $p_x(\cdot|\cdot) : X \times \Theta \to [0, +\infty)$ depends on a parameter vector taking values in a given set $\Theta \subset \mathbb{R}^\ell$ (namely, the parameter space).

This general problem is said to present data incompleteness when the data vector x cannot be directly observed, for instance, due to lacking or corrupted data [32]. Such incompleteness issues is formalized by assuming the "complete" data vector x not to be available, but to be observed only through an "incomplete" data vector $y = \Phi(x)$, obtained through a many-to-one mapping $\Phi : X \to Y \subset \mathbb{R}^m$ [14]. Hence, a given realization $y \in Y$ of the incomplete data may have been generated by any realization $x \in \Phi^{-1}(y) \subset X$ in the inverse image $\Phi^{-1}(y)$ of y, thus not allowing, for instance, a direct feasible computation of an ML estimate. SEM tries to avoid these difficulties by iteratively random sampling a complete data set and by using it to compute a ML standard estimate.

Specifically, we denote as $p_y(\cdot|\cdot) : Y \times \Theta \to [0, +\infty)$ the parametric incomplete data density function[2] and as $p_{x|y}(\cdot|y, \cdot) : \Phi^{-1}(y) \times \Theta \to [0, +\infty)$ the complete data parametric density conditioned to an incomplete data realization $y \in Y$, and we state that [14][8][33]:

$$p_y(y|\theta) = \int_{\Phi^{-1}(y)} p_x(x|\theta) d\zeta_y(x), \quad \theta \in \Theta, \tag{8}$$

$$p_{x|y}(x|y, \theta) = \frac{p_x(x|\theta)}{p_y(y|\theta)}, \quad x \in \Phi^{-1}(y),\ \theta \in \Theta, \tag{9}$$

where ζ_y is a suitable measure[3] on $\Phi^{-1}(y)$.

Given an observed incomplete data realization $y \in Y$, SEM computes a random sequence $\{\theta^t\}_{t=0}^{\infty}$, by performing at the t-th iteration ($t = 0, 1, \ldots$) the following three processing steps [8]:

[2] As with x, this is the density function of the random vector y with respect to a suitable σ-finite measure space $(Y, \mathcal{Y}, \eta)$ introduced on Y.

[3] More specifically, a σ-finite measure space $(\Phi^{-1}(y), \mathcal{Z}_y, \zeta_y)$ is introduced on the inverse image $\Phi^{-1}(y)$ of $y \in Y$, such that the restriction of $p_x(\cdot|\theta)$ to $\Phi^{-1}(y)$ is $\mathcal{Z}_y$-measurable for all $\theta \in \Theta$. This allows introducing the model described by Eq. (8) for the incomplete data density, which yields the expression (9) for the conditional complete data one [14][33][45].

- **E-step**: compute the conditional complete data density $p_{x|y}(\cdot|y,\theta^t)$ corresponding to the current parameter estimate $\theta^t \in \Theta$;
- **S-step**: sample a complete data realization $x^t \in X$ according to the conditional density computed in the E-step;
- **M-step**: update the parameter estimate, by computing a standard ML estimate $\theta^{t+1} \in \Theta$ according to the complete data realization x^t sampled in the S-step.

As previously stated, the resulting discrete time random process $\{\theta^t\}_{t=0}^{\infty}$ is not pointwise nor a.s. convergent, but has been proven to be a homogeneous Markov chain. If this sequence turns out to be also ergodic, it converges in distribution to the unique stationary distribution of the Markov chain and is expected to concentrate around the global maxima of the likelihood function [8].

The FMM case satisfies this ergodicity assumption [8], thus suggesting SEM as a promising estimation tool. The specific FMM framework can be viewed as affected by data incompleteness problems, since it is not known from which of the K available statistical populations involved in Eq. (1) a given image sample is drawn. This implicitly means that no training information about the possible thematic classes in the SAR image is exploited in the estimation process, i.e., the SAR amplitude pdf estimation problem is addressed in an unsupervised context. Thus, denoting as $\Sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_K\}$ the set of the K different populations, we assume the population label $s_k \in \Sigma$ of the k-th image pixel not to be known, thus suggesting the following definition of the complete and of the incomplete data vectors, respectively:

$$x = (r_1, s_1, r_2, s_2, \ldots, r_N, s_N), \qquad y = (r_1, r_2, \ldots, r_N). \tag{10}$$

Assuming the couples $\{(r_k, s_k) : k = 1, 2, \ldots, N\}$ of random variables to be i.i.d., we denote as $p_{r|s}, P_s, P_{s|r}$, and p_{rs} the parametric pdf of r_k conditioned to s_k, the parametric probability mass function (PMF) of s_k (i.e., the label prior probability), the parametric PMF of s_k conditioned to r_k (i.e., the label posterior probability), and the parametric joint density of (r_k, s_k), respectively $(k = 1, 2, \ldots, N)$. Hence [33][45]:

$$P_s(\sigma_i|\theta) = P_i, \quad p_{r|s}(r_k|\sigma_i,\theta) = p_i(r_k|\theta_i), \quad \theta \in \Theta,\ i = 1, 2, \ldots, K,\ k = 1, 2, \ldots, N, \tag{11}$$

$$p_{rs}(r_k, s_k|\theta) = p_{r|s}(r_k|s_k,\theta)P_s(s_k|\theta), \quad \theta \in \Theta,\ k = 1, 2, \ldots, N. \tag{12}$$

The i.i.d. assumption allows obtaining the following expressions for the density functions of the incomplete and of the complete data vectors, respectively:

$$p_y(y|\theta) = \prod_{k=1}^{N} p_r(r_k|\theta), \qquad p_x(x|\theta) = \prod_{k=1}^{N} p_{rs}(r_k, s_k|\theta) = \prod_{k=1}^{N} p_{r|s}(r_k|s_k,\theta)P_s(s_k|\theta), \tag{13}$$

$$p_{x|y}(x|y,\theta) = \frac{p_x(x|\theta)}{p_y(y|\theta)} = \prod_{k=1}^{N} p_{r|s}(r_k|s_k,\theta)\frac{P_s(s_k|\theta)}{p_r(r_k|\theta)} = \prod_{k=1}^{N} P_{s|r}(s_k|r_k,\theta) \tag{14}$$

Therefore, the t-th iteration of the SEM algorithm in the present FMM context involves the following operations ($t = 0, 1, \ldots$):

- **E-step:** compute the values of the posterior probabilities corresponding to the current parameter estimate $\theta^t \in \Theta$, i.e. ($k = 1, 2, \ldots, N$, $i = 1, 2, \ldots, K$):

$$P_{s|r}(\sigma_i|r_k, \theta^t) = p_{r|s}(r_k|\sigma_i, \theta^t)\frac{P_s(\sigma_i|\theta^t)}{p_r(r_k|\theta^t)} = \frac{P_i^t p_i(r_k|\theta_i^t)}{p_r(r_k|\theta^t)} = \tau_{ik}^t; \tag{15}$$

- **S-step:** sample randomly a complete data realization x^t, by sampling a label s_k^t for each k-th pixel according to the current estimated posterior probability distribution $\{\tau_{ik}^t : i = 1, 2, \ldots, K\}$ of the pixel ($k = 1, 2, \ldots, N$), thus implicitly partitioning the image $\mathcal{I}$ in K subsets;

- **M-step:** update the parameter estimates, by computing, according to the partition generated by the S-step, a standard supervised ML estimate $\theta^{t+1} \in \Theta$, i.e.[4]:

$$P_i^{t+1} = \frac{|Q_{it}|}{N}, \quad \theta_i^{t+1} = \arg\max_{\phi \in \Theta_i} \sum_{k \in Q_{it}} \ln p_i(r_k|\phi), \quad i = 1, 2, \ldots, K, \tag{16}$$

  where $Q_{it} = \{k : s_k^t = \sigma_i\}$ is the index set of the image samples assigned to the component σ_i ($i = 1, 2, \ldots, K$).

A common initialization procedure for the described iterative process consists of setting initially a uniform posterior distribution for all image pixels, but in order to provide the iterative process with a more reasonable starting point (thus avoiding time consuming burn-in iterations) we take into consideration the form of the amplitude histogram. Furthermore, since SEM is not pointwise or a.s. convergent, a specific termination procedure has to be defined in order to extract from the stationary steady distribution of the random sequence $\{\theta^t\}_{t=0}^{\infty}$ a single parameter estimate θ^*. Several strategies have been proposed to this purpose, for instance, in [4] and in [8].

3.3 The "dictionary" approach to SAR amplitude pdf estimation

The application of the FMM model described by Eq. (1) requires the definition of suitable parametric models for the mixture components. However, as described in Section 2, in the present SAR-specific context several parametric models have been proposed and proved to be effective descriptions of the statistics of the pixel intensities corresponding to different land cover typologies. In order to overcome the intrinsic difficulty of an *a priori* choice of a single suitable model and to improve the flexibility of the adopted FMM approach, we avoid selecting a specific parametric family for each component $p_i(\cdot|\cdot)$ ($i = 1, 2, \ldots, K$), and we adopt instead a finite dictionary $\mathcal{D} = \{f_1, f_2, \ldots, f_M\}$ of M SAR-specific distinct

[4] Given a finite set A, we denote by $|A|$ the cardinality (i.e., the number of elements) of A.

parametric pdfs $f_j(\cdot|\cdot) : [0,+\infty) \times \Xi_j \to [0,+\infty)$ $(j = 1, 2, \ldots, M)$ with parameter spaces Ξ_j $(j = 1, 2, \ldots, M)$. In [12] a similar approach was proposed, which applied EM, SEM, and ICE to a "generalized mixture model" with mixture components not restricted to belong to a specific parametric family, but selected inside the Pearson system of distributions. Here, dealing with SAR amplitude data pdf estimation, we adopt a dictionary $\mathcal{D}$ consisting of the following eight parametric pdfs:

- Lognormal, the empirical log-normal distribution [40]:

$$f_1(r|m,\sigma) = \frac{1}{\sigma r\sqrt{2\pi}} \exp\left[-\frac{(\ln r - m)^2}{2\sigma^2}\right], \quad r > 0; \tag{17}$$

- Weibull, the empirical Weibull distribution [37][40]:

$$f_5(r|\eta,\mu) = \frac{\eta}{\mu^\eta} r^{\eta-1} \exp\left[-\left(\frac{r}{\mu}\right)^\eta\right], \quad r \geq 0; \tag{18}$$

- Fisher, the empirical Fisher distribution [52]:

$$p(r|L,M,\mu) = \frac{\Gamma(L+M)}{\Gamma(L)\Gamma(M)} \frac{L}{M\mu} \frac{\left(\frac{Lr}{M\mu}\right)^{L-1}}{\left(1+\frac{Lr}{M\mu}\right)^{L+M}}, \quad r \geq 0, \tag{19}$$

  where $\Gamma(\cdot)$ is the Gamma function [48];

- GGamma, the empirical GΓD distribution [28]:

$$p(r|\nu,\sigma,\kappa) = \frac{\nu}{\sigma\Gamma(\kappa)} \left(\frac{r}{\sigma}\right)^{\kappa\nu-1} \exp\left\{-\left(\frac{r}{\sigma}\right)^\nu\right\}, \quad r \geq 0; \tag{20}$$

- Nakagami, the Nakagami distribution, proposed as an amplitude model for multi-look SAR data [38][51]:

$$f_2(r|L,\mu) = \frac{2}{\Gamma(L)} \left(\frac{L}{\mu}\right)^L r^{2L-1} \exp\left(-\frac{Lr^2}{\mu}\right), \quad r \geq 0; \tag{21}$$

- K-root, the amplitude distribution corresponding to a K-distributed intensity [40] (hereafter denoted simply as "K-root"):

$$f_6(r|L,M,\mu) = \frac{4}{\Gamma(L)\Gamma(M)} \left(\frac{LM}{\mu}\right)^{(L+M)/2} r^{L+M-1} K_{M-L}\left[2r\left(\frac{LM}{\mu}\right)^{1/2}\right], \quad r \geq 0; \tag{22}$$

- GGD, the generalized Gaussian Rayleigh (GGR) distribution, based on a generalized Gaussian model for the backscattered SAR signal [35]:

$$f_3(r|\lambda,\gamma) = \frac{\gamma^2 r}{\lambda^2\Gamma^2(\lambda)} \int_0^{\pi/2} \exp[-(\gamma r)^{1/\lambda}(|\cos\theta|^{1/\lambda} + |\sin\theta|^{1/\lambda})]d\theta, \quad r \geq 0; \tag{23}$$

- HT-Rayleigh, the SαS generalized Rayleigh distribution (hereafter simply denoted as SαSGR), based on a SαS model for the SAR backscattered signal [25]:

$$f_4(r|\alpha,\gamma) = r \int_0^{+\infty} \rho \exp(-\gamma\rho^\alpha) J_0(r\rho)d\rho, \quad r \geq 0. \tag{24}$$

Hence, with this specific choice, $M = 8$ distinct parametric families are involved in the estimation process with 2-parameter families (i.e., $\Xi_j = (0,+\infty)^2$) for $j = 1,2,5,7,8$ and a 3-parameter family for $j = 3,4,6$ (i.e., $\Xi_6 = (0,+\infty)^3$). We do not include the Rayleigh distribution in the dictionary, since this pdf is a particular case of almost all the pdfs above [25][28][35][40].

Specifically, we integrate this dictionary-based approach in the described SEM estimation framework, by exploiting at each SEM iteration the image partition induced by the sampling process in order to fit each parametric family in the dictionary to each mixture component, thus generating a set of M feasible candidate estimates per component. Hence, an optimality criterion has to be defined in order to choose, for each component, the optimal estimate among the available candidates. Specifically denoting as $\xi_{ij}^t \in \Xi_j$ the optimal parameter estimate computed for the j-th parametric model $f_j(\cdot|\xi_j)$ ($\xi_j \in \Xi_j$) in the dictionary according to the data samples assigned to the i-th component σ_i at the t-th SEM iteration, we adopt, as a selection criterion for σ_i, the corresponding log-likelihood, i.e.:

$$L_{ij}^t = \sum_{k \in Q_{it}} \ln f_j(r_k|\xi_{ij}^t), \quad i = 1,2,\ldots,K,\ j = 1,2,\ldots,M. \tag{25}$$

Hence, the pdf estimate for the component σ_i is updated as the candidate estimate $f_j(\cdot|\xi_{ij}^t)$ yielding the highest value of L_{ij}^t ($i = 1,2,\ldots,K$, $j = 1,2,\ldots,M$, $t = 0,1,\ldots$).

We stress however that the computation at the t-th iteration of the set $\{\xi_{ij}^t : i = 1,2,\ldots,K,\ j = 1,2,\ldots,M\}$ of the optimal parameter vectors according to the M-step of the SEM algorithm should be performed by using an ML procedure, i.e.:

$$\xi_{ij}^t = \arg\max_{\xi \in \Xi_j} \sum_{k \in Q_{it}} \ln f_j(r_k|\xi) \tag{26}$$

This approach turns out to be unfeasible for several SAR-specific pdfs, such as the K distribution [40]. Hence, we avoid using ML estimates and we adopt in the M-step the "method of log-cumulants" (MoLC) [38] instead, which has been proved to be a feasible and effective estimation tool for all usual SAR parametric models [37][38].

3.4 Parameter estimation with the Mellin transform and MoLC

The method of log-cumulants (MoLC) has been recently proposed as a parametric pdf estimation technique feasible for distributions defined on $[0, +\infty)$, and has been explicitly applied in the context of the usual parametric families employed for SAR amplitude and intensity data modelling (e.g. for the Nakagami-Gamma and the K distributions) [37][38][39][51]. MoLC is based on the generalization of the usual moment-based statistics, by using the Mellin transform in the computation of characteristic functions and moment generating functions, instead of the usual Fourier and Laplace transforms.

Given a generic random variable u, the moment generating function (MGF) Φ_u of u is defined as the bilateral Laplace transform of the pdf of u [43], i.e.:

$$\Phi_u(s) = \mathcal{L}(p_u)(s) = \int_{-\infty}^{+\infty} p_u(u) \exp(su) du, \quad s \in \mathbb{C}, \tag{27}$$

where $\mathcal{L}$ is the bilateral Laplace transform operator[5] on the Lebesgue space $L^1(\mathbb{R})$ [49]. The MGF is known to converge and to be analytical at least in a vertical strip of the complex plane and it turns out to be implicitly related to the MoM estimation approach. In fact, if the interior of the convergence strip contains a neighborhood of the origin, then the ν-th order moment ($\nu = 1, 2, \ldots$) can be expressed as

$$m_\nu = E\{u^\nu\} = \Phi_u^{(\nu)}(0), \tag{28}$$

where the superscript denotes a differentiation operator [43]. Related quantities are the characteristic function of u, defined as the Fourier transform of the pdf, the second moment generating function Ψ_u, defined as the complex logarithm of the MGF, and the ν-th order cumulant k_ν, defined as the ν-th order derivative of the second MGF computed in the origin of the complex plane:

$$\Psi_u(s) = \ln \Phi_u(s), \qquad k_\nu = \Psi_u^{(\nu)}(0). \tag{29}$$

In particular, the first and second order cumulants turn out to be equal to the distribution mean and variance, respectively [43].

The MoM estimates are actually computed by analytically expressing the moments (or the cumulants) of the parametric pdf under investigation as functions of the unknown parameters, and by estimating the moments as sample-moments, thus formulating the parameter estimation problem as the solution of a (typically non-linear) system of equations. In [38][39] this approach is specialized to non-negative random variables (such as SAR amplitude and intensity), corresponding to pdf defined on $[0, +\infty)$, by redefining MGFs and characteristic functions as Mellin transforms and resulting in a more feasible estimation.

[5] Actually, the bilateral Laplace operator would involve the exponential $\exp(-su)$ [43], but, in the context of statistics, the MGF is usually defined with the exponential $\exp(su)$ as reported in Eq. (27). However, this slight modification has no significant impact on the analytical properties of the resulting transform. Hence, hereafter we will refer to the bilateral Laplace transform as defined by Eq. (27).

Thus, given a non-negative random variable u, the second-kind characteristic function ϕ_u of u is defined as the Mellin transform [49] of the pdf of u, i.e.:

$$\phi_u(s) = \mathcal{M}(p_u)(s) = \int_0^{+\infty} p_u(u) u^{s-1} du, \quad s \in \mathbb{C}, \tag{30}$$

where $\mathcal{M}$ is the Mellin transform on $L^1(0, +\infty)$. Also ϕ_u is known to converge and to be analytical in a vertical strip $\mathcal{S}$ of the complex plane [38][49]. If the interior of the convergence strip contains a neighborhood of 1, then the following definitions are formulated by analogy with the Laplace-based case [38]:

- ν-th order second kind moment: $\mu_\nu = \phi_u^{(\nu)}(1)$, $\nu = 1, 2, \ldots$;
- second kind second characteristic function: $\psi_u(s) = \ln \phi_u(s)$, $s \in \mathcal{S}$;
- ν-th order second kind cumulant: $\kappa_\nu = \psi_u^{(\nu)}(1)$, $\nu = 1, 2, \ldots$

The expressions "log-moments" and "log-cumulants" are also employed for the second kind moments and cumulants, thanks to their relation with the moments of the logarithm of u, i.e. [38][51]:

$$\mu_\nu = E\{(\ln u)^\nu\}, \quad \kappa_1 = \mu_1 = E\{\ln u\}, \quad \kappa_2 = \mathrm{Var}\{\ln u\}, \quad \kappa_3 = E\{(\ln u - \kappa_1)^3\}. \tag{31}$$

Hence, the estimation method of log-cumulants is based on the analytical calculation of log-moments and log-cumulants as functions of the unknown parameters and on the inversion of the resulting equations. Therefore, MoLC estimates are obtained from sample-moments estimates of the log-moments or of the log-cumulants by solving a system of non-linear equations.

The MoLC equations for the pdfs comprising our dictionary can be found in Section 2. The solution of the resulting equations turns out to be feasible and fast for all the considered distributions. Specifically, log-normal does not require a real solution process, since the parameters of this distributions are exactly the first two log-cumulants. SαSGR and Weibull allow an analytical solution of the corresponding system of two equations. Nakagami, GGR, GΓD, Fisher and K-root require a numerical solution procedure, but, thanks to the strict monotonicity properties of the functions involved, this procedure has been proved to be simple and fast for all the three parametric families [35][37][38][28]. In addition, for several of the considered models, good estimation properties have been proved theoretically for the MoLC approach. In particular, the MoLC estimates exhibit a lower variance with respect to the MoM ones for the Nakagami distribution [38], and are consistent for the GGR one [35].

However, as pointed out in [35], the MoLC equations for GGR and K-root can yield no solutions for specific values of $\hat{\kappa}_2$ and $\hat{\kappa}_3$ (the sample estimates of moments $kappa_2$ and κ_3 respectively). In such situations, these parametric families are not compatible with the empirical data distributions and are not considered in the selection of the optimal model.

3.5 Overall architecture of the proposed method

Plugging MoLC together with the optimal model selection procedure in the iterative SEM estimation process, denoting as $p_i^t(\cdot)$ and $p^t(\cdot)$ the resulting t-th step σ_i-conditional and unconditional amplitude pdf estimates, respectively, and formulating SEM explicitly in terms of the histogram $\{h(z) : z = 0, 1, \ldots, Z-1\}$ of the image $\mathcal{I}$, we obtain at the t-th iteration $(t = 0, 1, \ldots)$:

- **E-step:** compute, for any greylevel z and any component σ_i, the posterior probability estimates corresponding to the current pdf estimates, i.e. $(z = 0, 1, \ldots, Z-1,\ i = 1, 2, \ldots, K)$:

$$\tau_i^t(z) = \frac{P_i^t p_i^t(z)}{p^t(z)}, \quad \text{where} \quad p^t(z) = \sum_{i=1}^{K} P_i^t p_i^t(z); \tag{32}$$

- **S-step:** sample the label $s^t(z)$ of each greylevel z according to the current estimated posterior probability distribution $\{\tau_i^t(z) : i = 1, 2, \ldots, K\}$ $(z = 0, 1, \ldots, Z-1)$;

- **MoLC-step:** for each mixture component σ_i, compute the following histogram-based estimates of the mixture proportion and of the first three log-cumulants:

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

$$\kappa_{2i}^t = \frac{\sum_{z \in Q_{it}} h(z)(\ln z - \kappa_{1i}^t)^2}{\sum_{z \in Q_{it}} h(z)}, \quad \kappa_{3i}^t = \frac{\sum_{z \in Q_{it}} h(z)(\ln z - \kappa_{1i}^t)^3}{\sum_{z \in Q_{it}} h(z)}, \tag{33}$$

  where $Q_{it} = \{z : s^t(z) = \sigma_i\}$ is the set of grey levels assigned to the component σ_i $(i = 1, 2, \ldots, K)$; then, solve the corresponding MoLC equations for each parametric family $f_j(\cdot|\xi_j)$ $(\xi_j \in \Xi_j)$ in the dictionary, thus computing the resulting MoLC estimate $\xi_{ij}^t \in \Xi_j (i = 1, 2, \ldots, K,\ j = 1, 2, \ldots, M)$;

- **MS-step** (Model Selection-step)**:** for each mixture component σ_i, compute the log-likelihood of each estimated pdf $f_j(\cdot|\xi_{ij}^t)$ (except, at least, GGR or K-root if the previous step yielded no solutions for the corresponding MoLC equations) according to the data assigned to σ_i:

$$L_{ij}^t = \sum_{z \in Q_{it}} h(z) \ln f_j(z|\xi_{ij}^t) \tag{34}$$

  and define $p_i^{t+1}(\cdot)$ as the estimated pdf $f_j(\cdot|\xi_{ij}^t)$ yielding the highest value of L_{ij}^t $(i = 1, 2, \ldots, K,\ j = 1, 2, \ldots, M)$.

Thus, the proposed selection of an optimal SAR-specific model for each mixture component operatively requires the substitution of the usual ML-based M-step with the described MoLC-step, and the integration in the iterative procedure of a further MS-step, devoted to

the model selection process. The resulting "dictionary-based" SEM method will be denoted hereafter simply as DSEM.

Note that, differently from the previously described general SEM approach, the adopted version of SEM, operating directly on the image histogram, implicitly assigns the same population label to all the image pixels presenting the same grey level. This specific histogram-based approach has been adopted in order to reduce the computation time of the proposed method. Considering, for example, a 1024×1024 pixel-sized image with 8 bpp (bit per pixel), the general SEM approach would involve calculating the posterior probability distribution and sampling a population label for each of the 2^{20} image pixels, whereas the adopted approach directly deals with only $2^8 = 256$ distinct grey levels. In addition, after computing the image histogram, the execution time of the adopted strategy is independent of the image size. Furthermore, we stress that fitting all the eight considered parametric families to each component does not yield a severe increase in the computation time, with respect to the usual single-model approach. In fact, the solution of the MoLC equations is very fast for all the models in the dictionary, thus not involving critical computational issues.

As in the general SEM framework, the resulting sequence of pdf estimates is expected not to converge pointwise nor a.s., but to reach a stationary behavior, thus requiring the definition of a specific procedure to extract a single optimal pdf estimate from the sequence itself. We adopt here the approach suggested in [4], which computes, at any iteration t ($t = 0, 1, \ldots$), the log-likelihood of the current pdf estimate $p_r^t(\cdot)$ over the whole image data set $\mathcal{I}$, i.e.:

$$L_{\mathcal{I}}^t = \sum_{z=0}^{Z-1} h(z) \ln p_r^t(z), \tag{35}$$

and chooses the estimate $p_r^t(\cdot)$ exhibiting the highest likelihood $L_{\mathcal{I}}^t$. This identifies a single estimate aiming at performing a "dictionary-based" ML estimation of the SAR amplitude pdf of the input image.

On the other hand, the log-likelihood cannot be adopted as a criterion to choose the optimal number K of components, due to the monotonic relation between these two quantities [46]. Several different validation functionals have been proposed in the literature as selection criteria for the parameter K, for instance, according to a Bayesian model-based framework [2][3], to discriminant analysis [1][23][31], or to the *minimum message length* (MML) approach [46]. The latter, promoting the idea of penalizing the logliklihood function with respect to the number of components has shown to be useless for our application because the number of observations (i.e. the number of pixels, often of the order of 100000) is far greater than K (commonly below 20) so that the change of K does not noticeably affect the message length. In order to overcome the problem of choosing the number of components K outside the SEM procedure, we plug the estimation into the SEM scheme as follows. We initialize the algorithmwith respect to the form of the histogram. Thus, first we find the number and locations of the histogram modes. The common way to do it is smoothing: the choice of a specific histogram smoothing procedure depends on the complexity of the histogram (specifically, on the level of noise-like behavior), however for all our test images the linear smoothing with window size of 10-20 provided accurate estimation of

the number of modes. Then we initialize the components with respect to the positions of the modes: on the grey levels corresponding to every mode of the histogram we randomly initialize several components (depending on the level of competitiveness). This level of competitiveness directly corresponds to the complexity of this particular image: if we expect to find many distinct types of land-cover there, this parameter should be set higher, say 3-4; whereas for the most of the cases this parameter could be set 2 without any loss in accuracy. The higher this parameter is set the more "thoroughly" the algorithm will tries to find the components in the image. Such an approach is justified by the observation that all the distributions in our dictionary $\mathcal{D}$ have single-mode pdfs. We remind here that the initial estimate K of the number of components for SEM should be an upper bound. With sufficient level of competitiveness the upper bound condition is obviously met.

One more critical issue of any iterative scheme is to define the number of iterations. As mentioned above, the SEM sequence of pdf estimates is expected not to converge pointwise nor a.s., but rather to reach a stationary behavior, so in order to stop it we apply the following stationarity criterion:

$$\sum_{iter=t-k+1}^{t} \left(\sum_{i=1}^{K} |P_i(iter) - P_i(iter-1)| \right) < \epsilon, \tag{36}$$

where K is the number of components, $P_i(iter)$ are the mixing proportions on the $iter^{th}$ step as in (1). So we control the proportion of pixels being reallocated into some other component during the previous k iterations and once this portion goes below some level (ϵ) we stop the iterations.

We stress, in particular, that the proposed pdf estimation method for SAR amplitude data turns out to be automatic. In fact, the selection of the number of mixture components, the choice of the optimal model for each component and the estimation of the parameters of the model are jointly and automatically performed by the algorithm without any need for user intervention. However, several parameters such as smoothing window size and the stationarity criterion constant ϵ have s strong impact on the runtime/accuracy ratio, so they need to be defined correctly for the optimal performance.

4 Experimental results

4.1 Data sets for experiments

The proposed DSEM algorithm for pdf estimation has been tested on a set of high-resolution SAR images, and compared with several usual SAR-specific parametric pdf estimation strategies.

The first seven images used for experiments are the portions of a 16 bpp singlelook X-band SAR-image with 5 meter esolution acquired in 2008 over the region of Piemonte (Italy) by COSMO-SkyMed (©Italian Space Agency):

- "River_ponds", 1400×1400 pixels;
- "Small_river", 300×500 pixels;
- "Mountain_lake1" and "Mountain_lake2", 400×400 pixels;
- "Mountain_town", 2200×1700 pixels;
- "Mountains8" and "Mountains16", both being the same image, the former 8 bpp, the latter 16 bpp, for the sake of comparison, 3000×1400

These portions have been selected in order to test the proposed method on different typologies of data histograms.

The last test image (700×700 pixels, 8 bpp) was acquired by the RAMSES airborne sensor, we further refer to it as "Ramses" for simplicity. This image was provided by the French Space Agency (©ONERA-CNES).

We stress, in particular, that some of the images exhibit bimodal histograms, whereas the others have fairly simple unimodal histograms. All images are shown in Appendix A, after histogram stretching and/or equalization, and all the corresponding histograms are reported in Appendix B.

4.2 Pdf estimation results

The proposed DSEM method has been applied to all the considered images and the resulting pdf estimates have been assessed both quantitatively, by computing their correlation coefficients with the image histograms, and qualitatively, by visually comparing the plots of the estimates and of the histograms. Appendix B shows all the plotted comparisons for all the 11 images.

Before presenting the results, several remarks about the implementation of our algorithm should be done. Firstly, When we deal with the 16 bpp images, in order to minimize the computational burden, according to the nature of the data, we apply the following procedure: instead of working with the whole 100% bins of the histogram, we work only with 99,9% of the data, the part located within the 99,9% quantile. This almost does not affect the estimation accuracy, while sharly reducing the computational burden. This is due to the fact

Image	ρ	$KS-dist$	K^*
River_ponds	99,95%	0,0001	3
Small_river	99,80%	0,0067	5
Mountain_lake1	99,91%	0,0011	4
Mountain_lake2	99,90%	0,0008	4
Mountain_town	99,89%	0,0049	4
Mountain8	99,70%	0,0082	3
Mountain16	99,80%	0,0017	3
Ramses	99,62%	0,0070	4

Table 1: Results for the DSEM algorithm applied to all the employed SAR images: correlation coefficient ρ and Kolmogorov-Smirnov distance $KS-dist$ between the estimated pdf and the image histogram and optimal number K^* of mixture components.

that for all the images in our test set, the data in histograms was compactly concentrated close to the origin, i.e. close to zero, so usually the 99,9% quantile was in range $300 \div 1500$, thus decresing the number of bins under study from 65536 to around 1000, thus reducing the computation weight 65 times. Secondly, the histogram of "Ramses" has shown irregularities, having picks for minimum intensity (shadows) and for maximum intensity (reflection); due to the fact that this kind of local effects (very high and very thin, with width 1, modes of histogram) cannot be adequately modeled by means of classical absolutely continuous pdfs (all pdfs in our dictionary), these spikes of histogram have been manually removed.

The correlation coefficients between the resulting estimated pdfs and the image histograms are very high (always greater than 99,5%) for all the considered images (see Table 1), thus assessing the effectiveness of the proposed method from the viewpoint of the estimation accuracy. The visual comparison between the pdf estimates and the corresponding image histograms confirms this conclusion, as shown, for example, in Fig. 1.

In order to further assess the capabilities of the method, a comparison has also been performed with several other standard parametric models for SAR amplitude data. Specifically, we present here the six (either theoretical or empirical) best-performing models from our dictionary, the corresponding parameters for all models have been estimated by using MoLC. The resulting correlation coefficients are listed in Table 2. A comparison between Tables 1 and 2 shows that the proposed DSEM algorithm yields the pdf estimate with the highest correlation coefficients with the image histograms of all images.

In particular, the good results achieved on "River_ponds", "Mountain_lake1" and "Mountain_lake2" which exhibit bimodal histograms, stress the usefulness of the adopted FMM approach. As shown in Fig. 1 for "Mountain_lake1", DSEM effectively describes the bimodal statistics; on the contrary, all the parametric models considered in Table 2 are intrinsically unimodal and provide poor estimates. We stress, in particular, the relevance, in this case, of the automatic selection of an optimal number K^* of components.

On the other hand, we also note from the comparison between Tables 1 and 2 that for most of the remaining images, presenting a unimodal histogram, the difference between the

Image	Theoretical models				Empirical models		
	GGR	Nakagami	HT-Rayleigh	K-root	Log-Normal	Weibull	GGamma
River_ponds	not defined	98,86%	93,06%	98,05%	93,84%	99,06%	98,21%
Small_river	not defined	98,66%	92,33%	not defined	96,83%	98,01%	98,74%
Mountain_lake1	78,82%	77,12%	73,00%	not defined	86,20%	77,62%	79,12%
Mountain_lake2	94,27%	97,14%	92,46%	not defined	96,40%	96,73%	97,40%
Mountain_town	not defined	97,87%	88,49%	not defined	97,57%	96,64%	88,32%
Mountain8	89,70%	93,30%	86,20%	95,30%	99,33%	92,52%	94,93%
Mountain16	90,43%	94,10%	88,43%	not defined	99,41%	93,34%	92,11%
Ramses	94,21%	93,65%	94,14%	not defined	73,68%	93,80%	90,09%

Table 2: Correlation coefficients between the estimated pdfs and the image histograms for seven parametric families and all SAR images.

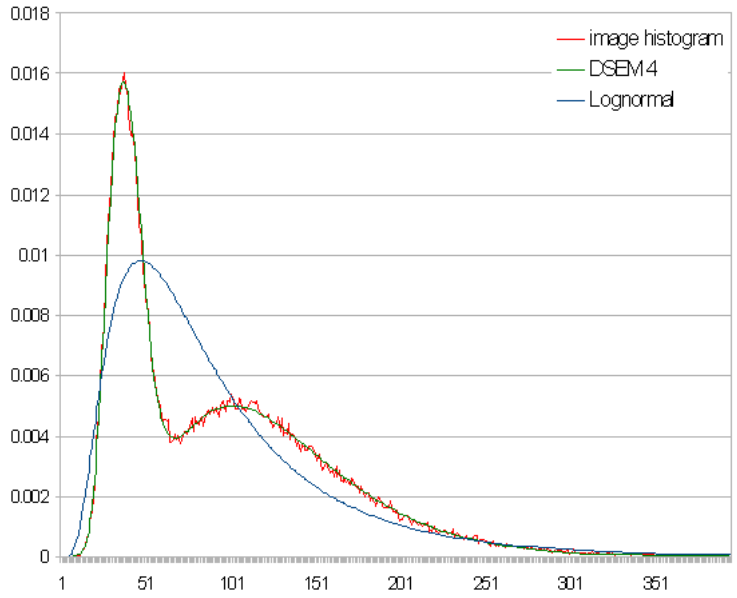

Figure 1: Plot of the image histogram and of the DSEM pdf estimates with $K = 4$ and the best fitting pdf from the dictionary for the "Mountain_lake1" data image.

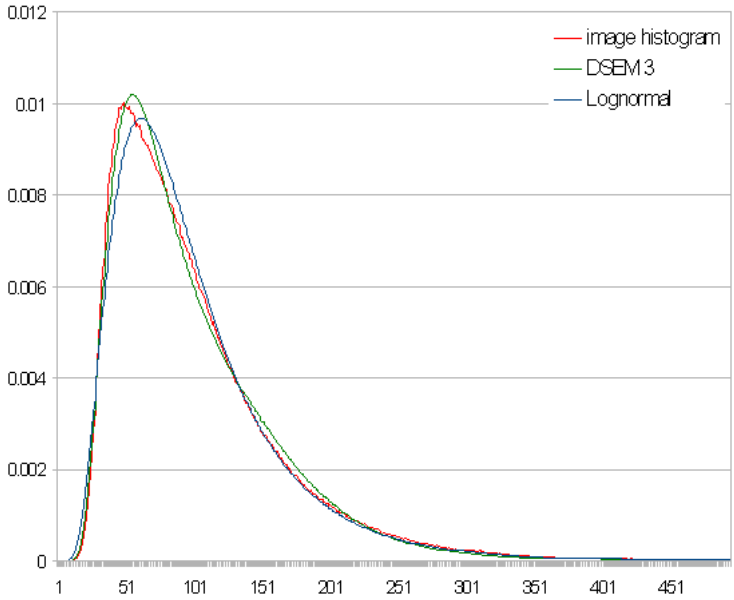

Figure 2: Plot of the image histogram and of the DSEM and lognormal pdf estimates for the "Mountain16" data image.

DSEM accuracy and the one achieved by the best single pdf estimate is not as big as for the multimodal cases. In particular, at least one among the Nakagami, GGR, log-normal, Weibull and GGamma distributions allows to obtain very accurate estimation results, though worse than the DSEM result. In these cases, thanks to the unimodal empirical distribution, a FMM-based approach is not mandatory and a single-component pdf estimate turns out to be quite efficient. For instance, in Fig. 2 we plot for the "Mountain16" image the pdf estimates provided by DSEM and by the best performing single parametric model for this image (namely, lognormal).

5 Conclusions

In this research report, an efficient finite mixture model (FMM) estimation algorithm has been developed for high resolution SAR amplitude data pdf, by integrating the SEM and the MoLC methods with an automatic technique to select, for each mixture component, an optimal parametric model inside a predefined dictionary of parametric pdfs. In particular, the developed estimation strategy is explicitly focused on the context of high resolution SAR image analysis and correspondingly a set of eight theoretic or empirical models for SAR amplitude data (i.e., Nakagami, log-normal, generalized Gaussian Rayleigh, SαS generalized

Rayleigh, Weibull, K-root, Fisher and generalized Gamma) has been adopted as a dictionary. The method generalizes to high resolution the previously designed approach [34] for low resolution SAR.

The numerical results of the application of the method to several real SAR images acquired by COSMO-SkyMed and airborne RAMSES sensor prove the proposed DSEM algorithm to provide very accurate pdf estimates, both from the viewpoint of a visual comparison between the estimates and the corresponding image histograms, and from the viewpoint of the quantitative correlation coefficient between them. We stress, in particular, that the method proves to be effective on all the considered images, despite of their different statistics (i.e., histogram unimodality or multimodality) and high heterogeneity. Correlation coefficients higher that 99% are obtained, in fact, in all cases, thus proving the flexibility of the method.

Specifically, the use of a mixture model is mandatory when dealing with multimodal statistics. Applied to the "River_ponds", "Mountain_lake1" and "Mountain_lake2" images, which exhibit a bimodal histogram, the developed DSEM algorithm correctly detects positions and sizes of both statistical modes. On the other hand, the results provided by DSEM in case of unimodal histograms usually provide only minor improvement as compared to the best single-component parametric models included in the dictionary.

The experiments have suggested the generalized Gaussian Rayleigh, Lognormal and generalized Gamma models to be the most effective single-component parametric families, providing higher correlation and often being the best fitting single component models. However, the analysis of more complicated images has shown that in case of more irregular histograms (multimodal, spiky) the single-component models from the dictionary fail in the task of accurate estimation (e.g. the "Mountain_lake1" image). That can be explained to some extent for the theoretical models as they are not designed to deal with multi-landcover and heterogenous images. This fact stresses out the necessity to work with mixture models as effective tools to achieve reasonable level of accuracy.

We stress here that the proposed DSEM algorithm is automatic, by performing both the FMM estimation process and the optimization of the number of mixture components without any need for user interaction. In addition, thanks to the specific histogram-based version of SEM it adopts, the computation time of DSEM is almost independent of the image size. These interesting operational properties, together with the estimation accuracy it provides for all the different considered images prove DSEM to be a flexible and effective pdf estimation tool for SAR data analysis.

Acknowledgments

This research has been conducted within a collaboration between the research team ARIANA of the Institut National de Recherche en Informatique et Automatique (INRIA) Sophia Antipolis, France, and the Dept. of Biophysical and Electronic Engineering (DIBE) of the University of Genoa, Italy.

It was carried out with the financial support of French Space Agency (CNES), whose support is gratefully acknowledged. The authors would also like to thank the Italian Space Agency (ASI) for providing the COSMO-SkyMed images of Piemonte (©ASI, 2008) and CNES for providing the RAMSES image (©ONERA-CNES, 2004).

Appendix

A SAR images employed for experiments

The present Appendix shows the 8 images used in the experiments (Figs. 1-7). Only for visualization purposes, their histograms have been stretched and/or equalized.

Figure 1: "River_ponds" image (©ASI).

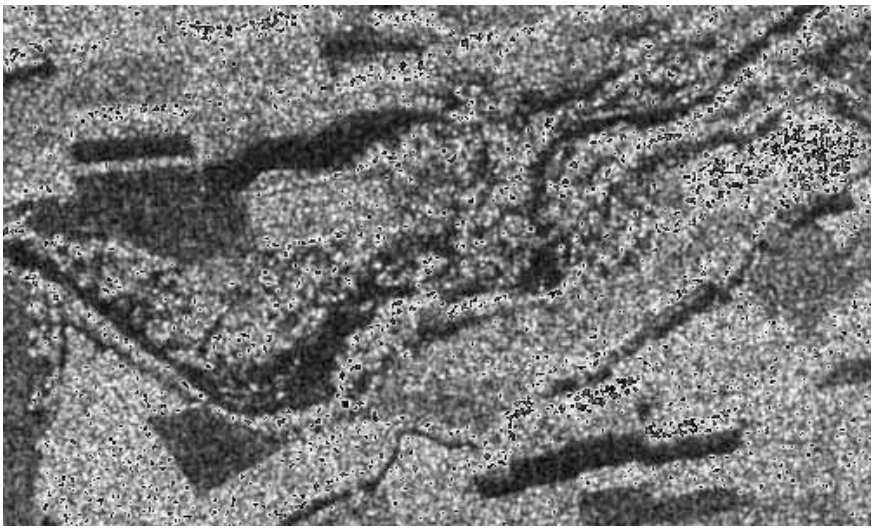

Figure 2: "Small_river" image (©ASI).

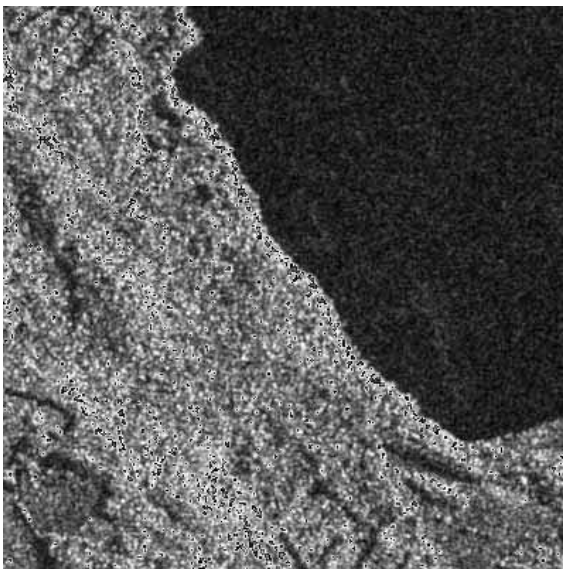

Figure 3: "Mountain_lake1" image (©ASI).

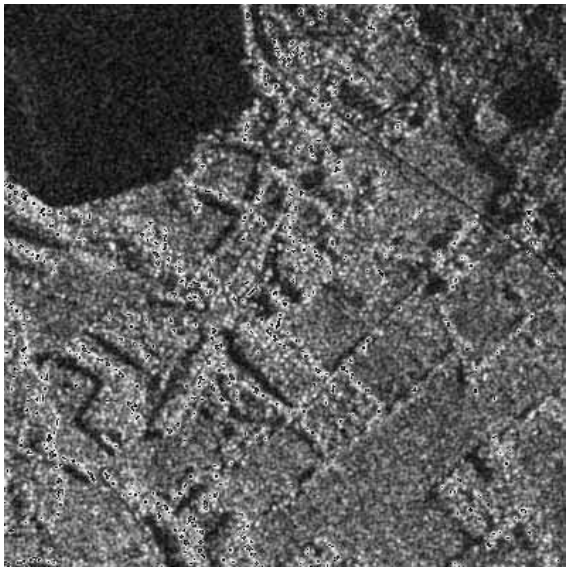

Figure 4: "Mountain_lake2" image (©ASI).

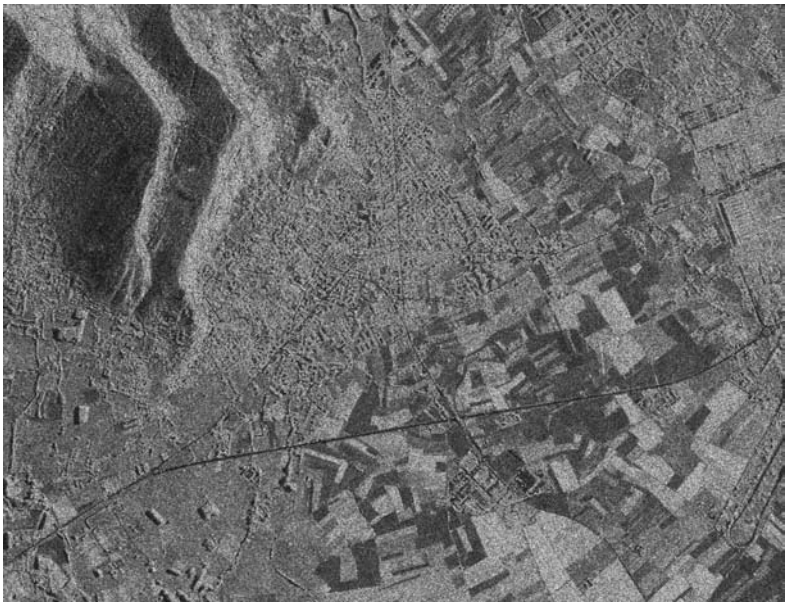

Figure 5: "Mountain_town" image (©ASI).

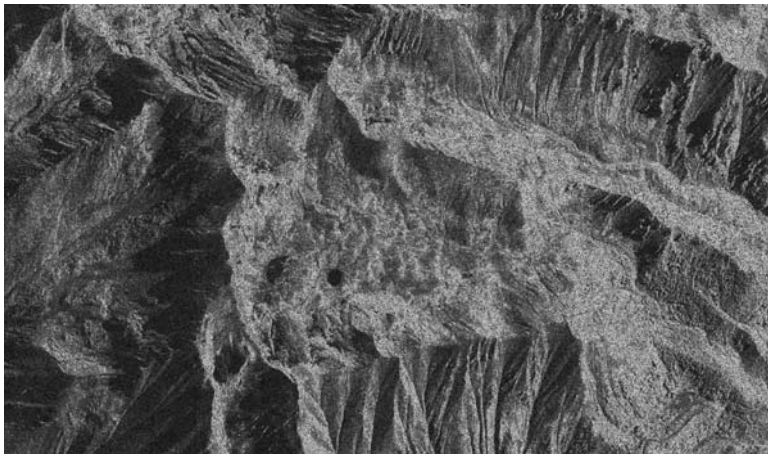

Figure 6: "Mountain8"/"Mountain16" image (©ASI).

Figure 7: "Ramses" image (©ONERA-CNES).

B Plots of the estimated pdfs

In Appendix B we report the estimated pdfs provided by the proposed DSEM algorithm (Figs. 1-8), together with the histograms of the corresponding images and the histogram of the best fitting pdf from the dictionary.

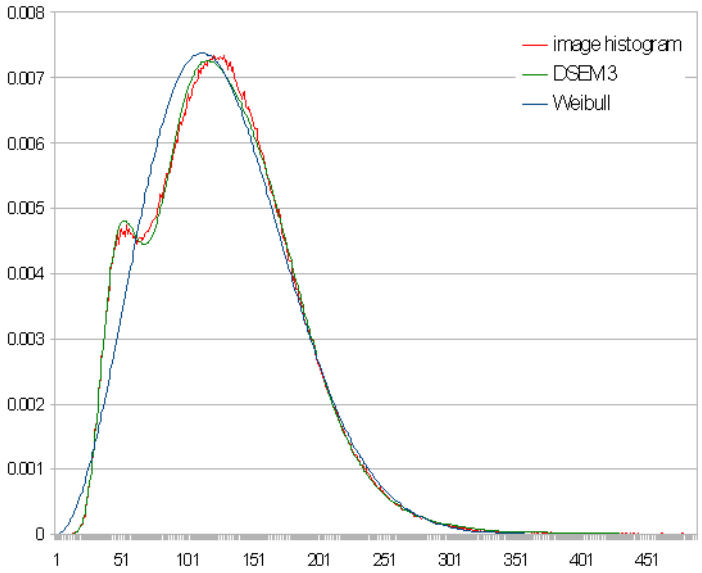

Figure 1: Plot of the histograms for the "River_ponds" image.

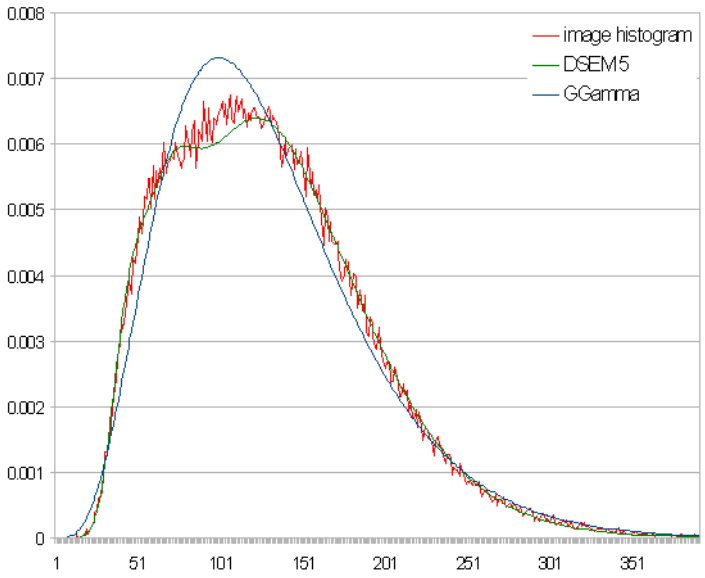

Figure 2: Plot of the histograms for the "River_ponds" image.

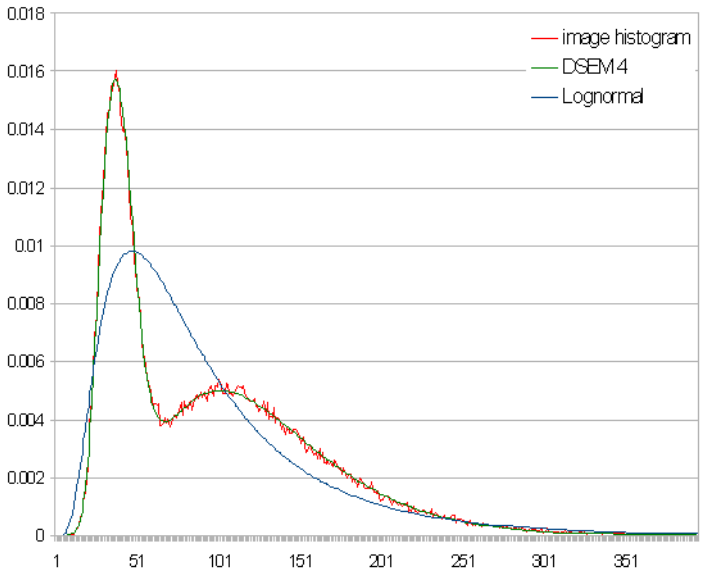

Figure 3: Plot of the histograms for the "Mountain_lake1" image.

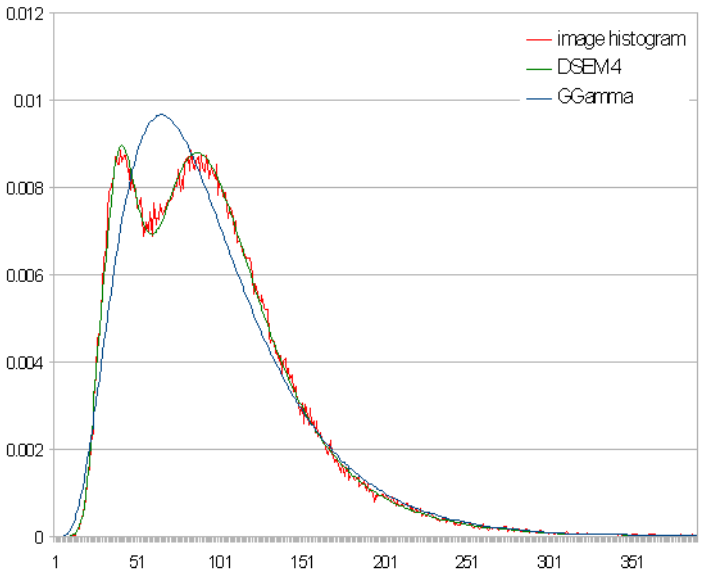

Figure 4: Plot of the histograms for the "Mountain_lake2" image.

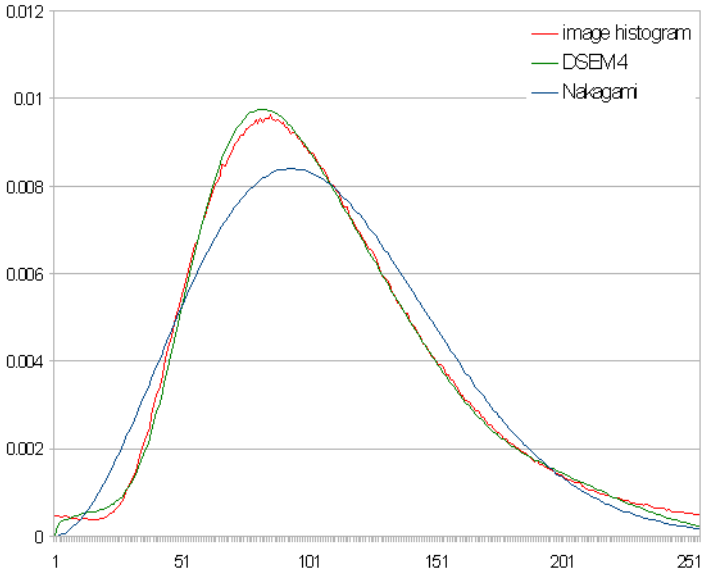

Figure 5: Plot of the histograms for the "Mountain_town" image.

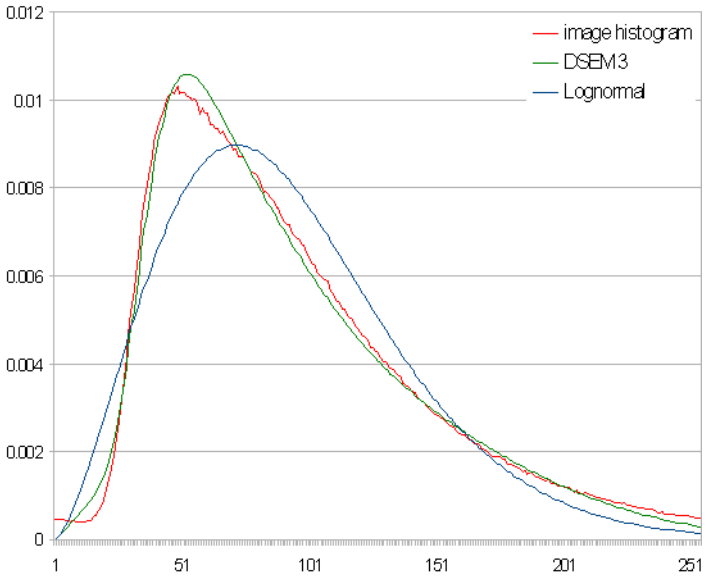

Figure 6: Plot of the histograms for the "Mountain8" image.

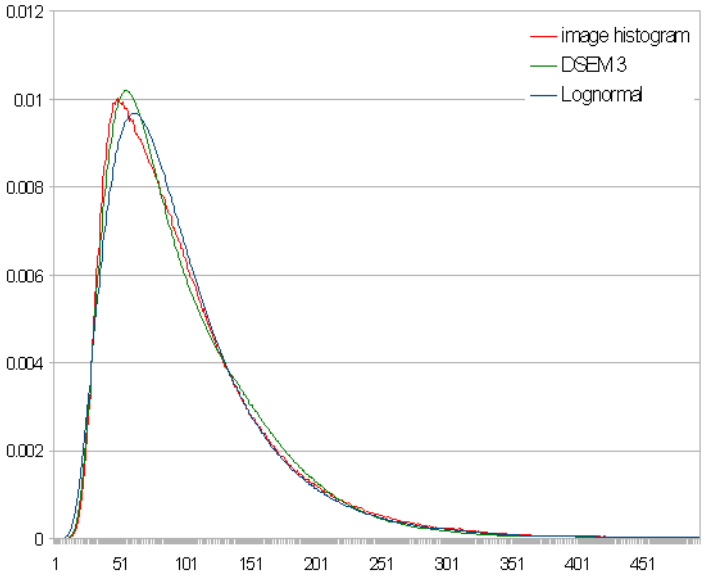

Figure 7: Plot of the histograms for the "Mountain16" image.

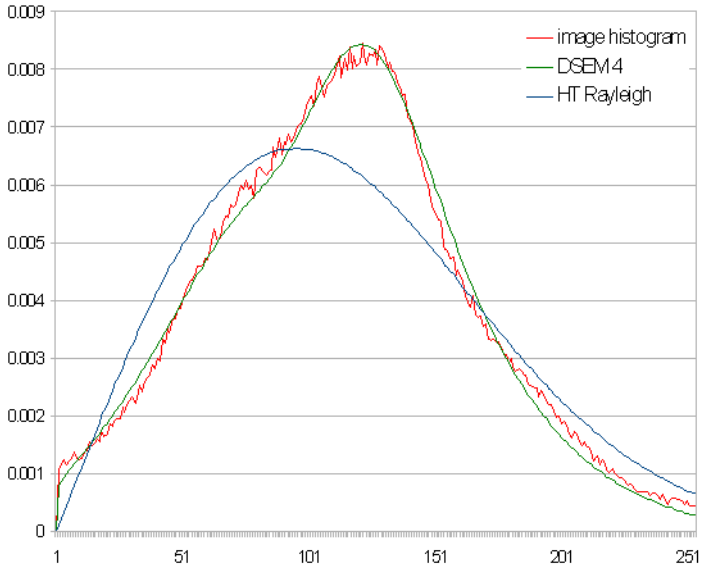

Figure 8: Plot of the histograms for the "Ramses" image.

References

[1] S. Bandyopadhyay and U. Maulik, *Nonparametric genetic clustering: comparison of validity indices*, IEEE Transactions on Systems, Man, and Cybernetics **31** (2001), no. 1, 120–125.

[2] C. Biernacki, G. Celeux, and G. Govaert, *An improvement of the NEC criterion for assessing the number of clusters in a mixture model*, Pattern Recognition Letters **20** (1999), 267–272.

[3] ———, *Assessing a mixture model for clustering with the integrated completed likelihood*, IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000), no. 7, 719–725.

[4] ———, *Strategies for getting highest likelihood in mixture models*, Research Report 4255, INRIA, September 2001.

[5] C. M. Bishop, *Neural networks for pattern recognition*, 2nd edition, Oxford University Press, 1996.

[6] F. Bowman, *Introduction to Bessel functions*, Dover Publications, New York, 1968.

[7] L. Bruzzone, M. Marconcini, U. Wegmuller, and A. Wiesmann, *An advanced system for the automatic classification of multitemporal SAR images*, IEEE Transactions on Geoscience and Remote Sensing **42** (2004), no. 6, 1321–1334.

[8] G. Celeux, D. Chauveau, and J. Diebolt, *On stochastic versions of the EM algorithm*, Research Report 2514, INRIA, March 1995.

[9] G. Celeux, S. Chretien, F. Forbes, and A. Mkhadri, *A component-wise EM algorithm for mixtures*, Research Report 3746, INRIA, August 1999.

[10] M. Cheney, *A mathematical tutorial on synthetic aperture radar*, SIAM Review **43** (2001), no. 2.

[11] ———, *An introduction to synthetic aperture radar (SAR) and SAR interferometry*, pp. 167–177, in "Approximation theory X: wavelets, splines, and applications", C. K. Chui, L. L. Schumacher, and J. Stockler editors, Vanderbilt University Press, Nashville, TN, U.S.A., 2002.

[12] Y. Delignon, A. Marzouki, and W. Pieczynski, *Estimation of generalized mixtures and its application to image segmentation*, IEEE Transactions on Image Processing **6** (2001), no. 10, 1364–1375.

[13] Y. Delignon and W. Pieczynski, *Modelling non-Rayleigh speckle distribution in SAR images*, IEEE Transactions on Geoscience and Remote Sensing **40** (2002), no. 6, 1430–1435.

[14] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data and the EM algorithm*, Journal of the Royal Statistical Society (1977), no. 39, 1–38.

[15] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, 2nd edition, Wiley, New York, 2001.

[16] T. Eltoft, *Speckle modeling and filtering*, Proc. NORSIG, Bergen, Norway, 2003.

[17] A. C. Frery, H.-J. Muller, C. C. F. Yanasse, and S. Sant'anna, *A model for extremely heterogeneous clutter*, IEEE Transactions on Geoscience and Remote Sensing **35** (1997), no. 3, 1017–1028.

[18] K. Fukunaga, *Introduction to statistical pattern recognition*, 2nd edition, Academic Press, 1990.

[19] Q. Jackson and D. A. Landgrebe, *An adaptive classifier design for high-dimensional data analysis with a limited training data set*, IEEE Transactions on Geoscience and Remote Sensing **39** (2001), no. 12, 2664–2679.

[20] E. Jakeman and P. N. Pusey, *A model for non-Rayleigh sea echo*, IEEE Transactions on Antennas and Propagation **24** (1976), 806–814.

[21] I.-R. Joughin, D. P. Winebrenner, and D. B. Percival, *Probabifty density functions for multilook polarimetric signatures*, IEEE Transactions on Geoscience and Remote Sensing **32** (1994), no. 3.

[22] G. K. Karagiannidis, D. A. Zogas, and S. A. Kotsopoulos, *On the multivariate Nakagami-m distribution with exponential correlation*, IEEE Transactions on Communications **51** (2003), no. 8, 1240–1244.

[23] R. Kothary and D. Pitts, *On finding the number of clusters*, Pattern Recognition Letters **20** (1999), 405–416.

[24] B.-C. Kuo and D. A. Landgrebe, *A covariance estimator for small sample size classification problems and its application to feature extraction*, IEEE Transactions on Geoscience and Remote Sensing **40** (2002), no. 4, 814–819.

[25] E. E. Kuruoglu and J. Zerubia, *Modelling SAR images with a generalization of the Rayleigh distribution*, Research Report 4121, INRIA, February 2001.

[26] ———, *Skewed α-stable distributions for modelling textures*, Pattern Recognition Letters **24** (2003), 339–348.

[27] 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 Transactions on Geoscience and Remote Sensing **32** (1994), no. 5, 648–659.

[28] H.-C. Li, W. Hong, and Y.-R. Wu, *Generalized gamma distribution with molc estimation for statistical modeling of sar images*, Proceedings of APSAR 2007. 1st Asian and Pacific Conference, 2007, pp. 525–528.

[29] P. Mantero, G. Moser, and S. B. Serpico, *Partially supervised classification of remote sensing images using svm-based probability density estimation*, IEEE honorary workshop for Prof. D. A. Landgrebe, 27-28 October 2003, 2003.

[30] P. Masson and W. Pieczynski, *SEM algorithm and unsupervised statistical segmentation of satellite images*, IEEE Transactions on Geoscience and Remote Sensing **31** (1993), no. 3, 618–633.

[31] U. Maulik and S. Bandyopadhyay, *Performance evaluation of some clustering algorithms and validity indices*, IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002), no. 12, 1650–1654.

[32] T. K. Moon, *The Expectation-Maximization algorithm*, IEEE Signal Processing Magazine **13** (1996), no. 6, 47–60.

[33] G. Moser, *Development of unsupervised change detection methods for remote sensing images*, "laurea" degree thesis, University of Genoa, November 2001.

[34] G. Moser, S. Serpico, and J. Zerubia, *Dictionary-based stochastic expectation-maximization for SAR amplitude density probability function estimation*, IEEE Transactions on Geoscience and Remote Sensing **44** (2006), no. 1, 188–199.

[35] G. Moser, J. Zerubia, and S. B. Serpico, *SAR amplitude probability density function estimation based on a generalized gaussian model*, IEEE Transactions on Geoscience and Remote Sensing **15** (2006), 1429–1442.

[36] H.-J. Muller and R. Pac, *G-statistics for scaled sar data*, Proc. IEEE Geoscience and Remote Sensing Symp., vol. 2, 1999, pp. 1297–1299.

[37] J.-M. Nicolas, *Introduction aux statistiques de deuxième espéce: application aux lois d'images RSO (introduction to second kind statistics: applications to SAR images laws)*, Research Report (in french) 2002D001, ENST, Paris, February 2002.

[38] ———, *Introduction aux statistiques de deuxième espéce: applications des logs-moments et des logs-cumulants a l'analyse des lois d'images radar*, Traitement du Signal (in french) **19** (2002).

[39] J.-M. Nicolas and F. Tupin, *Gamma mixture modeled with "second kind statistics": application to SAR image processing*, Proceedings of the IGARSS Conference, Toronto (Canada), 2002.

[40] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar images*, Artech House, Norwood, 1998.

[41] C. J. Oliver, *A model for non-rayleigh scattering statistics*, Opt. Acta **31** (1984), no. 6, 701–722.

[42] ———, *Correlated K-distributed scattering model*, Opt. Acta **32** (1985), 1515–1547.

[43] A. Papoulis, *Probability, random variables, and stochastic processes*, 3rd edition, McGraw-Hill International Editions, 1991.

[44] M. Petrou, F. Giorgini, and P. Smits, *Modelling the histograms of various classes in SAR images*, Pattern Recognition Letters **23** (2002), 1103–1107.

[45] R. A. Redner and H. F. Walker, *Mixture densities, maximum likelihood, and the EM algorithm*, SIAM Review **26** (1984), no. 2, 195–239.

[46] ———, *Unsupervised learning of finite mixture models*, IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002), no. 3, 381–396.

[47] J.A. Richards and X. Jia, *Remote sensing digital image analysis*, Springer-Verlag, Berlin, 1999.

[48] W. Rudin, *Principles of mathematical analysis*, 2nd edition, McGraw-Hill, 1976.

[49] I. Sneddon, *The use of integral transforms*, McGraw-Hill, New York, 1972.

[50] S. Tadjudin and D. A. Landgrebe, *Robust parameter estimation for mixture model*, IEEE Transactions on Geoscience and Remote Sensing **38** (2000), no. 1, 439–445.

[51] C. Tison, J.-M. Nicolas, and F. Tupin, *Accuracy of Fisher distributions and log-moment estimation to describe histograms of high-resolution SAR images over urban areas*, Proceedings of the IGARSS Conference, July 21-25, Toulouse (France), 2003.

[52] 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 Transactions on Geoscience and Remote Sensing **42** (2004), no. 10, 2046–2057.

[53] H. L. Van Trees, *Detection, estimation and modulation theory*, vol. 1, John Wiley & Sons., New York, 1968.

[54] J. Weston, A. Gammerman, M. Stitson, V. Vapnik, V. Vovk, and C. Watkins, *Support vector density estimation*, pp. 293–306, in "Advances in Kernel Methods Support Vector Learning", B. Schölkopf, C. J. C. Burges, and A. J. Smola editors, MIT Press, Cambridge, MA, U.S.A., 1999.

[55] E. Wong and B. Hajek, *Stochastic processes in engineering systems*, Springer-Verlag, NewYork, 1985.

[56] C. F. J. Wu, *On the convergence properties of the EM algorithm*, The Annals of Statistics **11** (1983), 95–103.

[57] S. H. Yueh and J. A. Kong, *K distribution and polarimetric terrain radar clutter*, J. Electromagn. Waves Applicat. **3** (1999), no. 8, 747–768.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399