Dictionary-Based Stochastic Expectation–Maximization for SAR Amplitude Probability Density Function Estimation

Gabriele Moser, Josiane Zerubia, Sebastiano B. Serpico

Dictionary-Based Stochastic Expectation– Maximization for SAR Amplitude Probability Density Function Estimation

Gabriele Moser, *Member, IEEE*, Josiane Zerubia, *Fellow, IEEE*, and Sebastiano B. Serpico, *Senior Member, IEEE*

Abstract—In remotely sensed data analysis, a crucial problem is represented by the need to develop accurate models for the statistics of the pixel intensities. This paper deals with the problem of probability density function (pdf) estimation in the context of synthetic aperture radar (SAR) amplitude data analysis. Several theoretical and heuristic models for the pdfs of SAR data have been proposed in the literature, which have been proved to be effective for different land-cover typologies, thus making the choice of a single optimal parametric pdf a hard task, especially when dealing with heterogeneous SAR data. In this paper, an innovative estimation algorithm is described, which faces such a problem by adopting a finite mixture model for the amplitude pdf, with mixture 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, and of optimization of the number of components by combining the stochastic expectation–maximization iterative methodology with the recently developed "method-of-log-cumulants" for parametric pdf estimation in the case of nonnegative random variables. Experimental results on several real SAR images are reported, showing that the proposed method accurately models the statistics of SAR amplitude data.

Index Terms—Finite mixture models (FMMs), parametric estimation, probability density function (pdf) estimation, stochastic expectation maximization (SEM), synthetic aperture radar (SAR) images.

I. INTRODUCTION

IN 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 [12], [13], [41], [47] data, this modeling process turns out to be a crucial task, for instance, to attain classification [18], denoising [41], or target-detection [41] goals.

From a methodological viewpoint, either parametric or nonparametric estimation strategies can be used for this task [18], [21], [54]. Specifically, a parametric approach postulates a given mathematical model for each class-conditional probability density function (pdf) and formulates the pdf estimation

problem as a parameter estimation problem. Several strategies have been proposed in the literature to deal with parameter estimation, e.g., the maximum-likelihood methodology [18] and the "method of moments" [41], [46], [53]. On the contrary, nonparametric pdf estimation approaches do not assume any specific analytical model for the unknown pdf, thus providing a higher flexibility, although usually involving internal architecture parameters to be set by the user [18]. In particular, several nonparametric kernel-based estimation and regression architectures have been described in the literature, that have proved to be effective estimation tools, such as standard Parzen window estimators [18], [43] artificial neural networks [7], and support vector machines (SVMs) [31], [55], [56].

In the present paper, we address the problem of the estimation of the pdf of SAR amplitude data. Several different theoretical and heuristic models for the pdfs of SAR data have been proposed in the literature. For instance, the Rayleigh [41] distribution has been proposed as a theoretical amplitude pdf based on a Gaussian model for the backscattering phenomena [41] involved in the generation of a single-look SAR image. The Nakagami–Gamma [41], [53] distribution generalizes the single-look Rayleigh model to multilook data and is a usually accepted model for nontextured image areas [41]. The K distribution [41] for SAR intensity data is obtained by a different model for the statistics of the backscattering phenomena [16], [23], [24] and is known to correctly describe the statistics of highly textured image areas [41]. The log-normal and the Weibull distributions [41] have been introduced as heuristic models of the amplitude or of the intensity statistics, and successfully applied to images of urban areas (log-normal) and of ocean, land, and sea-ice areas (Weibull) [41]. The Fisher distribution has been adopted as an empirical model for the SAR statistics over high-resolution urban regions [53], and the Pearson system of parametric families has been applied for SAR image segmentation purposes [15]. In [29], symmetric α-stable distributions [27], [28] have been employed to describe SAR backscattering phenomena, and the resulting amplitude pdfs have proved to correctly model the amplitude statistics over urban areas.

Hence, different parametric families among the above-mentioned ones have turned out to be effective models for different land cover typologies [41], which make the choice of a single optimal SAR amplitude parametric pdf a hard task. In addition, a remotely sensed image can, in general, show a varied scene, presenting several distinct land cover typologies.

In this paper, a SAR amplitude estimation algorithm is proposed, that addresses these problems by adopting a finite mixture model [19], [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 statistical population [16], [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, and we assume each component to belong to a given dictionary of SAR-specific pdfs, namely, the well-known log-normal, Nakagami–Gamma, Weibull, and K distributions, together with the symmetric-α-stable generalized Rayleigh pdf [29] and the recently proposed generalized Gaussian Rayleigh model [36], [38].

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) pdf estimation algorithm [6], [11], [32], [49], [57] with the "method-of-log-cumulants" (MoLC) [53]. In addition, the developed method automatically performs an optimal choice of the number K of mixture components by computing, for each value of K in a predefined search range, the corresponding pdf estimate and by choosing the estimate exhibiting the highest correlation coefficient with the empirical data distribution (i.e., the image histogram).

The novelty of the paper lies in the combination of SEM with a dictionary-based approach, which aims at improving the accuracy and the flexibility of the estimation process, and in the integration of MoLC into the SEM iterative procedure, which allows one to exploit, in the proposed estimation scheme, the good theoretical properties exhibited by MoLC, when applied to most SAR-specific pdfs [36], [53].

The proposed approach has been validated by using several real ERS-1, XSAR, E-SAR, and airborne SAR images. The experimental results show that the method accurately models the amplitude distributions of all the images considered, from both a qualitative viewpoint (i.e., visual comparison between the data histogram and the estimated pdf) and a quantitative viewpoint (i.e., correlation coefficient and Kolmogorov–Smirnov distance between the data histogram and the estimated pdf), thus proving its effectiveness and flexibility.

In Section II, an overview of the previous work about finite mixture models in remote sensing is provided, and in Section III, the proposed estimation scheme is presented and the SEM method is described. Section IV reports the results of the application of the proposed approach to the statistical modeling of the gray levels of several real SAR amplitude images, and points out that the method fits the amplitude distribution better than several parametric models developed for this purpose. Finally, conclusions are drawn in Section V.

II. PREVIOUS WORK ON FINITE MIXTURE MODELS

Finite mixture models (FMMs) have been widely used in the pattern recognition [4], [19], [46] and remote sensing [15], [30] literature, thanks to their ability to model an unknown pdf as a linear combination of parametric mixture components, each representing, for instance, the conditional statistics of a given land-cover typology, or, more generally, a single mode of a multimodal distribution. Specifically, the problem of FMM parametric estimation, i.e., the computation of a parameter vector

optimally representing the observed image data, has been addressed by using several different approaches [19], [46]. The computation of maximum-likelihood (ML) estimates would involve the maximization of the log-likelihood function of the whole image data, but the solution of this maximization problem is not feasible analytically [19] and also involves several numerical difficulties, due, for instance, to the usual presence of several local maxima [46].

In order to solve this issue, the use of the expectation–maximization (EM) algorithm has been proposed [17], [34], [46], which formalizes the problem of the estimation of the parameters of a mixture as an incomplete data problem and introduces a sequence of parameter estimates by iteratively maximizing a pseudolikelihood function [17], [46]. In the context of mixture densities, EM has been proved to converge (under mild assumptions) to a stationary point of the log-likelihood function [17], [46], [59], although it does not converge, in general, to a global maximum point, and sometimes requires long convergence times [9], [19]. In addition, the maximization of the pseudolikelihood function, even if analytically tractable in several applications [34] (e.g., mixtures of exponential families [46]), does not always yield a closed-form solution (e.g., when a K or a Weibull model is assumed for the mixture components [41]); moreover, the convergence point may belong to the boundary of the parameter space, thus possibly involving analytical singularities [19].

Several variants of EM have been introduced in order to overcome these difficulties. In [22], [26], and [52], modified versions of EM, as well as regularized covariance estimators, are proposed in order to increase the robustness of the estimation process in the context of Gaussian mixture models for hyperspectral data classification. A simplified version of EM, named classification EM [6], has been developed, that 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," is described in [10]: it aims at reducing the computation time and also at avoiding analytical singularities [19]. The stochastic EM (SEM) [6], [11], [32], [49] avoids the computation of the pseudolikelihood function and the related analytical maximization issues by integrating a stochastic sampling procedure into the estimation process. Hence, the sequence of parameter estimates generated by SEM is a discrete-time random process, that does not converge pointwise nor almost surely, 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 [9]. Simulated annealing EM [9] is a combination of EM with SEM that allows SEM to converge also almost surely [42], [58], although to a local maximum and requiring the preliminary definition of a suitable annealing schedule. Monte Carlo EM and simulated annealing Monte Carlo EM are further stochastic variants of EM that replace the computation of the pseudolikelihood function with a Monte Carlo sampling procedure and almost surely converge, under mild assumptions, to a local maximum [9]. In [15] the "iterated conditional expectation" method is used as an alternative to EM and SEM to perform mixture estimation for image segmentation and clustering purposes. In [19], the minimum message length principle is applied in order to formu-

late a mixture-based unsupervised parametric estimator, aimed at avoiding convergence to the boundary of the parameter space, as well as at improving, as compared with EM, robustness to initialization, but it still involves the problem of the pseudolikelihood maximization.

III. DICTIONARY-BASED FINITE MIXTURE MODEL FOR SAR DATA PDF ESTIMATION

A. Overview of the Proposed Method

In order to take explicitly into account the possible presence, in a given SAR amplitude image $\mathcal{I}$ [41], of several distinct land-cover typologies yielding different contributions to the statistics of the pixel intensities, we assume the use of a finite mixture model for the pdf. Specifically, we model a SAR amplitude image composed of N pixels 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 probability density function (pdf):[1]

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

where $p_i(\cdot|\theta_i)$ is a density function dependent on a vector θ_i of parameters, taking values from a set $\Theta_i \subset \mathbb{R}^{\ell_i}$ (ℓ_i being the number of parameters for the ith component; $i = 1, 2, \ldots, K$), $\{P_1, P_2, \ldots, P_K\}$ is a set of mixing proportions such that

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

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

$$\theta = (\theta_1, P_1, \theta_2, P_2, \ldots, \theta_K, P_K). \qquad (3)$$

Each component $p_i(\cdot|\theta_i)$ ($i = 1, 2, \ldots, K$) of the above FMM is modeled by resorting to a finite dictionary $\mathcal{D} = \{f_1, f_2, \ldots, f_M\}$ of M SAR-specific distinct parametric pdfs $f_j(\cdot|\xi_j)$ ($j = 1, 2, \ldots, M$), parameterized by a vector ξ_j with parameter spaces Ξ_j (i.e., $\xi_j \in \Xi_j; j = 1, 2, \ldots, M$). The proposed method aims at integrating and automating the processes of: 1) optimizing the number K of mixture components; 2) selecting, for each mixture component, the optimal model inside the dictionary; 3) estimating, for each mixture component, the parameters of the optimal model.

As far as the optimization of the number of components is concerned, given a maximum number $K_{\max}$ of components, the proposed method computes (as described below) a K-component pdf estimate for each $K = 1, 2, \ldots, K_{\max}$. Several different validation functionals have been proposed in the literature as criteria to select the best value K^* of the parameter K, for instance, according to a Bayesian model-based framework [3], [4], to discriminant analysis [1], [25], [33], or to the minimum message length approach [19]. In the context of the proposed method, we use the correlation coefficient ρ_K [42] between the K-component pdf estimate and the empirical data distribution

(i.e., the image histogram) as a simple quantitative measure of the estimation quality: hence, the proposed technique chooses the number K^* of components yielding the largest value ρ^* of ρ_K. We note that the log-likelihood cannot be adopted as a criterion to choose K^*, due to the monotonic behavior of the log-likelihood with respect to K [19].

The key idea of the proposed dictionary-based method is to compute iteratively, for each $K = 1, 2, \ldots, K_{\max}$, a K-component pdf estimate by fitting, at every iteration, each parametric model in the dictionary to each mixture component. This approach generates a set of M candidate estimates per component; then, the optimal estimate (i.e., the optimal model in the dictionary) is selected according to an ML criterion, thus integrating at each iteration the processes of model selection and parameter estimation. In particular, the latter process is based on the integration of the SEM and the MoLC [53] estimation techniques. Here we adopt SEM both because it can avoid local maxima of the log-likelihood function and because it does not require the analytical maximization of the EM pseudolikelihood function. The adoption of several of the usual SAR amplitude or intensity parametric models (e.g., the Weibull or K distributions) for the mixture components yields no closed-form solution to this optimization problem, and complicates the application of most of the above-mentioned mixture estimation strategies.

We stress that the proposed "dictionary-based" SEM method (which, in the following, will be denoted by DSEM) turns out to be completely automatic. The selection of the number of mixture components, the choice of the optimal model for each component, and the estimation of the model parameters are jointly and automatically performed by the algorithm without need for the user's intervention. Only the maximum number $K_{\max}$ of components and the maximum number $t_{\max}$ of SEM iterations have to be defined prior to the application of the method, but the choice of $K_{\max}$ is not critical, as $K_{\max}$ is just an upper bound to the number of components [19], [33], and $t_{\max}$ only has to be large enough to let the iterative SEM process reach stationarity.

A flowchart of DSEM will be shown in Section III-C; in order to complete the overview of the algorithm, the following subsections address in greater detail the methodological issues involved by the method. In Section III-B, we briefly recall the key ideas of the SEM methodology, and in Section III-C, we describe the processing stages performed by DSEM. Comments on the specific models included in the dictionary $\mathcal{D}$ are made in Section III-D.

B. SEM for Mixture 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 $p_x(\cdot|\theta)$ depends on a parameter vector θ taking values from a given set $\Theta \subset \mathbb{R}^\ell$ (i.e., the parameter space). Here we refer to a generic "density function" since the adopted estimation procedure is suitable for both probability density function estimation in the case of continuous random vectors and for probability mass function estimation in the case of discrete vectors. Further details about the mathematical framework of this estimation problem can be found in [9], [35], and [37].

Such a general problem is said to be affected by data incompleteness when the data vector x cannot be directly observed,

---

[1]This approach is widely accepted in the context of estimation theory [18], [21], [54], and operatively corresponds to discarding, in the estimation process, the contextual information associated with the correlations between neighboring pixels in an image, thus exploiting only information about single pixels.

for instance, due to lacking or corrupted data [34]. The incompleteness issue is formalized by assuming the "complete" data vector x not to be available but to be observable only through an "incomplete" data vector $y = \Phi(x)$ obtained by a many-to-one mapping $\Phi : X \to Y \subset \mathbb{R}^m$ [17]. 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 computation of an ML estimate. SEM tries to avoid these difficulties by iteratively and randomly sampling a complete dataset and using it to compute an ML standard estimate.

Specifically, the FMM framework can be regarded as being affected by data incompleteness problems, since it is not known from which of the K available statistical populations (each corresponding to a mixture component) involved in (1) a given image sample has been drawn. This implicitly means that no training information about the possible land-cover types 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 by $\Sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_K\}$ the set of the K different populations, we assume to know the amplitude value of the hth pixel but not its population label $s_h \in \Sigma$ ($h = 1, 2, \ldots, N$). This suggests the following definitions of the complete and incomplete data vectors, respectively:

$$x = (r_1, s_1, r_2, s_2, \ldots, r_N, s_N) \quad y = (r_1, r_2, \ldots, r_N) \quad (4)$$

and of the parameter space

$$\Theta = \left\{ \theta : \sum_{j=1}^{K} P_j = 1, \ \theta_i \in \Theta_i, \ 0 \le P_i \le 1, \ i = 1, 2, \ldots, K \right\}. \quad (5)$$

As described in [9] and [37], assuming the couples $\{(r_h, s_h) : h = 1, 2, \ldots, N\}$ of random variables to be i.i.d., the tth iteration of the SEM algorithm for FMM parameter estimation involves the following operations ($t = 0, 1, \ldots$):

• *E-step:* For each pixel $h = 1, 2, \ldots, N$ and for each mixture component $i = 1, 2, \ldots, K$ compute the values of the posterior probabilities of the component labels given the observation r_h and the current parameter estimate $\theta^t \in \Theta$, i.e.,

$$P_{s|r}(\sigma_i | r_h, \theta^t) = \frac{P_i^t p_i(r_h | \theta_i^t)}{\sum_{j=1}^{K} P_j^t p_j(r_h | \theta_j^t)} = \tau_{ih}^t. \quad (6)$$

• *S-step:* Randomly sample a complete data realization x^t by sampling a label s_h^t for each hth pixel ($h = 1, 2, \ldots, N$) according to the current estimated posterior probabilities τ_{ih}^t ($i = 1, 2, \ldots, K$) of the pixel, thus implicitly partitioning the image $\mathcal{I}$ into 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.,[2]

$$P_i^{t+1} = \frac{|Q_i^t|}{N}$$
$$\theta_i^{t+1} = \arg\max_{\phi \in \Theta_i} \sum_{h \in Q_i^t} \ln p_i(r_h | \phi), \qquad i = 1, 2, \ldots, K \quad (7)$$

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

where $Q_i^t = \{h : s_h^t = \sigma_i\}$ is the index set of image samples assigned to the component σ_i ($i = 1, 2, \ldots, K$) at the tth iteration and $|Q_i^t|$ gives the number of pixels assigned to σ_i at the tth iteration.

C. Proposed "Dictionary" Approach to SAR Amplitude PDF Estimation

As mentioned in Section III-A, for each $K = 1, 2, \ldots, K_{\max}$, the proposed DSEM method generates a K-component pdf estimate and computes the correlation coefficient ρ_K between this estimate and the image histogram as a validation criterion. In order to compute the K-component pdf estimate, we integrate the dictionary-based approach into the described SEM estimation framework. In particular, at each SEM iteration, we exploit 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. Then, the optimal candidate is selected by integrating at each SEM iteration a further "model selection (MS) step."

In the M-step of the SEM algorithm, the computation (at the tth iteration) of the optimal parameter vector θ^{t+1} is performed by using an ML procedure. This approach turns out not to be feasible for several SAR-specific pdfs, such as the K distribution [41]. Hence, we avoid using ML estimates and, in the M-step, we adopt the MoLC approach, which has been proven to be a feasible and effective estimation tool for all usual SAR parametric models [40], [53].

MoLC has recently been proposed as a parametric pdf estimation technique suitable 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 modeling (e.g., the Nakagami–Gamma and the K distributions) [53]. MoLC is based on the generalization of the usual moment-based statistics by using the Mellin transform [51] for the computation of characteristic functions and moment generating functions, instead of the usual Fourier and Laplace transforms, and allows stating a set of (typically nonlinear) equations relating the unknown parameters of a given parametric model with one or more logarithmic moments or logarithmic cumulants (shortly "log-moments" and "log-cumulants"). The solution of such equations allows computing the desired parameter estimates [53]. Further details about the mathematical formulations of MoLC can be found in [37], [39], and [40].

Concerning the MS-step, we have adopted the log-likelihood function as a criterion to select the best model in the dictionary for each component.

Finally, in order to reduce the computation time of the proposed method, a histogram-based approach has been used. In particular, each iteration of the proposed DSEM algorithm is expressed explicitly in terms of the (nonnormalized) histogram $\{h(z) : z = 0, 1, \ldots, Z - 1\}$ of the image $\mathcal{I}$. Therefore, a component label drawn from $\{\sigma_1, \sigma_2, \ldots, \sigma_K\}$ is assigned to each gray level (and not to each image pixel), thus implicitly associating the same population label with all the pixels presenting the same gray level. For example, in the case of a 1024×1024 pixel image with 8 bpp (bits per pixel), the general SEM approach would involve both calculating the posterior probabilities of the mixture components and sampling a population label for each

of the $2^{20} \simeq 10^6$ image pixels, whereas the adopted approach deals directly with only $2^8 = 256$ distinct gray levels.

In particular, denoting by $p_i^t(\cdot)$ and $p^t(\cdot)$ the tth step σ_i-conditional $(i = 1, 2, \ldots, K)$ and unconditional amplitude pdf estimates, respectively, the tth DSEM iteration $(t = 0, 1, \ldots)$ can be summarized as follows:

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

$$\tau_i^t(z) = \frac{P_i^t p_i^t(z)}{p^t(z)}, \quad \text{where} \quad p^t(z) = \sum_{j=1}^{K} P_j^t p_j^t(z) \quad (8)$$

and in $p_i^t(\cdot)$ and $p^t(\cdot)$ the dependence on the parameter vectors of the selected models has been dropped for notational ease.

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

- *MoLC-step:* for each mixture component σ_i, compute the following histogram-based estimates of the mixture proportion and of the first three σ_i-conditional log-cumulants [53]

$$P_i^{t+1} = \frac{\sum_{z \in Q_i^t} h(z)}{\sum_{z=0}^{Z-1} h(z)}$$

$$\kappa_{1i}^t = \frac{\sum_{z \in Q_i^t} h(z) \ln z}{\sum_{z \in Q_i^t} h(z)}$$

$$\kappa_{2i}^t = \frac{\sum_{z \in Q_i^t} h(z) (\ln z - \kappa_{1i}^t)^2}{\sum_{z \in Q_i^t} h(z)}$$

$$\kappa_{3i}^t = \frac{\sum_{z \in Q_i^t} h(z) (\ln z - \kappa_{1i}^t)^3}{\sum_{z \in Q_i^t} h(z)} \quad (9)$$

where $Q_i^t = \{z : s^t(z) = \sigma_i\}$ is the set of gray-levels assigned to the component $\sigma_i (i = 1, 2, \ldots, K)$; 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 $f_j(\cdot|\xi_{ij}^t)$ according to the data assigned to σ_i

$$L_{ij}^t = \sum_{z \in Q_i^t} h(z) \ln f_j \left(z|\xi_{ij}^t\right) \quad (10)$$

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

A feasible initialization procedure for the described iterative process lies in assigning randomly and with the same probability $(1/K)$ the image samples to the mixture components in the initialization stage [9]. This is equivalent to setting initially a uniform posterior distribution for all the image pixels, and to using
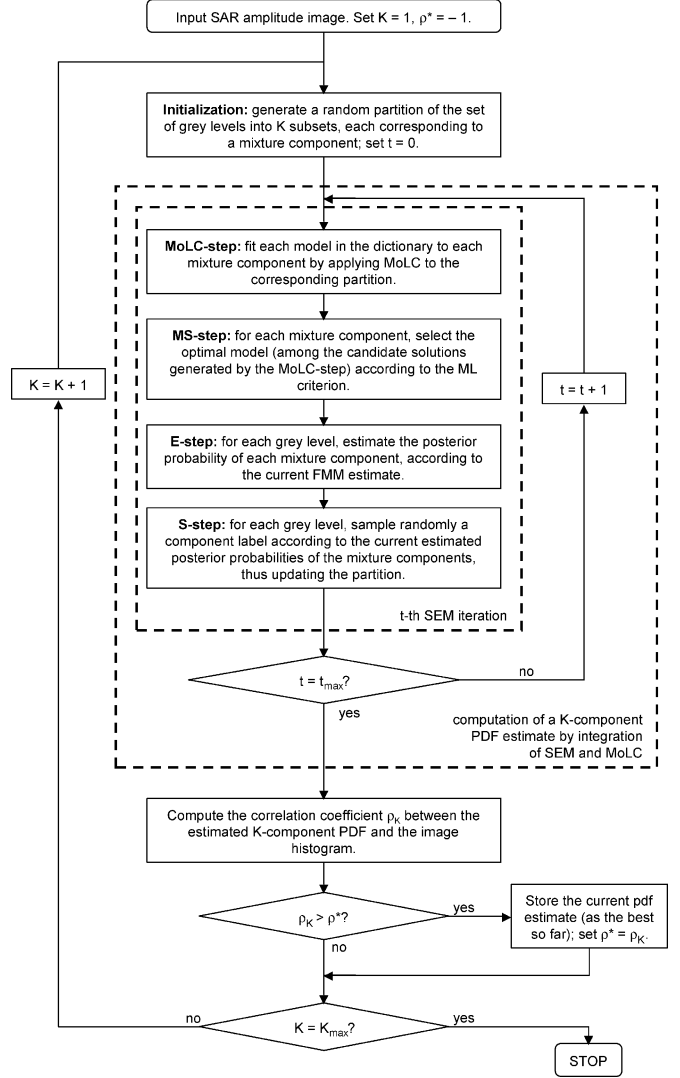

Fig. 1. Flowchart of the proposed DSEM algorithm.

this distribution in the random sampling process performed by the S-step.

The flowchart of the proposed technique in displayed in Fig. 1. Note that in the flowchart the order of the processing steps performed at each DSEM iteration is not the same as in the above methodological description (i.e., the first processing step in the flowchart is the MoLC-step, whereas it is the E-step in the description), since the flowchart refers to the order of the operations in the implementation of the method; anyhow, the two orders are equivalent.

As in the general SEM framework, the resulting sequence of pdf estimates is expected not to converge pointwise nor almost surely but to reach a stationary behavior, concentrating around the global maxima of the log-likelihood function [9]. This requires the definition of a specific procedure to extract a single optimal pdf estimate from the sequence itself [5], [9]. Here we adopt the approach proposed in [5] which computes, at any iteration $t(t = 0, 1, \ldots, t_{\max})$, the log-likelihood of the current pdf estimate $p^t(\cdot)$ over the whole image dataset $\mathcal{I}$, i.e.,

$$L_{\mathcal{I}}^t = \sum_{k=1}^{N} \ln p^t(r_k) = \sum_{z=0}^{Z-1} h(z) \ln p^t(z) \quad (11)$$

and chooses the estimate $p^t(\cdot)$ exhibiting the highest log-likelihood $L_{\mathcal{I}}^t (t = 0, 1, \ldots, t_{\max})$. This procedure selects, from among the $(t_{\max}+1)$ pdf estimates, the best "dictionary-based" ML estimate of the SAR amplitude pdf of the input image. As stated in Section III-A, the number $t_{\max}$ of iterations merely has to be large enough to let DSEM reach stationarity; this can be established by checking if the sequence $\{L_{\mathcal{I}}^t : t = 0, 1, \ldots\}$ of the log-likelihood values exhibits a steady stationary behavior.

D. Dictionary Content

Dealing with SAR amplitude data, we use a dictionary $\mathcal{D}$ consisting of the following six parametric pdfs:

- the empirical log-normal distribution [41]:

$$f_1(r|m,\sigma) = \frac{1}{\sigma r \sqrt{2\pi}} \exp\left[-\frac{(\ln r - m)^2}{2\sigma^2}\right], \qquad r > 0; \quad (12)$$

- the Nakagami distribution, proposed as an amplitude model for multilook SAR data [41], [53]:

$$f_2(r|L,\mu) = \frac{2}{\Gamma(L)} \left(\frac{L}{\mu}\right)^L r^{2L-1} \exp\left(-\frac{Lr^2}{\mu}\right), \qquad r \geq 0 \quad (13)$$

where $\Gamma(\cdot)$ is the Gamma function [48];
- the generalized Gaussian Rayleigh (GGR) distribution, based on a generalized Gaussian model for backscattered SAR signals [36], [38]:

$$f_3(r|\lambda,\gamma) = \frac{\gamma^2 r}{\lambda^2 \Gamma^2(\lambda)} \int_0^{\frac{\pi}{2}} \exp\left[-(\gamma r)^{\frac{1}{\lambda}}\right.$$
$$\left. \times \left(|\cos\theta|^{\frac{1}{\lambda}} + |\sin\theta|^{\frac{1}{\lambda}}\right)\right] d\theta, \qquad r \geq 0; \quad (14)$$

- the symmetric-α-stable (SαS) generalized Rayleigh distribution (here after simply denoted by SαSGR), based on an SαS model [2] for SAR backscattered signals [29]:

$$f_4(r|\alpha,\gamma) = r \int_0^{+\infty} \rho \exp(-\gamma\rho^\alpha) J_0(r\rho) d\rho, \qquad r \geq 0 \quad (15)$$

where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind [51];
- the empirical Weibull distribution [41]:

$$f_5(r|\eta,\mu) = \frac{\eta}{\mu^\eta} r^{\eta-1} \exp\left[-\left(\frac{r}{\mu}\right)^\eta\right], \qquad r \geq 0; \quad (16)$$

- the amplitude distribution corresponding to a K-distributed intensity [41] (hereafter denoted by "K-root"):

$$f_6(r|L,M,\mu) = \frac{4}{\Gamma(L)\Gamma(M)} \left(\frac{LM}{\mu}\right)^{\frac{(L+M)}{2}}$$
$$\times r^{L+M-1} K_{M-L}\left[2r\left(\frac{LM}{\mu}\right)^{\frac{1}{2}}\right], \qquad r \geq 0 \quad (17)$$

where $K_\nu(\cdot)$ ($\nu > 0$) is the νth-order modified Bessel function of the second kind [51].

TABLE I
MoLC EQUATIONS FOR ALL THE PARAMETRIC FAMILIES INCLUDED IN THE ADOPTED DICTIONARY. $\Psi(\cdot)$ IS THE DIGAMMA FUNCTION [8]. $\Psi(\nu, \cdot)$ IS THE νth-ORDER POLYGAMMA FUNCTION [8], AND $G_\nu(\cdot)$ IS THE INTEGRAL FUNCTION INTRODUCED IN [38] FOR GGR PARAMETRIC ESTIMATION

Parametric family		MoLC equation
f_1	Log-Normal	$\kappa_1 = \mu$
		$\kappa_2 = \sigma^2$
f_2	Nakagami	$2\kappa_1 = \ln\mu + \Psi(L) - \ln L$
		$4\kappa_2 = \Psi(1, L)$
f_3	GGR	$\kappa_1 = \lambda\Psi(2\lambda) - \ln\gamma - \lambda G_1(\lambda) G_0(\lambda)^{-1}$
		$\kappa_2 = \lambda^2\Psi(1, 2\lambda) + \lambda^2 G_2(\lambda) G_0(\lambda)^{-1} - \lambda^2 G_1(\lambda)^2 G_0(\lambda)^{-2}$
f_4	SαSGR	$\alpha\kappa_1 = \Psi(1)(\alpha - 1) + \alpha\ln 2 + \ln\gamma$
		$\kappa_2 = \Psi(1, 1)\alpha^{-2}$
f_5	Weibull	$\kappa_1 = \ln\mu + \Psi(1)\eta^{-1}$
		$\kappa_2 = \Psi(1, 1)\eta^{-2}$
f_6	K-root	$2\kappa_1 = \ln\mu + \Psi(L) - \ln L + \Psi(M) - \ln M$
		$4\kappa_2 = \Psi(1, L) + \Psi(1, M)$
		$8\kappa_3 = \Psi(2, L) + \Psi(2, M)$

Hence, with this specific choice, $M = 6$ distinct parametric families are involved in the estimation process. We do not include the Rayleigh distribution in the dictionary, as this pdf is a particular case of almost all the aforesaid pdfs [29], [38], [41]. Table I shows the MoLC nonlinear equations for the parametric families adopted in the proposed dictionary-based method [38]–[40], [53].

At each DSEM iteration and for each component σ_i, the application of MoLC to the models $f_1, f_2, \ldots, f_5$ requires the computations of sample estimates of the first two σ_i-conditional log-cumulants, whereas f_6 (i.e., K-root) also needs the sample estimate of the third σ_i-conditional log-cumulant [$i = 1, 2, \ldots, K$; see (9)]. The solution of the resulting equations turns out to be feasible and fast for all the considered distributions. Specifically, the log-normal distribution does not require a real solution process, since the parameters of this distribution are exactly the first two log-cumulants. SαSGR and Weibull allow analytical solutions of the corresponding systems of two equations. Nakagami, GGR, 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 three parametric families [38], [53]. Therefore, fitting all the six considered parametric models to each component does not result in a sharp increase in the computation time, as compared with the usual single-model approach. In addition, for several of the considered pdfs, good estimation properties have been proved theoretically for the MoLC approach. In particular, the MoLC estimates exhibit a lower variance than the ones given by the "method-of-moments" for the Nakagami distribution [40], and are consistent for the GGR one [38]. However, as pointed out in [38], the MoLC equations for GGR and K-root can yield no solutions for specific values of the sample log-cumulants. 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.

IV. EXPERIMENTAL RESULTS

A. Datasets for Experiments

The proposed DSEM algorithm for pdf estimation was tested on 11 real SAR images, and compared with several usual SAR-specific parametric pdf estimation strategies. The first six images used for the experiments were single bands acquired in August 1989 over the agricultural region of Feltwell (U.K.) by a fully polarimetric PLC-band NASA/JPL airborne sensor (for further details on this dataset, we refer the reader to [50]). Specifically, all three polarizations [horizontal, horizontal–vertical, vertical (HH, HV, VV)] acquired at band C, the HV and VV polarizations acquired at band L, and the HH polarization acquired at band P were utilized. The remaining channels (i.e., L-HH, P-HV, and P-VV) were discarded, since (as reported in [37]) their histograms exhibited strong irregularities. Hereafter, the adopted Feltwell bands will be denoted synthetically by "Feltwell-CHH," "Feltwell-CHV," . . ., "Feltwell-PHH."

The other five employed images were the following:

- a single-look ERS-1 image acquired in April 1993 over the urban and agricultural regions around Bourges, France;
- an ERS-1 image of the agricultural region of Flevoland, Netherlands;
- a three-look XSAR scene (hereafter denoted by "Suisse-Lake") of a portion of the Swiss territory: it includes a mountain area, a lake, and an urban area (for further details on this image, we refer the reader to [14]);
- a three-look XSAR image (hereafter denoted by "Suisse-Mountain") of a mountain area in the Swiss territory;
- a three-look E-SAR image of the area of Oberpfaffenhofen near Munich, Germany.

In particular, we stress that "Suisse-Mountain" exhibits a bimodal histogram, whereas all the other images show a unimodal statistics. In Figs. 2 and 3, we present, as examples, the "Feltwell-LHV" and the "Suisse-Mountain" images after histogram stretching and/or equalization.

B. PDF Estimation Results

The proposed DSEM method has been applied to the 11 considered images, and the resulting pdf estimates have been assessed both quantitatively (by computing their correlation coefficients and their Kolmogorov–Smirnov (KS) distances [42], [45] with the image histograms) and qualitatively (by visually comparing the plots of the estimates with those of the histograms).

Assuming a maximum number of mixture components equal to 10, the correlation coefficients between the resulting estimated pdfs and the image histograms are very high (always greater than 99%) for all 11 images considered (see Table II), thus demonstrating the effectiveness of the proposed method in terms of estimation accuracy. Correspondingly, very small values of the KS distances between the histograms and the related pdf estimates were obtained (note that the KS distance takes on values in the range [0, 1]). The visual comparison
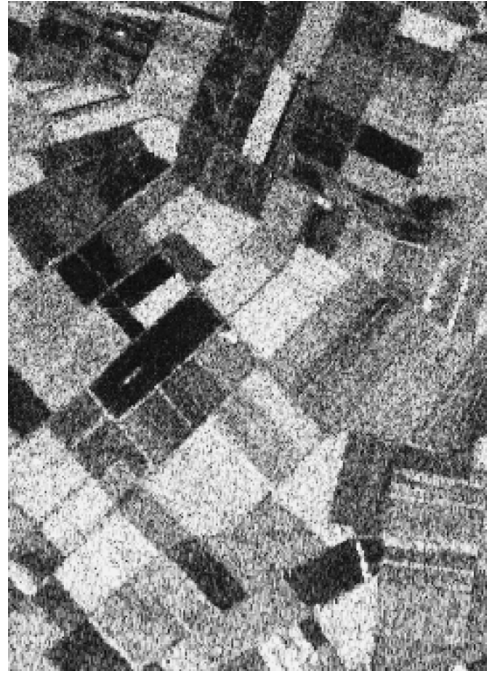

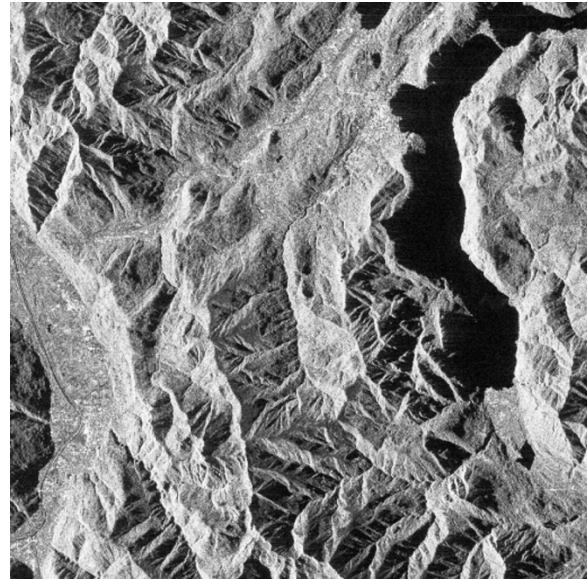Fig. 2. "Feltwell-LHV" image employed for the experiments.

Fig. 3. "Suisse-Mountain" image used for the experiments.

between the pdf estimates and the corresponding image histograms confirms this conclusion, as shown, for example, in Figs. 4, 5, and 7. In all the experiments, 100 iterations were sufficient for DSEM to reach stationarity: a further increase in t_{max} took longer computation times (which grow linearly with t_{max}), and provided no significant improvements in the correlation coefficients, also thanks to the fact that such coefficients were already very close to 100% (see Table II).

In order to further assess the capabilities of the method, a comparison has also been made with the six (either theoretical or empirical) models included in the dictionary, and the corresponding parameters have been estimated by using MoLC. The resulting correlation coefficients and KS distances are given in Tables III and IV, respectively. A comparison between Table II and Table III shows that the proposed DSEM

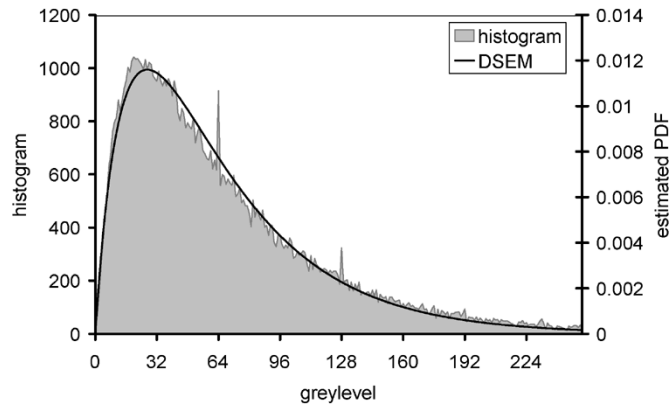| Image	ρ^*	KS distance	K^*	Selected models	Estimated mixing proportions
Feltwell-CHH	99,42%	0.0187	2	(f_3, f_3)	(0.49, 0.51)
Feltwell-CHV	99,41%	0.0103	5	$(f_3, f_1, f_1, f_1, f_2)$	(0.23, 0.11, 0.23, 0.16, 0.27)
Feltwell-CVV	99,62%	0.0178	3	(f_1, f_1, f_3)	(0.19, 0.39, 0.42)
Feltwell-LHV	99,41%	0.0253	3	(f_1, f_4, f_1)	(0.44, 0.16, 0.40)
Feltwell-LVV	99,70%	0.0210	6	$(f_1, f_3, f_4, f_4, f_3, f_1)$	(0.24, 0.10, 0.01, 0.16, 0.26, 0.23)
Feltwell-PHH	99,44%	0.0184	2	(f_3, f_1)	(0.39, 0.61)
Bourges	99,81%	0.0094	3	(f_6, f_2, f_3)	(0.35, 0.13, 0.52)
Flevoland	99,88%	0.0273	6	$(f_4, f_1, f_5, f_5, f_1, f_6)$	(0.17, 0.26, 0.19, 0.10, 0.16, 0.11)
Suisse-Lake	99,70%	0.0118	4	(f_1, f_1, f_4, f_1)	(0.34, 0.13, 0.15, 0.38)
Suisse-Mountain	99,88%	0.0096	4	(f_1, f_1, f_3, f_1)	(0.28, 0.23, 0.12, 0.36)
Oberpfaffenhofen	99,86%	0.0154	1	f_6	1

Fig. 4. Plots of the image histogram and of the DSEM pdf estimate for the "Feltwell-CHH" dataset.

Fig. 5. Plots of the image histogram and of the DSEM pdf estimates with 4 ($= K^*$) and 6 components, respectively, for the "Suisse-Mountain" dataset.

Image	Theoretical models				Empirical models	
	GGR	Nakagami	SαSGR	K-root	Log-Normal	Weibull
Feltwell-CHH	**99,42%**	95,80%	89,98%	not defined	98,52%	97,63%
Feltwell-CHV	**98,95%**	97,13%	88,87%	not defined	96,11%	98,72%
Feltwell-CVV	**99,11%**	94,56%	87,54%	not defined	98,86%	96,84%
Feltwell-LHV	**98,74%**	94,14%	75,55%	not defined	97,98%	97,25%
Feltwell-LVV	**99,28%**	93,43%	84,96%	not defined	99,26%	96,54%
Feltwell-PHH	98,61%	92,54%	81,57%	not defined	**99,27%**	95,71%
Bourges	98,97%	98,97%	95,57%	**99,77%**	95,65%	99,37%
Flevoland	not defined	98,70%	88,32%	**99,77%**	99,00%	97,35%
Suisse-Lake	**99,63%**	97,90%	94,30%	not defined	97,43%	98,78%
Suisse-Mountain	**93,70%**	92,76%	87,07%	not defined	91,25%	93,34%
Oberpfaffenhofen	not defined	**99,88%**	91,00%	99,86%	98,41%	99,60%

Image	Theoretical models				Empirical models	
	GGR	Nakagami	SαSGR	K-root	Log-Normal	Weibull
Feltwell-CHH	**0.0343**	0.0624	0.1164	not defined	0.0644	0.0485
Feltwell-CHV	0.0412	0.0547	0.1077	not defined	0.0854	**0.0377**
Feltwell-CVV	**0.0349**	0.0693	0.1306	not defined	0.0629	0.0525
Feltwell-LHV	**0.0562**	0.0961	0.1700	not defined	0.0790	0.0620
Feltwell-LVV	**0.0464**	0.0870	0.1428	not defined	0.0614	0.0649
Feltwell-PHH	**0.0359**	0.0903	0.1850	not defined	0.0501	0.0642
Bourges	0.0274	0.0399	0.0763	0.0301	0.0842	0.0354
Flevoland	not defined	0.0272	0.1511	**0.0127**	0.0349	0.0407
Suisse-Lake	**0.0241**	0.0513	0.1070	not defined	0.0559	0.0413
Suisse-Mountain	**0.0430**	0.0514	0.1152	not defined	0.0777	0.0482
Oberpfaffenhofen	not defined	**0.0150**	0.1312	0.0154	0.0484	0.0252

algorithm yields the pdf estimate with the highest correlation coefficients with the image histograms of 10 out of the 11 images, with the exception of "Oberpfaffenhofen." Anyway, for this image, the best result is provided by the Nakagami distribution with a correlation coefficient of 99.88%, which is very similar to the 99.86% coefficient provided by K-root and DSEM. Specifically, "Oberpfaffenhofen" is the only dataset for which the automatic method for the selection of the optimal number of components chooses $K^* = 1$ (see Table II). Among the six adopted parametric models in the dictionary, in this case K-root is selected by DSEM, since it yields the highest value for the log-likelihood ($L_\mathcal{I} = -3.0105 \cdot 10^6$), whereas Nakagami gives a slightly lower value ($L_\mathcal{I} = -3.0296 \cdot 10^6$), although providing a slightly better correlation coefficient. In any case, the difference between the performance of the Nakagami distribution and the DSEM/K-root one is almost negligible. Similarly, DSEM provides smaller values of the KS distances between the histograms and the estimated pdfs for all the images except "Oberpfaffenhofen" and "Flevoland." However, also in the cases of these two images, the KS distance values obtained by DSEM are close to the minimum ones achieved by Nakagami and K-root, respectively.
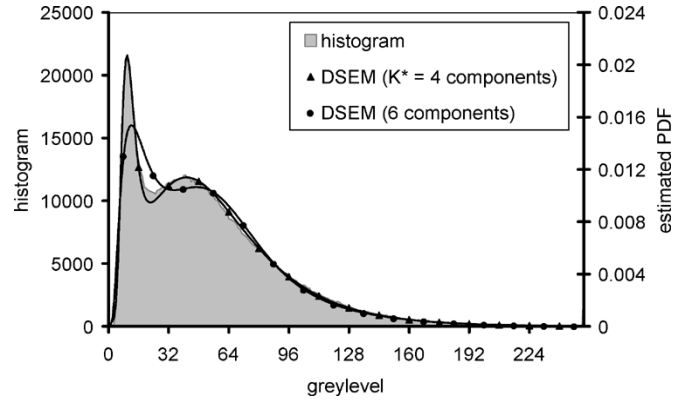
In particular, the good results obtained on "Suisse-Mountain," which exhibits a bimodal histogram, point out the usefulness of the adopted FMM approach. As shown in Fig. 5, DSEM effectively describes the bimodal statistics; by contrast, all the
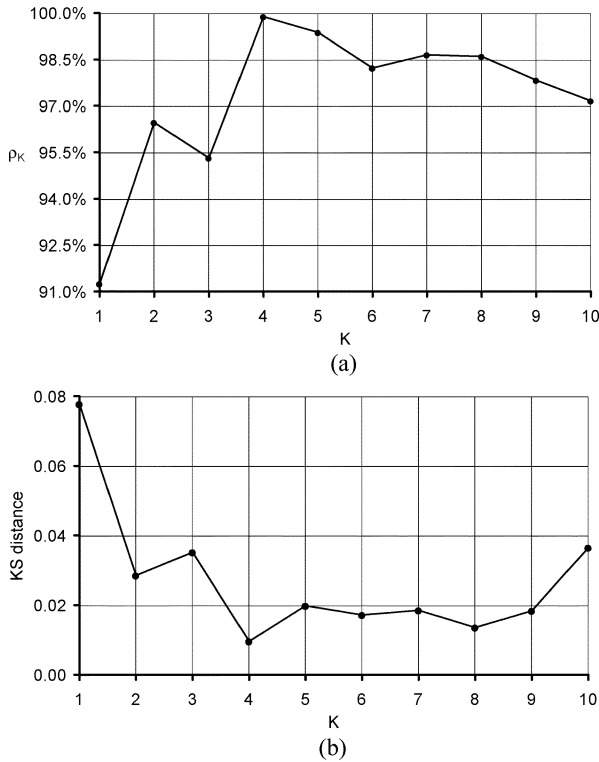
Fig. 6.   Plot of (a) the correlation coefficient ρ_K and of (b) the KS distance as functions of the number K of components for the "Suisse-Mountain" dataset.

Fig. 7.   Plots of the image histogram and of the DSEM and GGR pdf estimates for the "Feltwell-CHV" dataset.

parametric models considered for comparison are intrinsically monomodal and provide poor estimates (see Tables III and IV). DSEM selects the optimal number of components $K^* = 4$, which yields a high correlation coefficient (see Table II) and a visually accurate identification of both modes of the pdf (see Fig. 5); a different number of components would result in a worse estimate. Fig. 5 shows, for instance, also the six-component pdf estimate, which detects the bimodal structure of the histogram but strongly underestimates the height of the left mode and presents a bias in the identification of the position of the right one. In particular, in Fig. 6, we show the behaviors of the correlation coefficient ρ_K and of the KS distance as functions of the number K of components $(K = 1, 2, \ldots, 10)$ for the "Suisse-Mountain" image. In this case, a single-component model yields a poor 91.25% correlation coefficient, whereas a bimodal mixture allows achieving a 96.46% coefficient, and a four-component density gives a 99.88% result. A local maximum of ρ_K is obtained by two components, which is consistent with the visual bimodal behavior of the "Suisse-Mountain" histogram, although a more flexible four-component mixture allows achieving an even more precise estimate. Note also that ρ_K exhibits a decreasing behavior for $K \geq 7$, which further confirms that the choice of the upper bound $K_{\max} = 10$ is not critical. The same conclusions can be drawn from the plot of the KS distance, which exhibits a well-defined global minimum for $K = 4$ but also a local minimum for $K = 2$.

On the other hand, a comparison between Table II and Tables III and IV shows that, for most of the remaining images presenting unimodal histograms at least one of the single-model parametric distributions listed in Tables III and IV achieves results similar to the ones provided by DSEM. In particular, at least one among the GGR, K-root, and log-normal distributions
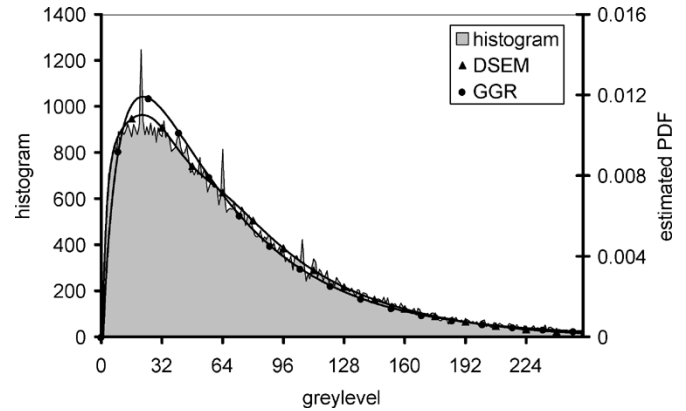
allows obtaining very accurate estimation results. In these cases, an FMM-based approach is not mandatory and a single-component pdf estimate turns out to be effective. A multicomponent model allows one to obtain, also in such situations, a better result than a single-component one, but the improvement is not very significant. For instance, in Fig. 7 we plot, for the "Feltwell-CHV" image, the pdf estimates provided by DSEM and by the best performing single parametric model (i.e., GGR).

In particular, the experiments proved that the generalized Gaussian Rayleigh and K-root models are the most effective single-component parametric families, as they achieved correlation coefficients higher than 98% in all the experiments. However, as discussed in [38], the estimation processes for both models can provide no solutions in the case of specific combinations of the values of the sample-log-cumulants. On the other hand, quite good results are also obtained by using a log-normal or a Weibull distribution, which can be fitted to any image histogram without restrictions and which give correlation coefficients often higher than 98%. Similar conclusions are also suggested by the corresponding KS distance values, as GGR provided, among the considered single-component models, the smallest KS distances for all the images except "Feltwell-CHV," for which the smallest KS distance was given by Weibull. We also note that the effectiveness of the GGR and log-normal models is also pointed out by the selection process performed by DSEM. As shown in Table II, log-normal (i.e., f_1) and GGR (i.e., f_3) turn out to be the most frequently chosen pdfs (among the parametric families in the dictionary) as models for the mixture components. In addition, an analysis of the mixing proportions estimated by DSEM (see Table II) suggests that, for almost all the images, all the selected components presented a proportion above 0.1, thus being likely to give a nonnegligible contribution to the mixture density representing the pdf estimate (the only exception turned out to be "Feltwell-LVV," for which the third component showed a small 0.01 proportion).

Finally, focusing on the meanings of the component labels, in Fig. 8 we show the label configuration generated by DSEM, when applied, for instance, to "Suisse-Mountain." Specifically, at each DSEM iteration, a component label is assigned to each gray level (i.e., all the image pixels exhibiting the same gray level are given the same label). Fig. 8(a) and (c) presents the optimal component labels in the "image space" and in the "gray-
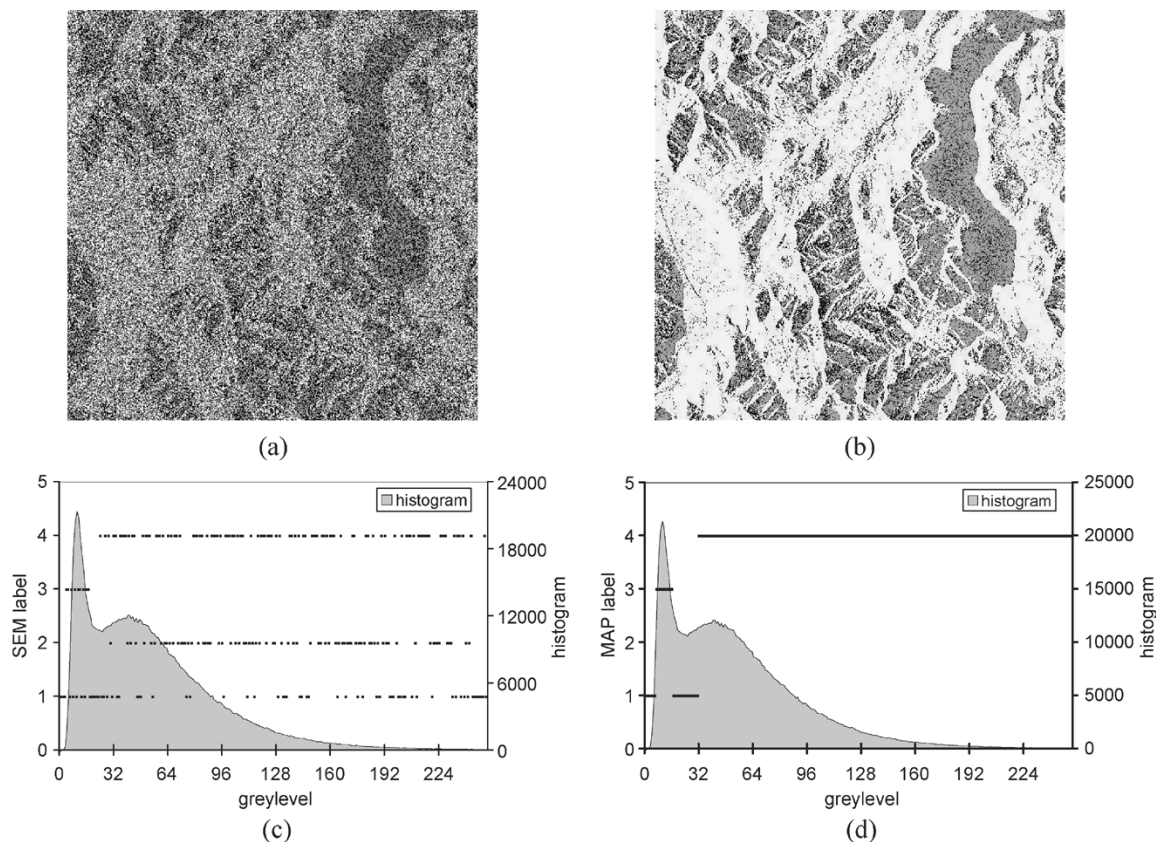
Fig. 8. Component labels for the "Suisse-Mountain" dataset ($K^* = 4$ components). (a) Map of the DSEM labels. (b) Label map generated by the MAP classification rule fed with the DSEM-based pdf estimates conditioned to each component. (c) DSEM labels in the gray-level space. (d) MAP labels in the gray-level space (color legend for the label maps: black $= \sigma_1$, dark gray $= \sigma_2$, light gray $= \sigma_3$, white $= \sigma_4$. In the gray-level plots, the component σ_i is denoted by the numeric label $i \in \{1, 2, 3, 4\}$).

level space," respectively. Due to the random sampling process generating such labels, the resulting map [Fig. 8(a)] turns out to be very noisy. Similarly, also the label configuration on the gray-level axis does not clearly identify four "decision regions" [each assigned to one of the components; see Fig. 8(c)]. In order to derive an unsupervised classification map (i.e., a clustering map) from the DSEM estimation result, the pdf and prior-probability estimates computed by DSEM for the four components can be fed, for example, to a "maximum *a posteriori*" (MAP) classifier [21], thus reassigning to each gray level a component label according to the MAP rule: a much less noisy map is obtained [Fig. 8(b)] that is consistent with the fact that the MAP labels effectively identify a set of well-defined decision regions on the gray-level axis [Fig. 8(d)] In particular, most gray levels are assigned to σ_3 and σ_4, which correspond to the two modes of the "Suisse-Mountain" distribution, while the decision region for σ_1 includes two intervals of the gray-level axis where both modes have small contents [left interval in Fig. 8(d)] or there is a transition between the two modes [right interval in Fig. 8(d)], and σ_2 is assigned no gray levels by the MAP rule. A visual analysis of the map suggests that the component σ_4 corresponds to light pixels in the mountain areas, σ_3 corresponds to the majority of the dark pixels, and σ_1 corresponds to very dark pixels and medium dark pixels.

A further experimental analysis has also been made by comparing the DSEM results with the ones yielded by the nonparametric Parzen window estimation method [43], applied using two different strategies to select the kernel-width parameter (see

[37]). For all the considered images, DSEM obtained estimation accuracies similar to the best ones given by these two versions of the Parzen approach (further details about this experimental comparison can be found in [37]).

V. CONCLUSION

In this paper, an innovative finite mixture model (FMM) estimation algorithm has been developed for SAR amplitude data pdf by integrating the SEM and MoLC methods with an automatic technique to select, for each mixture component, an optimal parametric model from a predefined dictionary of parametric pdfs. In particular, the developed estimation strategy has been explicitly focused on the context of SAR image analysis, and accordingly, a set of six theoretical or empirical models (see Table I) for SAR amplitude data have been adopted as a dictionary.

The numerical results of the application of the method to several real SAR images acquired by various SAR sensors prove that the proposed DSEM algorithm provides 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 coefficients between them. It is worth noting that the method was effective on all the considered images, despite the differences in their statistics (i.e., histogram unimodality or multimodality), dynamics range, and multilooking characteristics. KS distances smaller than 0.03 and correlation coefficients higher than 99%

were obtained in all the experiments, thus pointing out the flexibility of the method.

Specifically, the use of a mixture model appears to be an effective choice when dealing with multimodal statistics, for which the classical unimodal parametric models cannot provide satisfactory results. When applied to the "Suisse-Mountain" image, which exhibits a bimodal histogram, the developed DSEM algorithm correctly detected the positions and the heights of both modes. On the other hand, the results provided by DSEM in the case of unimodal histograms are usually comparable with the ones obtained by the best of the single-component parametric models included in the dictionary. In these cases, the multicomponent DSEM estimator generally yields only minor improvements.

The adopted dictionary consists of six (either theoretical or empirical) parametric pdfs, usually employed to characterize the statistics of different land-covers in amplitude SAR imagery in order to ensure that a mixture density with such components can be an appropriate model for a SAR scene. The high accuracies of the resulting estimates suggest that the proposed dictionary provides a sufficiently general model for SAR amplitude statistics; consequently, we have not increased the set of pdfs in the dictionary (e.g., including Fisher [53] or inverse Gaussian pdfs [20]). On the other hand, interesting future developments of the present research activity would be to extend the dictionary and to analyze the role of each pdf in order to establish if all the models are strictly necessary to achieve the obtained high-accuracy values or if some model(s) may be removed from the dictionary without affecting the quality of the resulting pdf estimates (this could allow a reduction in the overall computation time).

We note that the proposed DSEM algorithm turns out to be completely automatic, as it performs both the FMM estimation process and the optimization of the number of mixture components without any need for the user's intervention. In addition, thanks to the specific histogram-based SEM version adopted, the computation time of DSEM is small and almost independent of the image size (i.e., the image dimensions affect only the time required to compute the image histogram). These operational properties represent further elements in favor of the proposed method.

It is worth noting that such a histogram-based approach implicitly assumes the SAR amplitude data to be quantized on a discrete scale (e.g., corresponding to 8 or 16 bits per pixel), which may be a critical issue, due to the wide dynamic range typical of SAR data, especially in the presence of point scatterers. In such cases, this approach needs to be applied with care in order to ensure that the statistics of the structures in a given scene are correctly captured by the method. In particular, the adopted mixture model implicitly represents a continuous distribution; discrete components (e.g., that modeling the effect of clipping operations) can be taken into account by including impulsive terms in the mixture itself [44].

ACKNOWLEDGMENT

REFERENCES

[1] S. Bandyopadhyay and U. Maulik, "Nonparametric genetic clustering: Comparison of validity indices," *IEEE Trans. Syst., Man, Cybern., C*, vol. 31, no. 1, pp. 120–125, Feb. 2001.

[2] 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, Dec. 1999.

[3] C. Biernacki, G. Celeux, and G. Govaert, "An improvement of the NEC criterion for assessing the number of clusters in a mixture model," *Pattern Recognit. Lett.*, vol. 20, pp. 267–272, 1999.

[4] ——, "Assessing a mixture model for clustering with the integrated completed likelihood," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 7, pp. 719–725, Jul. 2000.

[5] ——, "Strategies for getting highest likelihood in mixture models," INRIA, Res. Rep. 4255, Sep. 2001. [Online]. Available: http://www.inria.fr/rrrt/rr-4255.html.

[6] ——, "Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models," *Computational Statistics and Data Analysis*, vol. 41, no. 3, pp. 561–575, 2003.

[7] C. M. Bishop, *Neural Networks for Pattern Recognition*, 2nd ed. Oxford, U.K.: Oxford Univ. Press, 1996.

[8] F. Bowman, *Introduction to Bessel Functions*. New York: Dover, 1968.

[9] G. Celeux, D. Chauveau, and J. Diebolt, "On Stochastic Versions of the EM Algorithm," INRIA, Res. Rep. 2514, Mar. 1995. [Online]. Available: http://www.inria.fr/rrrt/rr-2514.html.

[10] G. Celeux, S. Chretien, F. Forbes, and A. Mkhadri, "A Component-Wise EM Algorithm for Mixtures," INRIA, Res. Rep. 3746, Aug. 1999. [Online]. Available: http://www.inria.fr/rrrt/rr-3746.html.

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

[12] M. Cheney, "A mathematical tutorial on synthetic aperture radar," *SIAM Rev.*, vol. 43, no. 2, 2001.

[13] ——, "An introduction to synthetic aperture radar (SAR) and SAR interferometry," in *Approximation Theory X: Wavelets, Splines, and Applications*, C. K. Chui, L. L. Schumacher, and J. Stockler, Eds. Nashville, TN: Vanderbilt Univ. Press, 2002, pp. 167–177.

[14] M. Datcu, K. Seidel, and M. Walessa, "Spatial information retrieval from remote sensing images: Part I. information theoretical perspective," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, pp. 1431–1445, 1998.

[15] Y. Delignon, A. Marzouki, and W. Pieczynski, "Estimation of generalized mixtures and its application to image segmentation," *IEEE Trans. Image Process.*, vol. 6, no. 10, pp. 1364–1375, Oct. 1997.

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

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data and the EM algorithm," *J. R. Statist. Soc.*, no. 39, pp. 1–38, 1977.

[18] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[19] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.

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

[21] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. Orlando, FL: Academic, 1990.

[22] Q. Jackson and D. A. Landgrebe, "An adaptive classifier design for high-dimensional data analysis with a limited training dataset," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 12, pp. 2664–2679, Dec. 2001.

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

[24] ——, "Significance of K distributions in scattering experiments," *Phys. Rev. Lett.*, vol. 40, pp. 546–550, 1978.

[25] R. Kothary and D. Pitts, "On finding the number of clusters," *Pattern Recognit. Lett.*, vol. 20, pp. 405–416, 1999.

[26] B.-C. Kuo and D. A. Landgrebe, "A covariance estimator for small sample size classification problems and its application to feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 4, pp. 814–819, Apr. 2002.

[27] E. E. Kuruoglu, "Density parameter estimation of skewed alpha-stable distributions," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2192–2201, Oct. 2001.

[28] E. E. Kuruoglu and J. Zerubia, "Skewed α-stable distributions for modeling textures," *Pattern Recognit. Lett.*, vol. 24, pp. 339–348, 2003.

[29] ——, "Modeling SAR images with a generalization of the Rayleigh distribution," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 527–533, Apr. 2004.

[30] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. New York: Wiley, 2003.

[31] 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, Mar. 2005.

[32] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised statistical segmentation of satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 3, pp. 618–633, Mar. 1993.

[33] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1650–1654, Dec. 2002.

[34] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, 1996.

[35] G. Moser, "Development of unsupervised change detection methods for remote sensing images," "Laurea" degree thesis, Univ. Genoa, Genoa, Italy, Nov. 2001.

[36] 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, 2006, to be published.

[37] ——, "Dictionary-based stochastic expectation-maximization for SAR amplitude probability density function estimation," INRIA, Res. Rep. 5154, Mar. 2004. [Online]. Available: http://www.inria.fr/rrrt/rr-5154.html.

[38] ——, "SAR amplitude probability density function estimation based on a generalized Gaussian scattering model," INRIA, Res. Rep. 5153, Mar. 2004. [Online]. Available: http://www.inria.fr/rrrt/rr-5153.html.

[39] J. M. Nicolas, "Introduction aux statistiques de deuxième espéce: Application aux lois d'images RSO," ENST, Paris, Res. Rep. 2002D001, Feb. 2002. [Online]. Available: http://www.tsi.enst.fr/~nicolas/note.ps.

[40] J.-M. Nicolas, "Introduction aux statistiques de deuxième espéce: Applications des logs-moments et des logs-cumulants a l'analyze des lois d'images radar," *Trait. Signal*, vol. 19, no. 3, pp. 139–167, 2002.

[41] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Norwood: Artech House, 1998.

[42] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[43] E. Parzen, "On estimation of probability density function and mode," *Signal Process.*, vol. 33, pp. 267–281, 1962.

[44] M. Petrou, F. Giorgini, and P. Smits, "Modeling the histograms of various classes in SAR images," *Pattern Recognit. Lett.*, vol. 23, pp. 1103–1107, 2002.

[45] W. H. Press, S. A. Teukolsky, W. T. Wetterling, and B. P. Flannery, *Numerical Recipes in C*, Cambridge Univ. Press, Cambridge, U.K., 2002. [Online]. Available: http://www.library.cornell.edu/nr/bookcpdf.html.

[46] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood, and the EM algorithm," *SIAM Rev.*, vol. 26, no. 2, pp. 195–239, 1984.

[47] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis*. Berlin, Germany: Springer-Verlag, 1999.

[48] W. Rudin, *Principles of Mathematical Analysis*, 2nd ed. New York: McGraw-Hill, 1976.

[49] S. M. Schweizer and J. M. F. Moura, "Efficient detection in hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 584–597, Apr. 2001.

[50] S. B. Serpico, L. Bruzzone, and F. Roli, "An experimental comparison of neural and statistical nonparametric algorithms for supervised classification of remote sensing images," *Pattern Recognit. Lett.*, vol. 17, pp. 1331–1341, 1996.

[51] I. Sneddon, *The Use of Integral Transforms*. New York: McGraw-Hill, 1972.

[52] S. Tadjudin and D. A. Landgrebe, "Robust parameter estimation for mixture model," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 439–445, Jan. 2000.

[53] 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, Oct. 2004.

[54] H. L. Van Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968, vol. 1.

[55] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

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

[57] E. M. Winter and S. G. Beaven, "Effect of sensor defects on hyperspectral mixture analysis algorithms," in *Proc. SPIE Conf. Imaging Spectrometry V*, vol. 3753, 1999, pp. 152–157.

[58] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*. New York: Springer Verlag, 1985.

[59] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, pp. 95–103, 1983.

Gabriele Moser (S'03–M'05) received the laurea (M.S.) degree (summa cum laude) in telecommunications engineering and the Ph.D. degree in space sciences and engineering from the University of Genoa, Genoa, Italy, in 2001 and 2005, respectively.

Since 2001, he has cooperated with the Signal Processing and Telecommunications Research Group (SP&T) of the Department of Biophysical and Electronic Engineering (DIBE), University of Genoa, in the field of remote sensing image analysis. From January to March 2004, he was a Visiting Student at the Institut National de Recherche en Informatique et en Automatique, Sophia Antipolis, France, working with the "Ariana" research group on the problem of SAR data modeling. He is currently a Research Fellow at DIBE. His research activity is focused on image processing and image analysis methodologies for remote sensing data interpretation. In particular, his current research interests include SAR data analysis, multitemporal image classification, partially supervised classification, hyperspectral image analysis, and contextual classification. He has been a reviewer for several international journals.

Dr. Moser is a student member of the International Society for Optical Engineering (SPIE).

Josiane B. Zerubia (S'78–M'81–SM'99–F'03) received the M.Sc. degree from the Department of Electrical Engineering at Ecole Nationale Supérieure d'Ingénieurs Electriciens de Grenoble, Grenoble, France, in 1981, and the Dr.Eng. degree, the Ph.D. degree, and the "Habilitation," all from the University of Nice, Sophia-Antipolis, France, in 1986, 1988, and 1994, respectively.

She is a Permanent Research Scientist at l'Institut National de Recherche en Informatique et en Automatique, (INRIA), Sophia Antipolis, since 1989. She has been Director of Research since July 1995. She was Head of a remote sensing laboratory (PASTIS, INRIA Sophia-Antipolis) from mid-1995 to 1997. Since January 1998, she has been in charge of a new research group working on remote sensing (ARIANA, INRIA-CNRS-University of Nice). She has been an Adjunct Professor at Sup'Aero (ENSAE), Toulouse, France, since 1999. Before, she was with the Signal and Image Processing Institute, University of Southern California (USC), Los Angeles as a Postdoctoral Researcher. She also worked as a Researcher for the LASSY (University of Nice and CNRS) from 1984 to 1988 and in the Research Laboratory of Hewlett Packard in France and in Palo Alto, CA, from 1982 to 1984. Her current research interest is image processing using probabilistic models or variational methods. She also works on parameter estimation and optimization techniques. She has been a member of the editorial board of the French Society for Photogrammetry and Remote Sensing (SFPT) since 1998 and of the *International Journal of Computer Vision* since 2004.

Dr. Zerubia has been part of the IEEE BISP Technical Committee (SP Society) since 2005. She was Co-Chair of two workshops on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR'01, Sophia Antipolis, France, and EMMCVPR'03, Lisbon, Portugal), Co-Chair of a workshop on Image Processing and Related Mathematical Fields (IPRM'02, Moscow, Russia), Chair of a workshop on Photogrammetry and Remote Sensing for Urban Areas, Marne La Vallee, France, 2003. She was part of the IEEE IMDSP Technical Committee (SP Society) from 1997 to 2003, Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING from 1998 to 2002, Guest Coeditor of a special issue of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE in 2003, member-at-large of the Board of Governors of IEEE SP Society from 2002 to 2004. She has been area editor of IEEE TRANSACTIONS ON IMAGE PROCESSING since 2003.

Sebastiano B. Serpico (M'87–SM'00) received the laurea degree in electronic engineering and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1982 and 1989, respectively.

Since 1982, he has cooperated with the Department of Biophysical and Electronic Engineering (DIBE), University of Genoa, in the field of image processing and recognition. He was an Assistant Professor from 1990 to 1998 and an Associate Professor of telecommunications from 1998 to 2004 with the Faculty of Engineering, University of Genoa, where he taught signal theory, pattern recognition, telecommunication systems, and electrical communications He is currently a Full Professor of telecommunications. From 1995 to 1998, he was Head of the Signal Processing and Telecommunications Research Group (SP&T) at DIBE and is currently Head of the SP&T Laboratory. He is the Chairman of the Institute of Advanced Studies in Information and Communication Technologies. His current research interests include the application of pattern recognition (feature selection, classification, change detection, and data fusion) to remotely sensed images. He is the author or coauthor of more than 150 scientific publications, including journals and conference proceedings.

Dr. Serpico was the recipient of the *Recognition of TGARS Best Reviewers* from the IEEE Geoscience and Remote Sensing Society in 1998. He was a Guest Editor (with Professor D. A. Landgrebe) of a special issue of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING on the subject of the analysis of hyperspectral image data (July 2001) and a special issue on Advances in Techniques for Analysis of Remotely Sensed Data (March 2005). Since 2001, he has been an Associate Editor of the TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is a member of the International Association for Pattern Recognition Society (IAPR).