



HAL
open science

Unsupervised Learning of Generalized Gamma Mixture Model with Application in Statistical Modeling of High-Resolution SAR Images

Heng-Chao Li, Vladimir A. Krylov, Ping-Zhi Fan, Josiane Zerubia, William J. Emery

► **To cite this version:**

Heng-Chao Li, Vladimir A. Krylov, Ping-Zhi Fan, Josiane Zerubia, William J. Emery. Unsupervised Learning of Generalized Gamma Mixture Model with Application in Statistical Modeling of High-Resolution SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54 (4), pp.2153-2170. hal-01217654

HAL Id: hal-01217654

<https://inria.hal.science/hal-01217654v1>

Submitted on 20 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Unsupervised Learning of Generalized Gamma Mixture Model with Application in Statistical Modeling of High-Resolution SAR Images

Heng-Chao Li, *Senior Member, IEEE*, Vladimir A. Krylov, Ping-Zhi Fan, *Fellow, IEEE*, Josiane Zerubia, *Fellow, IEEE*, and William J. Emery, *Fellow, IEEE*

Abstract

The accurate statistical modeling of synthetic aperture radar (SAR) images is a crucial problem in the context of effective SAR image processing, interpretation and application. In this paper a semi-parametric approach is designed within the framework of finite mixture models based on the generalized Gamma distribution (GFD) in view of its flexibility and compact form. Specifically, we develop a generalized Gamma mixture model (GFMM) to implement an effective statistical analysis of high-resolution SAR images and prove the identifiability of such mixtures. A low-complexity unsupervised estimation method is derived by combining the proposed histogram-based expectation-conditional maximization (ECM) algorithm and the Figueiredo-Jain algorithm. This results in a numerical maximum likelihood (ML) estimator that can simultaneously determine the ML estimates of component parameters and the optimal number of mixture components. Finally, the state-of-the-art performance of this proposed method is verified by experiments with a wide range of high-resolution SAR images.

Index Terms

Synthetic aperture radar (SAR) images, finite mixture model, generalized Gamma distribution, expectation-conditional maximization (ECM) algorithm, minimum message length (MML), probability density function estimation, unsupervised learning.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) has become a very important Earth observation technique because of its all-weather, day-and-night, high spatial resolution and subsurface imaging capabilities [1]. As an active imaging system, SAR generates acquisitions via an imaging algorithm over the received echo signals from interactions of electromagnetic waves emitted by the antenna system of sensor with the illuminated ground area. With the rapid advancement of SAR technologies, such imagery is becoming more accessible, further greatly promoting the potential of SAR applications. In the context of SAR image processing and applications, a crucial problem is represented by the need to develop accurate models for the statistics of pixel intensities. To date, a lot of work related to the statistical nature of SAR images has been published in the literature.

From a methodological point of view, the strategies of statistical analysis for SAR images can be divided into three categories: non-parametric, semi-parametric and parametric approaches [2]. The non-parametric approach (e.g., Parzen window, support vector machine) is completely data driven but typically computationally expensive [3]. On the other hand, the parametric approach has the substantial advantages of simplicity and applicability. Given a

This version of the manuscript was accepted to the IEEE Transactions on Geoscience and Remote Sensing on October 9, 2015. This work was supported in part by the Program for New Century Excellent Talents in University under Grant NCET-11-0711, by the National Natural Science Foundation of China under Grant 61371165, by the Innovative Intelligence Base Project (111 Project No. 111-2-14), and by the Fundamental Research Funds for the Central Universities under Grant SWJTU12CX004.

H.-C. Li and P.-Z. Fan are with the Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 610031, China (e-mail: lihengchao_78@163.com). Now H.-C. Li is also with the Department of Aerospace Engineering Sciences, University of Colorado, Boulder, CO 80309 USA.

V. A. Krylov has participated in the work during his postdoc at the Ayin team, INRIA. He is currently with Dept. of Electrical, Electronic, Telecom. Engineering and Naval Architect. (DITEN), University of Genoa, 16145, Genoa, Italy (e-mail: vladimir.krylov@unige.it).

J. Zerubia is with Ayin research team, INRIA, Sophia Antipolis F-06902, France (e-mail: josiane.zerubia@inria.fr).

W.-J. Emery is with the Department of Aerospace Engineering Sciences, University of Colorado, Boulder, CO 80309 USA (e-mail: emery@colorado.edu).

mathematical model, it formulates the probability density function (PDF) estimation as a parameter estimation problem. The well-known examples in this category include the log-normal [1], Weibull [1], Fisher [4], and generalized Gamma [5] empirical models as well as several theoretical ones, such as Rayleigh [1], Nakagami-Gamma [1], [6], heavy-tailed Rayleigh [7], generalized Gaussian Rayleigh (GGR) [8], generalized Gamma Rayleigh (GTR) [2], \mathcal{K} [9] and \mathcal{G} [10]. Nowadays, the new generation of high-resolution spaceborne SAR satellites like TerraSAR-X, COSMO-SkyMed, and RADARSAT-2, together with modern airborne SAR systems, enables the systematic acquisition of data with spatial resolutions reaching metric/submetric values. Increasing SAR resolution implies a reduction of the number of scatterers per resolution cell and also an enhancement of the appreciability of backscattering responses from distinct ground cover materials. Therefore, the histograms of high-resolution SAR images, that commonly contain complex land-cover typologies, exhibit heavy-tailed or bi/multimodal characteristics. Under such conditions, it is impossible to apply a single parametric PDF model to accurately describe the statistics.

To address this issue, the semi-parametric approach is a good choice, which is designed as a compromise between nonparametric and parametric ones, and is related to the finite mixture model (FMM) [11] in the sense that the underlying PDF is defined as a weighted sum of parametric components. An example is the SAR amplitude model presented in [12], [13], where the components belong to a given dictionary of SAR-specific parametric models. However, this dictionary-based semi-parametric method suffers from several drawbacks. To begin with, it is not designed in general for the intensity data and can be adapted to intensities only after dictionary revision. In addition, some adopted parametric components are usually valid for the low or medium-resolution SAR images, while others take the PDFs in non-analytical form. Furthermore, the method of log-cumulants (MoLC) [4], [14], [40] for component parameter estimation, integrated in stochastic expectation-maximization (SEM) algorithm [15], cannot guarantee the likelihood (or log-likelihood) function to be monotonically increasing, thus resulting in the difficulty of effectively identifying the optimal/true number of mixture components.

In light of these limitations, we employ the generalized Gamma distribution (G Γ D) as the type-specific mixture components within the framework of FMM to formulate the generalized Gamma mixture model (G Γ MM). Our aim in this paper is to develop an efficient semi-parametric statistical analysis procedure for high-resolution SAR images. As an empirical parametric model for SAR images, the G Γ D has been demonstrated to be competitive, and performs commonly better than the majority of the previously developed parametric models in fitting SAR image data histograms for most cases [5]. Its further advantages include a compact analytical form and a rich family of distributions. Most importantly, it has the flexibility to model SAR images covering different kinds of scenes in both amplitude and intensity formats. Thanks to these merits of G Γ D, the G Γ MM offers high flexibility, robustness, and can be employed as a universally applicable tool for accurately representing arbitrarily complex PDFs of high-resolution SAR image data. Before proposing an estimator, we demonstrate the theoretical property of identifiability of G Γ MM. This ensures that any such mixture representation is unique and, therefore, the estimation problem is well-posed. To proceed with the estimation, we first reduce the complexity of parameter estimation by expressing the log-likelihood function as a function of image gray levels, rather than the commonly used image pixels. Then we derive a histogram-based expectation-conditional maximization (ECM) algorithm for iteratively finding the maximum likelihood (ML) solutions of each component parameters. This simplifies the complete-data ML estimation of G Γ MM by decomposing a complicated maximization step (M-step) of the expectation-maximization (EM) algorithm into several computationally simpler conditional M-steps (CM-steps). This histogram-based ECM version is incorporated in the Figueiredo-Jain (FJ) algorithm with component annihilation [16] to learn G Γ MM in an unsupervised way. Due to the FJ algorithm's properties, the proposed unsupervised learning method for the G Γ MM (hereafter referred to as HECM-FJ-G Γ MM) can automatically infer the optimal number of components, and avoid the issues of initialization and possible convergence to the boundary of the parameter space associated with the standard EM algorithm. Additionally, and very importantly, given a SAR image histogram, the computational complexity of HECM-FJ-G Γ MM is independent of image size. Finally, performance evaluations are also conducted to verify its validity on a wide range of real high-resolution space/airborne SAR images.

The remainder of this paper is organized as follows. In Section II, we give the definition of G Γ MM in detail, and demonstrate its identifiability. Section III derives the histogram-based ECM algorithm for iterative estimation of mixture component parameters on the basis of EM algorithm. Section IV presents the HECM-FJ-G Γ MM for unsupervised learning of G Γ MM, and describes its complete implementation. Experimental results are presented in Section V. Section VI ends this paper with some concluding remarks.

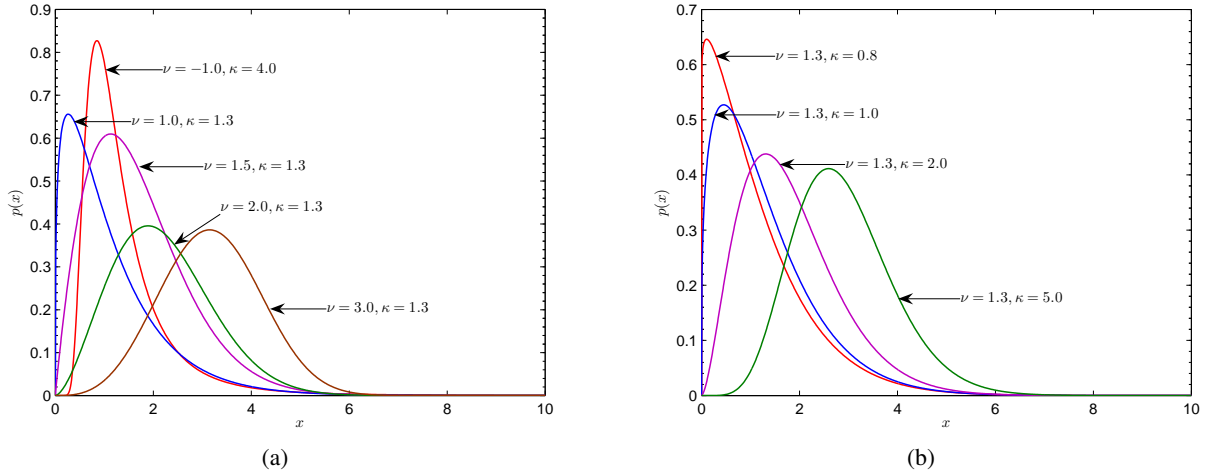


Fig. 1. PDFs of the G Γ D with unit variance for different pairs of ν and κ : (a) $\nu = -1, \kappa = 4.0$ and $\nu = \{1.0, 1.5, 2.0, 3.0\}, \kappa = 1.3$, (b) $\nu = 1.3, \kappa = \{0.8, 1.0, 2.0, 5.0\}$.

II. MODEL FORMULATION

A. Generalized Gamma Distribution

The generalized Gamma distribution was first introduced in 1962 by Stacy [17], and recently has been suggested for use as a flexible empirical statistical model of SAR images by the authors in [5]. The PDF of G Γ D is defined as

$$p(x) = \frac{|\nu|}{\sigma \Gamma(\kappa)} \left(\frac{x}{\sigma}\right)^{\kappa\nu-1} \exp\left\{-\left(\frac{x}{\sigma}\right)^\nu\right\}, x \in \mathbb{R}^+ \quad (1)$$

where ν is nonzero, κ and σ are positive real values, corresponding to the power, shape and scale parameters, respectively, and $\Gamma(\bullet)$ denotes the Gamma function. Its m^{th} -order moment is given by $E(x^m) = \sigma^m \frac{\Gamma(\kappa+m/\nu)}{\Gamma(\kappa)}$ if $m/\nu > -\kappa$, $E(x^m) = \infty$ otherwise. The G Γ D family contains a large variety of alternative distributions, including the Rayleigh ($\nu = 2, \kappa = 1$), exponential ($\nu = 1, \kappa = 1$), Nakagami ($\nu = 2$), Gamma ($\nu = 1$), Weibull ($\kappa = 1$) distributions commonly used for the PDFs of SAR images as special cases and log-normal ($\kappa \rightarrow \infty$) as an asymptotic case.

For illustrative purposes, Fig. 1 shows some examples of equation (1) with unit variance for different pairs of ν and κ , in which from the unit variance condition the third parameter (i.e., σ) is calculated by $\sigma = \frac{\Gamma(\kappa)}{\sqrt{\Gamma(\kappa+2/\nu)\Gamma(\kappa)-\Gamma(\kappa+1/\nu)^2}}$. It can be seen that the extra index power parameter ν is able to provide more flexibility to control the model shape, which combines with κ to make (1) mimic the PDFs with many behaviors of the mode and tails. When ν becomes smaller, the G Γ D exhibits some heavy-tailed characteristics and vice versa. Note that the G Γ D can often describe SAR images with unimodal histograms in both amplitude and intensity format better than the majority of the previously developed parametric models [5]. As such, it provides a good candidate for the mixture component to perform an efficient semi-parametric statistical analysis of high-resolution SAR images.

B. Generalized Gamma Mixture Model

Assume a set of data $X = \{x_i | i = 1, \dots, N\}$, where each x_i is a realization of variable x , here denoting the pixel value defined in an alphabet $\mathcal{D} = \{0, 1, \dots, L-1\}$ for a given high-resolution SAR image. Considering that the pixel value is a random variable, we propose a generalized Gamma mixture model (G Γ MM) to characterize it, which is defined as the weighted sum of M G Γ D components i.e.,

$$p(x|\Theta) = \sum_{m=1}^M \pi_m p(x|\theta_m) \quad (2)$$

where π_m corresponds to the mixture weight of each component, and satisfies

$$\pi_m > 0, \quad \sum_{m=1}^M \pi_m = 1. \quad (3)$$

$p(x|\theta_m)$ is a GFD describing the m^{th} component in the form of (1) with $\theta_m = \{\nu_m, \kappa_m, \sigma_m\}$. The symbol Θ refers to the whole set of parameters of a given M -mixture to be estimated, denoted by $\Theta = \{\pi_1, \pi_2, \dots, \pi_M, \theta_1, \theta_2, \dots, \theta_M\}$.

An unsupervised learning task for a GFMM is equivalent to the estimation of \widehat{M} and $\widehat{\Theta}$. In order for this estimation problem to be well-posed, in the following theorem we demonstrate the identifiability of GFMM. This property means that any GFMM allows a unique representation within the class of GFD mixtures. We stress that this is not a trivial statement, and analogous results have been obtained for some special cases of GFMM [18], [19].

Theorem 1: Let $\mathcal{F} = \{f : f_\theta(x) = p(x|\nu, \kappa, \sigma), \nu \neq 0, \kappa > 0, \sigma > 0\}$ be the family of GFDs. With π_m satisfying (3), the class $\mathcal{H}_{\mathcal{F}} = \left\{ H : H(x) = \sum_{m=1}^M \pi_m f_{\theta_m}(x), f_{\theta_{1 \leq m \leq M}}(x) \in \mathcal{F} \right\}$ of all finite mixtures of \mathcal{F} is identifiable. That is to say, for any two GFMMs $H_{1,2} \in \mathcal{H}_{\mathcal{F}}$, i.e.,

$$H_1 = \sum_{m=1}^{M_1} \pi_{1m} f_{\theta_{1m}}(x), \quad H_2 = \sum_{m=1}^{M_2} \pi_{2m} f_{\theta_{2m}}(x) \quad (4)$$

with $\theta_{im} = \theta_{in} \Leftrightarrow m = n$ for $i = 1, 2$, if $H_1 = H_2$, then $M_1 = M_2$ and $\{(\pi_{1m}, f_{\theta_{1m}})\}_{m=1}^{M_1}$ is a permutation of $\{(\pi_{2m}, f_{\theta_{2m}})\}_{m=1}^{M_2}$.

Proof: To state the identifiability of GFMM, according to the work of Atienza in [19], we will prove that given a linear transform $\mathcal{M}: f_\theta(x) \rightarrow \phi_f$ with domain $S(f)$ and a point u_0 in $S_0(f) = \{u \in S(f) : \phi_f \neq 0\}$, there exists a total order \prec on \mathcal{F} such that

$$f_1 \prec f_2 \iff \lim_{t \rightarrow t_0} \frac{\phi_{f_2}(u)}{\phi_{f_1}(u)} = 0 \quad (5)$$

for any two GFDs $f_1, f_2 \in \mathcal{F}$.

Let \mathcal{M} be a linear mapping which transforms a distribution $f \in \mathcal{F}$ into the moment generating function ϕ_f of $\log X$, i.e.,

$$\begin{aligned} \mathcal{M}[f_\theta(x)] : \phi_f(u) &= E(e^{u \log x}) \\ &= E(x^u) = \int_0^{+\infty} x^u f_\theta(x) dx. \end{aligned} \quad (6)$$

Then by analogy with the derivation process for a Mellin transform of GFD in [5], substituting $f_\theta(x) = p(x|\nu, \kappa, \sigma)$ of (1) into (6) yields

$$\phi_f(u) = \sigma^u \frac{\Gamma(\kappa + \frac{u}{\nu})}{\Gamma(\kappa)}, \quad u \in (-\kappa\nu, +\infty). \quad (7)$$

Clearly, $S_0(f) = (-\kappa\nu, +\infty)$ and $u_0 = +\infty$ verify (2) in **Corollary 1** of [19]. Next, we proceed to show that (5) defines a total order on \mathcal{F} .

From Stirling's formula $\Gamma(z+1) \sim \sqrt{2\pi z}(z/e)^z$ for $z \rightarrow +\infty$, we have

$$\begin{aligned} \phi_f(u) &\sim \frac{\sigma^u}{\Gamma(\kappa)} \sqrt{2\pi} \left(\frac{u}{\nu}\right)^{\kappa + \frac{u}{\nu} - \frac{1}{2}} \\ &\quad \times \left(1 + \frac{\kappa - 1}{u} \nu\right)^{\kappa + \frac{u}{\nu} - \frac{1}{2}} \exp\left\{1 - \kappa - \frac{u}{\nu}\right\} \\ &\sim \frac{\sqrt{2\pi}}{\Gamma(\kappa)} \exp\left\{u \log \sigma - \frac{u}{\nu}\right\} \\ &\quad \times \exp\left\{\left(\kappa + \frac{u}{\nu} - \frac{1}{2}\right) (\log u - \log \nu)\right\}, \end{aligned} \quad (8)$$

whenever $u \rightarrow +\infty$. The sign \sim means that the expressions on both sides are equivalent at $u \rightarrow +\infty$ up to a constant multiplicative factor. Therefore, for $u \rightarrow u_0$

$$\begin{aligned} \frac{\phi_{f_2}(u)}{\phi_{f_1}(u)} \sim C \exp \left\{ \left(\frac{1}{\nu_2} - \frac{1}{\nu_1} \right) u \log u + \left[(\log \sigma_2 - \log \sigma_1) \right. \right. \\ \left. \left. - \left(\frac{1}{\nu_2} - \frac{1}{\nu_1} \right) + \left(\frac{1}{\nu_2} \log \frac{1}{\nu_2} - \frac{1}{\nu_1} \log \frac{1}{\nu_1} \right) \right] u \right. \\ \left. + (\kappa_2 - \kappa_1) \log u \right\} \end{aligned} \quad (9)$$

with C a positive constant. Note, that the summands in the exponent are written in descending order, i.e., the first being the greatest for large u . So the ordering (5) can be constructed as $f_1 \prec f_2$ if and only if $[\nu_2 > \nu_1]$, or $[\nu_2 = \nu_1, \sigma_2 < \sigma_1]$, or $[\nu_2 = \nu_1, \sigma_2 = \sigma_1, \kappa_2 < \kappa_1]$, which is obviously a total order in \mathcal{F} . Thus, the GFMMs are identifiable. ■

III. ML ESTIMATION OF Θ WITH HISTOGRAM-BASED ECM

In the context of applications in modeling high-resolution SAR images, the crucial step for a M -component GFMM is to estimate the underlying parameters Θ . The general choice for deriving parameters from observations X is the maximum-likelihood (ML) estimation due to its desirable mathematical properties, such as consistency, asymptotic normality, and efficiency. The goal is to estimate Θ by maximizing the log-likelihood function such that

$$\hat{\Theta}_{ML} = \arg \max_{\Theta} \mathcal{L}(X, \Theta) \quad (10)$$

with

$$\begin{aligned} \mathcal{L}(X, \Theta) &= \log p(X|\Theta) \\ &= \sum_{i=1}^N \log \left(\sum_{m=1}^M \pi_m p(x_i|\theta_m) \right). \end{aligned} \quad (11)$$

To achieve a significant reduction of computational cost, (11) can be equivalently reformulated as a function of image gray levels, i.e.,

$$\begin{aligned} \mathcal{L}(X, \Theta) &= \mathcal{L}(Y, \Theta) = \log p(Y|\Theta) \\ &= \sum_{r=0}^{L-1} h(r) \log \left(\sum_{m=1}^M \pi_m p(r|\theta_m) \right) \end{aligned} \quad (12)$$

where $Y = \{h(r) : r \in \mathcal{D}\}$ is the non-normalized histogram of X , denoting the number of pixels with $x_i = r$. For example, there are $N = 262144$ image pixels for the case of a 512×512 image with 8-bit quantization, whereas only $L = 256$ distinct gray levels are involved in calculating the log-likelihood function. Clearly, the direct maximization of (12) is a difficult task because of the form of (1) and a logarithmic function of a sum of terms involved in (12).

To address this issue, we resort to the EM algorithm [11], [20], [21], which interprets Y as incomplete data with the missing information being a corresponding set of labels $Z = \{z^{(r)} | r = 1, \dots, L-1\}$. Each label is a binary vector $z^{(r)} = [z_1^{(r)}, \dots, z_M^{(r)}]$, where $z_m^{(r)} = 1$ and $z_k^{(r)} = 0$ for $k \neq m$ indicate that gray level r is produced by the m^{th} component of the mixture. Accordingly, Y is augmented by Z to form a complete data set. In such case, the resulting complete-data log-likelihood is given by

$$\begin{aligned} \mathcal{L}(Y, Z, \Theta) &= \log p(Y, Z|\Theta) \\ &= \sum_{m=1}^M \sum_{r=0}^{L-1} z_m^{(r)} h(r) \log \left(\pi_m p(r|\theta_m) \right). \end{aligned} \quad (13)$$

The EM algorithm generates a sequence of approximate ML estimations of the set of parameters $\{\hat{\Theta}(t), t = 1, 2, \dots\}$ by alternating an expectation step and a maximization one, i.e.,

$$\frac{\partial Q(\Theta, \hat{\Theta}(t))}{\partial \nu_m} = \sum_{r=0}^{L-1} \left\{ \left[\left(\kappa_m - \left(\frac{r}{\sigma_m} \right)^{\nu_m} \right) \log \left(\frac{r}{\sigma_m} \right) + \frac{1}{\nu_m} \right] \cdot p(m|r, \hat{\Theta}(t)) h(r) \right\} \quad (20)$$

$$\frac{\partial Q(\Theta, \hat{\Theta}(t))}{\partial \kappa_m} = \sum_{r=0}^{L-1} \left\{ \left[\nu_m \log \left(\frac{r}{\sigma_m} \right) - \Phi_0(\kappa_m) \right] \cdot p(m|r, \hat{\Theta}(t)) h(r) \right\} \quad (21)$$

$$\frac{\partial Q(\Theta, \hat{\Theta}(t))}{\partial \sigma_m} = \frac{\nu_m}{\sigma_m} \sum_{r=0}^{L-1} \left\{ \left[\left(\frac{r}{\sigma_m} \right)^{\nu_m} - \kappa_m \right] \cdot p(m|r, \hat{\Theta}(t)) h(r) \right\} \quad (22)$$

1) **E-step**: compute the conditional expectation of the complete log-likelihood

$$Q(\Theta, \hat{\Theta}(t)) = E \left[\log p(Y, Z|\Theta) | Y, \hat{\Theta}(t) \right], \quad (14)$$

given Y and the current estimate $\hat{\Theta}(t)$;

2) **M-step**: update the parameter estimates according to

$$\hat{\Theta}(t+1) = \arg \max_{\Theta} Q(\Theta, \hat{\Theta}(t)); \quad (15)$$

until convergence is achieved.

Specifically, in the E-step, Q function can be written as follows

$$\begin{aligned} Q(\Theta, \hat{\Theta}(t)) &= \sum_{m=1}^M \sum_{r=0}^{L-1} p(m|r, \hat{\Theta}(t)) h(r) \log(\pi_m) \\ &+ \sum_{m=1}^M \sum_{r=0}^{L-1} p(m|r, \hat{\Theta}(t)) h(r) \log(p(r|\theta_m)) \end{aligned} \quad (16)$$

with $p(m|r, \hat{\Theta}(t))$ being the posterior probability of r belonging to the m^{th} component, given by

$$p(m|r, \hat{\Theta}(t)) = E \left[z_m^{(r)} | r, \hat{\Theta}(t) \right] = \frac{\hat{\pi}_m(t) p(r|\hat{\theta}_m(t))}{\sum_{k=1}^M \hat{\pi}_k(t) p(r|\hat{\theta}_k(t))}. \quad (17)$$

In the M-step, we see from (16) that the Q function contains two independent terms, one depending on π_m and the other on $\theta_m = \{\nu_m, \kappa_m, \sigma_m\}$, which, hence, can be maximized separately. Firstly, we introduce the Lagrange multiplier λ with the constraint $\sum_{m=1}^M \pi_m = 1$, and find the estimate of π_m as follows

$$\begin{aligned} \frac{\partial}{\partial \pi_m} \left[\sum_{m=1}^M \sum_{r=0}^{L-1} p(m|r, \hat{\Theta}(t)) h(r) \log(\pi_m) \right. \\ \left. + \lambda \left(\sum_{m=1}^M \pi_m - 1 \right) \right] = 0. \end{aligned} \quad (18)$$

Using the fact that $\sum_{m=1}^M p(m|r, \hat{\Theta}(t)) = 1$, we get that $\lambda = -\sum_{r=0}^{L-1} h(r) = -N$ resulting in

$$\hat{\pi}_m(t+1) = \frac{1}{N} \sum_{r=0}^{L-1} p(m|r, \hat{\Theta}(t)) h(r). \quad (19)$$

Then, substituting (1) into (16) and taking the first derivative of $Q(\Theta, \hat{\Theta}(t))$ with respect to θ_m (i.e., ν_m , κ_m , and σ_m) give us the equations (20), (21), (22) (at the top of next page) with $\Phi_0(\bullet)$ being the Digamma function [22]. Owing to the parameter coupling and the presence of some complex terms in (20), (21), and (22), we cannot obtain the closed-form solution of θ_m from $\frac{\partial Q(\Theta, \hat{\Theta}(t))}{\partial \theta_m} = \mathbf{0}$, and thus numerical iteration is required. As such, the EM algorithm becomes less attractive for the case of GFMM.

$$\hat{\nu}_m(t+1) = \frac{\sum_{r=0}^{L-1} \left[\hat{\nu}_m^2(t) \left(\frac{r}{\hat{\sigma}_m(t)} \right)^{\hat{\nu}_m(t)} \log^2 \left(\frac{r}{\hat{\sigma}_m(t)} \right) + 1 \right] p(m|r, \hat{\Theta}(t)) h(r)}{\sum_{r=0}^{L-1} \left[\gamma_m(t) \left(\frac{r}{\hat{\sigma}_m(t)} \right)^{\hat{\nu}_m(t)} - \hat{\kappa}_m(t) \right] \log \left(\frac{r}{\hat{\sigma}_m(t)} \right) p(m|r, \hat{\Theta}(t)) h(r)} \quad (26)$$

Nevertheless, from (20), (21), and (22), it is not difficult to find that the complete-data ML estimation of θ_m is relatively simple, if maximization is undertaken conditionally on some of the parameters. As a consequence, we focus on the use of the ECM algorithm [23] for the estimation of GFMM, which takes advantage of the simplicity of the complete-data conditional ML estimation. On the $(t+1)^{th}$ iteration of the ECM algorithm, the E-step is the same as given above for the EM algorithm, while a complicated M-step of the latter is replaced by S computationally simpler CM-steps, i.e.,

$$\hat{\Theta}(t+s/S) = \arg \max_{\Theta} Q(\Theta, \hat{\Theta}(t)), s = 1, 2, \dots, S \quad (27)$$

subject to the constraint $g_s(\Theta) = g_s(\hat{\Theta}(t+(s-1)/S))$. Here $G = \{g_s(\Theta), s = 1, 2, \dots, S\}$ is a set of S preselected (vector) functions of Θ . Naturally, $\hat{\Theta}(t+s/S)$ satisfies

$$Q(\hat{\Theta}(t+s/S), \hat{\Theta}(t)) \geq Q(\Theta, \hat{\Theta}(t)) \quad (24)$$

for all $\Theta \in \Omega_s(\hat{\Theta}(t+(s-1)/S))$, with

$$\begin{aligned} \Omega_s(\hat{\Theta}(t+(s-1)/S)) \\ \equiv \left\{ \Theta \in \Omega : g_s(\Theta) = g_s(\hat{\Theta}(t+(s-1)/S)) \right\}. \end{aligned} \quad (25)$$

The output of the final CM-step is then defined as $\hat{\Theta}(t+S/S) = \hat{\Theta}(t+1)$.

In so doing, we partition the vector Θ of unknown parameters in GFMM into $S = 4$ subvectors $\Theta_1 = \{\pi_1, \pi_2, \dots, \pi_M\}$, $\Theta_2 = \{\nu_1, \nu_2, \dots, \nu_M\}$, $\Theta_3 = \{\kappa_1, \kappa_2, \dots, \kappa_M\}$, and $\Theta_4 = \{\sigma_1, \sigma_2, \dots, \sigma_M\}$. Also let us define the constraint spaces $\Omega_s(\hat{\Theta}(t+(s-1)/S))$ for $s = 1, 2, 3, 4$ having $g_1(\Theta) = \{\hat{\Theta}_2(t), \hat{\Theta}_3(t), \hat{\Theta}_4(t)\}$, $g_2(\Theta) = \{\hat{\Theta}_1(t+1), \hat{\Theta}_3(t), \hat{\Theta}_4(t)\}$, $g_3(\Theta) = \{\hat{\Theta}_1(t+1), \hat{\Theta}_2(t+1), \hat{\Theta}_4(t)\}$, and $g_4(\Theta) = \{\hat{\Theta}_1(t+1), \hat{\Theta}_2(t+1), \hat{\Theta}_3(t+1)\}$. From (23), the s^{th} CM-step requires the maximization of $Q(\Theta, \hat{\Theta}(t))$ with respect to the s^{th} subvector Θ_s with the other $(S-1)$ subvectors held fixed at their current values. Then the CM-step 1 for calculating $\hat{\Theta}_1(t+1)$ can be implemented by proceeding as with (19). As for other three CM-steps, it is not difficult to show that for $\forall m = 1, 2, \dots, M$, $\hat{\nu}_m(t+1)$ can be calculated using the generalized Newton method with a non-quadratic approximation [25] for $\Theta \in \Omega_2(\hat{\Theta}(t))$, given by¹ (26) (at the top of next page) where $\gamma_m(t) = \hat{\nu}_m(t) \log \left(\frac{r}{\hat{\sigma}_m(t)} \right) + 1$. Similarly, $\hat{\kappa}_m(t+1)$ is a solution of the following equation

$$\Phi_0(\kappa_m) = \frac{\hat{\nu}_m(t+1) \sum_{r=0}^{L-1} \log \left(\frac{r}{\hat{\sigma}_m(t)} \right) p(m|r, \hat{\Theta}(t)) h(r)}{\sum_{r=0}^{L-1} p(m|r, \hat{\Theta}(t)) h(r)}. \quad (27)$$

This equation can be iteratively solved using a simple bisection method [26] to yield $\hat{\kappa}_m(t+1)$, since the Digamma function $\Phi_0(\bullet)$ is strictly monotonically increasing on \mathbb{R}^+ . Further, we also have

$$\hat{\sigma}_m(t+1) = \left[\frac{\sum_{r=0}^{L-1} r^{\hat{\nu}_m(t+1)} p(m|r, \hat{\Theta}(t)) h(r)}{\hat{\kappa}_m(t+1) \sum_{r=0}^{L-1} p(m|r, \hat{\Theta}(t)) h(r)} \right]^{\frac{1}{\hat{\nu}_m(t+1)}} \quad (28)$$

for each m ($m = 1, 2, \dots, M$).

In detail, the proposed histogram-based ECM algorithm for the estimation of GFMM parameters performs an E-step followed by four successive CM-steps.

1) **E-step:** Calculate the posterior probabilities $\left\{ p(m|r, \hat{\Theta}(t)), m = 1, 2, \dots, M \right\}$ using (17), given Y and $\hat{\Theta}(t)$.

¹Here, with $\frac{\partial^2}{\partial \nu_m^2} Q(\Theta, \hat{\Theta}(t)) < 0$, this update resembles Newton-Raphson, but converges faster and fails less often [25].

2) CM-steps

- **CM-step 1:** Calculate $\hat{\Theta}_1(t+1)$ by using (19) with Θ_2 fixed at $\hat{\Theta}_2(t)$, Θ_3 fixed at $\hat{\Theta}_3(t)$, and Θ_4 fixed at $\hat{\Theta}_4(t)$.
- **CM-step 2:** Calculate $\hat{\Theta}_2(t+1)$ by using (26) with Θ_1 fixed at $\hat{\Theta}_1(t+1)$, Θ_3 fixed at $\hat{\Theta}_3(t)$, and Θ_4 fixed at $\hat{\Theta}_4(t)$.
- **CM-step 3:** Calculate $\hat{\Theta}_3(t+1)$ by solving (27) via bisection method with Θ_1 fixed at $\hat{\Theta}_1(t+1)$, Θ_2 fixed at $\hat{\Theta}_2(t+1)$, and Θ_4 fixed at $\hat{\Theta}_4(t)$.
- **CM-step 4:** Calculate $\hat{\Theta}_4(t+1)$ by using (28) with Θ_1 fixed at $\hat{\Theta}_1(t+1)$, Θ_2 fixed at $\hat{\Theta}_2(t+1)$, and Θ_3 fixed at $\hat{\Theta}_3(t+1)$.

A CM-step of the algorithm described above might be in closed form or it might itself require iteration, but because these CMs are over smaller dimensional spaces, they are simpler, faster, and more stable than the full maximizations (e.g., jointly solve $\frac{\partial Q(\Theta, \hat{\Theta}(t))}{\partial \pi_m = 0}$ and $\frac{\partial Q(\Theta, \hat{\Theta}(t))}{\partial \theta_m = 0}$) in the original M-step of the EM algorithm. Meanwhile, we have the following theorem to fulfill.

Theorem 2: Let $\mathcal{A} : \hat{\Theta}(t+1) = \mathcal{A}(\hat{\Theta}(t))$ denotes an update operation of the ECM algorithm for parameter set Θ of GFMM, which increases the log-likelihood of the observed-data model after each iteration, i.e., $\log p(Y|\hat{\Theta}(t+1)) \geq \log p(Y|\hat{\Theta}(t))$ for all t . If $\log p(Y|\Theta)$ is bounded, then $\log p(Y|\Theta(t)) \rightarrow \mathcal{L}^* = \mathcal{L}(Y, \Theta^*)$ for some stationary point Θ^* .

Proof: From (24) and with $S = 4$, we have

$$\begin{aligned} Q(\hat{\Theta}(t+1), \hat{\Theta}(t)) &\geq Q(\hat{\Theta}(t+3/4), \hat{\Theta}(t)) \geq Q(\hat{\Theta}(t+2/4), \hat{\Theta}(t)) \\ &\geq Q(\hat{\Theta}(t+1/4), \hat{\Theta}(t)) \geq Q(\hat{\Theta}(t), \hat{\Theta}(t)). \end{aligned} \quad (29)$$

The inequality (29) is a sufficient condition [20] for

$$\mathcal{L}(Y, \hat{\Theta}(t+1)) \geq \mathcal{L}(Y, \hat{\Theta}(t)) \quad (30)$$

to hold. In other words, this ECM algorithm monotonically increases the log-likelihood after each CM-step, and hence, after each iteration. Further, any ECM algorithm is a generalized EM (GEM) [23] and therefore, any property established for GEM holds for ECM. Thus, according to Theorem 1 of [27], when the ECM sequence of log-likelihood values $\{\mathcal{L}(Y, \hat{\Theta}(t))\}$ is bounded above, $\mathcal{L}(Y, \hat{\Theta}(t))$ converges monotonically to a finite limit \mathcal{L}^* . ■

IV. UNSUPERVISED LEARNING OF GFMM

In Section III, during the derivation of the histogram-based ECM algorithm for ML estimation of Θ , we assumed that the number of mixture components M is known in advance. Often in practice, this is not the case. Indeed, such prior information is generally unavailable, and a mixture model with an inappropriate number of components tends to overfit or underfit the data. The determination of M is a typical model selection problem. To automatically identify an optimal M , many deterministic approaches have been suggested from different perspectives, for example, Bayesian information criterion (BIC) [28], Akaike's information criterion (AIC) [29], minimum description length (MDL) [30], minimum message length (MML) [31], [32], and integrated completed likelihood (ICL) [33]. They all adopt the "model-class/model" hierarchy to select M , i.e.,

$$\hat{M} = \arg \min_M \left\{ \mathcal{C}(\hat{\Theta}(M), M), M = M_{min}, \dots, M_{max} \right\} \quad (31)$$

where a set of candidate mixtures are induced for each model-class (i.e., to estimate $\hat{\Theta}(M)$ for $M \in [M_{min}, M_{max}]$), and then the "best" model is selected according to the underlying selection criterion $\mathcal{C}(\hat{\Theta}(M), M)$. To abandon such a hierarchy, Figueiredo and Jain [16] developed an unsupervised algorithm based on a MML-like criterion for learning a FMM, whose idea is to directly find the "best" overall model by integrating parameter estimation and model selection in a single EM algorithm. We now discuss how to extend this algorithm to perform an unsupervised learning of GFMM based on histogram.

To simultaneously determine M and Θ , we start with the following criterion adopted in [16]

$$\hat{\Theta}(M) = \arg \min_{M, \Theta} \left\{ -\log p(\Theta) - \log p(X|\Theta) + \frac{1}{2} \log |I(\Theta)| + \frac{c}{2} \left(1 + \log \frac{1}{12} \right) \right\} \quad (32)$$

where $I(\Theta)$ is the expected Fisher information matrix, $|I(\Theta)|$ denotes its determinant, and $c = 3M + M$ is the number of free parameters in G Γ MM. Similarly, to reduce the difficulty of computing $I(\Theta)$, $I(\Theta)$ is replaced by the complete-data Fisher information matrix $I_c(\Theta) \equiv -E [\nabla_{\Theta}^2 \log p(X, Z|\Theta)]$ having the block-diagonal structure [34]

$$I_c(\Theta) = N \text{block-diag} \left\{ \pi_1 I^{(1)}(\theta_1), \dots, \pi_M I^{(1)}(\theta_M), \mathbf{A} \right\}. \quad (33)$$

Here, $I^{(1)}(\theta_m)$ is the Fisher information matrix for a single observation associated with the m^{th} component, and \mathbf{A} is the Fisher information matrix of a multinomial distribution with parameters (π_1, \dots, π_M) . Based on the independency assumption, the prior on the parameter set is taken as

$$p(\Theta) = p(\pi_1, \dots, \pi_M) \prod_{m=1}^M p(\theta_m) \quad (34)$$

where a non-informative Jeffreys' prior [35] is imposed on θ_m and π_m s as $p(\theta_m) \propto \sqrt{|I^{(1)}(\theta_m)|}$ and $p(\pi_1, \dots, \pi_M) \propto \sqrt{|\mathbf{A}|} = (\pi_1 \pi_1 \dots \pi_M)^{-1/2}$ in absence of any other knowledge.

By substituting (33), (34) into (32), we finally obtain

$$\hat{\Theta}(M) = \arg \min_{M, \Theta} \mathcal{L}_{MML}(X, \Theta) \quad (35)$$

with

$$\begin{aligned} \mathcal{L}_{MML}(X, \Theta) &= \frac{3}{2} \sum_{m: \pi_m > 0} \log \left(\frac{N \pi_m}{12} \right) \\ &+ \frac{K_{nz}}{2} \log \frac{N}{12} + 2K_{nz} - \log p(X|\Theta) \end{aligned} \quad (36)$$

where $K_{nz} = \sum_m [\pi_m > 0]$ denotes the number of non-zero-probability components. With an equivalence relation defined in (12), (36) becomes

$$\begin{aligned} \mathcal{L}_{MML}(X, \Theta) &= \mathcal{L}_{MML}(Y, \Theta) \\ &= \frac{3}{2} \sum_{m: \pi_m > 0} \log \left(\frac{N \pi_m}{12} \right) \\ &+ \frac{K_{nz}}{2} \log \frac{N}{12} + 2K_{nz} - \log p(Y|\Theta) \end{aligned} \quad (37)$$

which is essentially an incomplete-data penalized log-likelihood function [36].

Building on the maximization for (12) established in Section III, the EM algorithm is applied to maximize $-\mathcal{L}_{MML}(Y, \Theta)$ (equivalent to minimizing (37)). With data augmentation, the complete-data penalized log-likelihood is similarly found to be

$$\begin{aligned} -\mathcal{L}_{MML}(Y, Z, \Theta) &= \sum_{m: \pi_m > 0} \sum_{r=0}^{L-1} z_m^{(r)} h(r) \log \left(\pi_m p(r|\theta_m) \right) \\ &- \frac{3}{2} \sum_{m: \pi_m > 0} \log \left(\frac{N \pi_m}{12} \right) - \frac{K_{nz}}{2} \log \frac{N}{12} - 2K_{nz}. \end{aligned} \quad (38)$$

Concerning the E-step, the corresponding Q function now is

$$\begin{aligned}
 Q_{MML}(\Theta, \hat{\Theta}(t)) &= \sum_{m:\pi_m>0} \sum_{r=0}^{L-1} p(m|r, \hat{\Theta}(t)) h(r) \log(\pi_m) \\
 &+ \sum_{m:\pi_m>0} \sum_{r=0}^{L-1} p(m|r, \hat{\Theta}(t)) h(r) \log(p(r|\theta_m)) \\
 &- \frac{3}{2} \sum_{m:\pi_m>0} \log\left(\frac{N\pi_m}{12}\right) - \frac{K_{nz}}{2} \log \frac{N}{12} - 2K_{nz}
 \end{aligned} \tag{39}$$

with $p(m|r, \hat{\Theta}(t))$ given in (17).

Then we proceed to the M-step. From (39), it is observed that, compared with (16), none of the three new added terms involves θ_m , only the term $-\frac{3}{2} \sum_{m:\pi_m>0} \log\left(\frac{N\pi_m}{12}\right)$ depends on π_m . Thus, maximizing Q_{MML} with respect to θ_m is equivalent to maximizing the Q function of (16), and similarly, as above, π_m and θ_m still can be estimated separately. Proceeding as before with derivation of (19), the updated estimate of π_m at the $(t+1)^{th}$ iteration becomes

$$\hat{\pi}_m(t+1) = \frac{\max\left\{0, \left[\sum_{r=0}^{L-1} p(m|r, \hat{\Theta}(t)) h(r)\right] - \frac{3}{2}\right\}}{\sum_{k=1}^M \max\left\{0, \left[\sum_{r=0}^{L-1} p(m|r, \hat{\Theta}(t)) h(r)\right] - \frac{3}{2}\right\}}, \tag{40}$$

by maximizing the objective

$$\begin{aligned}
 &\sum_{m:\pi_m>0} \sum_{r=0}^{L-1} p(m|r, \hat{\Theta}(t)) h(r) \log(\pi_m) \\
 &- \frac{3}{2} \sum_{m:\pi_m>0} \log\left(\frac{N\pi_m}{12}\right)
 \end{aligned} \tag{41}$$

under the constraint $\sum_{m=1}^M \pi_m = 1$. Further, the update process of θ_m with $\hat{\pi}_m(t+1) > 0$ is the same as that proposed in Section III.

It is worth noting that (40), unlike (19), can automatically annihilate components that are not supported by the data during the learning process, which also inexplicitly avoids the possibility of convergence to the boundary of the parameter space. In the pruning paradigm, this algorithm starts with a large number of components all over the space to determine the true/optimal M , thus alleviating the need for a good initialization associated with the EM/ECM algorithms. Because of this, a component-wise EM algorithm [37] has to be adopted to sequentially update π_m and θ_m , as in [16]. Once $\hat{\pi}_m(t+1) = 0$ for one component, its probability mass is immediately redistributed to the remaining components for increasing their chance of survival. We summarize the unsupervised learning algorithm of GFMM in Algorithm 1.

From Algorithm 1, the specific implementation of HECM-FJ-GFMM includes two nested loops: the inner loop and the outer loop. The former, corresponding to lines 9-23, performs the histogram-based ECM algorithm to obtain the ML estimate of Θ in a component-wise manner for each K_{nz} , and implements the component annihilation of (40). The absolute variation of $\mathcal{L}_{MML}(Y, \Theta)$ is used as a stopping criterion for this inner loop, in which the tolerance ϵ determines the tradeoff between accuracy and time complexity. To consider the additional decrease of $\mathcal{L}_{MML}(Y, \Theta)$ caused by the decrease in K_{nz} , the outer loop, in lines 8-29, iterates over K_{nz} from M_{max} to M_{min} by successively annihilating the least probable component. Finally, we choose the estimates that lead to the minimum value of $\mathcal{L}_{MML}(Y, \Theta)$.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Datasets for Experiments

In order to demonstrate the effectiveness of our proposed model, we perform experiments on high-resolution amplitude SAR datasets. To provide a relatively large validation set, different typologies of data are considered, including satellite and airborne SAR systems, with different resolutions, frequencies, and polarimetric modes:

Algorithm 1 HECM-FJ-GFMM

```

1: Inputs:  $Y, N, M_{min}, M_{max}, \epsilon$ 
2: Output: Mixture model in  $\hat{\Theta}_{best}$ 
3: Initialization
4:  $t \leftarrow 0, K_{nz} \leftarrow M_{max}, \mathcal{L}_{min} \leftarrow +\infty$ 
5: Set initial parameters  $\hat{\Theta}(0) = \{\hat{\pi}_1, \dots, \hat{\pi}_{M_{max}}, \hat{\theta}_1, \dots, \hat{\theta}_{M_{max}}\}$ 
6: Calculate  $p(r|\hat{\theta}_m)$  for  $m = 1, \dots, M_{max}$ , and  $r = 0, \dots, L-1$ 
7: Main Loop
8: While:  $K_{nz} \geq M_{min}$  do
9:   Repeat
10:    for  $m = 1$  to  $M_{max}$  do
11:      Calculate  $p(m|r, \hat{\Theta}(t))$  using (17), for  $r = 0, \dots, L-1$ 
12:      Calculate  $\hat{\pi}_m(t+1)$  using (40)
13:       $\{\hat{\pi}_1, \dots, \hat{\pi}_{M_{max}}\} \leftarrow \{\hat{\pi}_1, \dots, \hat{\pi}_{M_{max}}\} (\sum_{m=1}^{M_{max}} \hat{\pi}_m)^{-1}$ 
14:      if  $\hat{\pi}_m > 0$ 
15:        Update  $\hat{\theta}_m(t+1)$  using (26), (27), (28) successively
16:        Calculate  $p(r|\hat{\theta}_m)$  for  $r = 0, \dots, L-1$ 
17:      else  $K_{nz} \leftarrow K_{nz} - 1$ 
18:      end if
19:    end for
20:     $\hat{\Theta}(t+1) \leftarrow \{\hat{\pi}_1, \dots, \hat{\pi}_{M_{max}}, \hat{\theta}_1, \dots, \hat{\theta}_{M_{max}}\}$ 
21:    Calculate  $\mathcal{L}_{MML}(Y, \hat{\Theta}(t+1))$  using (37)
22:     $t \leftarrow t + 1$ 
23:  until  $\mathcal{L}_{MML}(Y, \hat{\Theta}(t-1)) - \mathcal{L}_{MML}(Y, \hat{\Theta}(t)) < \epsilon$ 
24:  if  $\mathcal{L}_{MML}(Y, \hat{\Theta}(t)) \leq \mathcal{L}_{min}$  then
25:     $\mathcal{L}_{min} \leftarrow \mathcal{L}_{MML}(Y, \hat{\Theta}(t))$ 
26:     $\hat{\Theta}_{best} \leftarrow \hat{\Theta}(t)$ 
27:  end if
28:   $m^* \leftarrow \arg \min_m \{\hat{\pi}_m > 0\}, \hat{\pi}_{m^*} \leftarrow 0, K_{nz} \leftarrow K_{nz} - 1$ 
29: end while

```

- An excerpt from the TerraSAR-X image acquired in high-resolution SpotLight mode over the Port of Visakhapatnam on India's eastern coast (India, VV polarization, geocoded ellipsoid corrected, $0.50\text{m} \times 0.50\text{m}$ pixel spacing, 0.99m (ground range) $\times 1.12\text{m}$ (azimuth) resolution).
- Two sections of the X-band StripMap SAR image taken over Sanchagang near Poyang Lake also by the TerraSAR-X sensor (China, respectively HH- and VV-polarized, enhanced ellipsoid corrected, $2.75\text{m} \times 2.75\text{m}$ pixel spacing, 5.98m (ground range) $\times 6.11\text{m}$ (azimuth) resolution).
- A single-look scene of the COSMO-SkyMed image for the SpotLight acquisition mode in the region of Livorno (Italy, HH polarization, geocoded ellipsoid corrected, 1m (ground range) $\times 1\text{m}$ (azimuth) resolution).
- A multilook portion of an airborne RAMSES sensor acquisition over Toulouse suburbs (France, single polarization, downsampled to approximately 2m ground resolution).
- A portion of the HH channel amplitude image extracted from complex covariance format of L-band fully polarimetric data at the Foulum agricultural test site (Denmark) acquired by the Danish airborne EMISAR system whose nominal single-look spatial resolution is 2m for both ground range and azimuth directions.

Hereafter, for simplicity we refer to these test SAR images (shown in Fig. 2 and Fig. 3) as “India-Port”, “Sanchagang-HH”, “Sanchagang-VV”, “Livorno”, “Toulouse”, and “Foulum”.

B. PDF Estimation Results and Analysis

The proposed semi-parametric analysis method has been applied to all the selected SAR images and the resulting PDF estimates have been assessed both quantitatively and qualitatively. Measures for quantitative performance evaluation are the Kolmogorov-Smirnov (KS) distance [26] and the symmetric Kullback-Leibler (sKL) divergence [38]. Specifically, the KS distance, $D_{KS} = \max_{x \in \Omega} |F(x) - G(x)|$, defines the maximum absolute difference between the fitted cumulative distribution function (CDF) $F(x)$ and the empirical CDF $G(x)$, while the symmetric KL divergence, $D_{sKL} = \sum_{x \in \Omega} f(x) \log \frac{f(x)}{h(x)} + \sum_{x \in \Omega} h(x) \log \frac{h(x)}{f(x)}$, measures the dissimilarity between the estimated PDF $f(x)$ and the normalized histogram $h(x)$ from an information theory perspective. Both metrics reflect the significance level of discrepancy between two distributions, for which a small value indicates a better fit of the

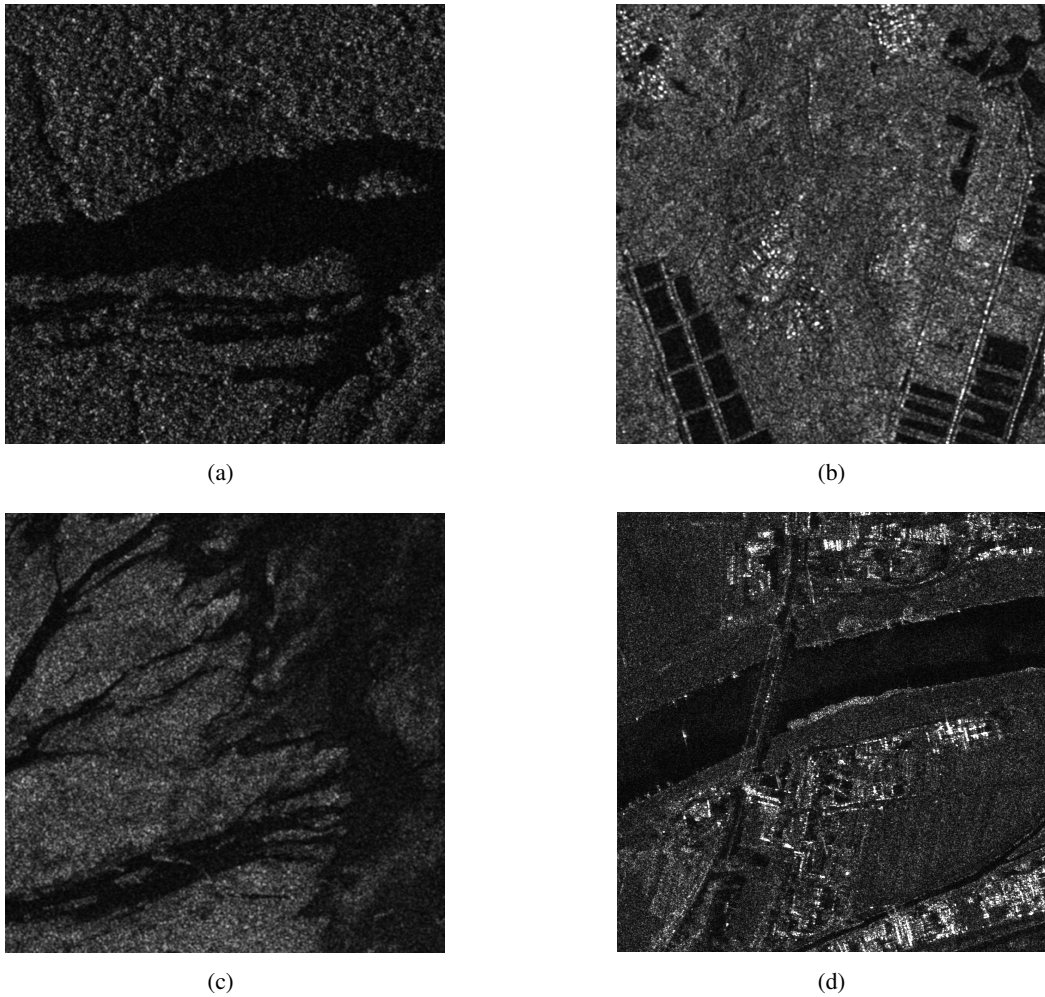


Fig. 2. Four spaceborne SAR images: (a) “India-Port” TerraSAR-X image (size: 512×512), (b) “Sanchagang-HH” TerraSAR-X image (size: 512×512), (c) “Sanchagang-VV” TerraSAR-X image (size: 512×512), and (d) “Livorno” COSMO-SkyMed image (size: 1024×1024).

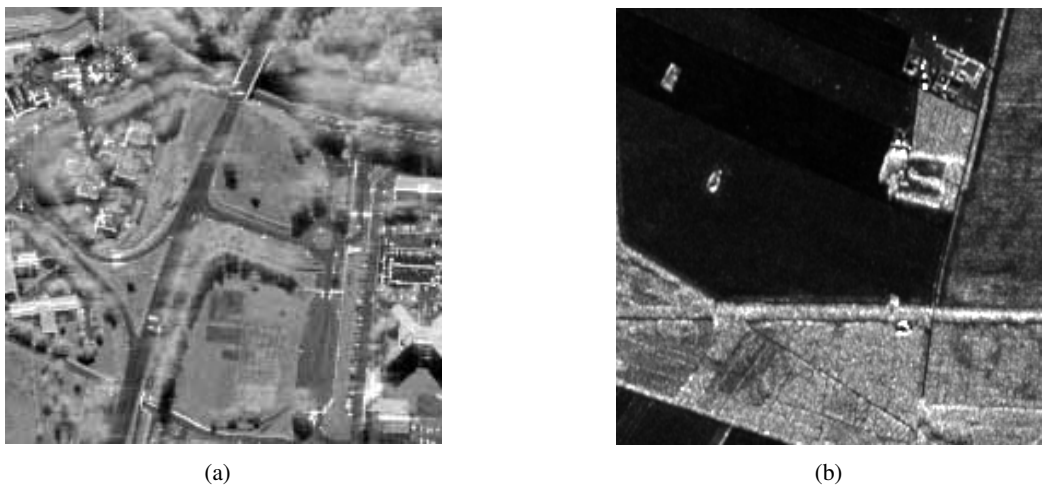


Fig. 3. Two airborne SAR images: (a) “Toulouse” RAMSES image (size: 200×200), and (b) “Foulum” EMISAR image (size: 256×256).

TABLE I
 QUANTITATIVE MEASURES OBTAINED BY GFD, 2NMM, EDSEM AND HECM-FJ-GFMM WITH THE OPTIMAL NUMBER OF MIXTURE COMPONENTS FOR SIX ACTUAL EXPERIMENTAL AMPLITUDE SAR IMAGES

| Image | GFD | | 2NMM | | EDSEM | | | | HECM-FJ-GFMM | | |
|-----------------|----------|-----------|----------|-----------|-------|--------------------------|----------|---------------|--------------|----------------|----------------|
| | D_{KS} | D_{sKL} | D_{KS} | D_{sKL} | M^* | Selected Models | D_{KS} | D_{sKL} | M^* | D_{KS} | D_{sKL} |
| “India-Port” | 0.0276 | 0.0395 | 0.0233 | 0.0679 | 2 | $\{f_6, f_1\}$ | 0.0027 | 0.0022 | 2 | 0.0020 | 0.0018 |
| “Sanchagang-HH” | 0.0319 | 0.0786 | 0.0686 | 0.3126 | 2 | $\{f_4, f_1\}$ | 0.0032 | 0.0048 | 4 | 0.0010 | 0.0019 |
| “Sanchagang-VV” | 0.0417 | 0.0946 | 0.0444 | 0.1387 | 2 | $\{f_1, f_4\}$ | 0.0027 | 0.0014 | 3 | 0.0024 | 0.0019 |
| “Livorno” | 0.0053 | 0.0060 | 0.0563 | 0.1122 | 4 | $\{f_4, f_6, f_1, f_4\}$ | 0.0023 | 0.0024 | 4 | 7.74e-4 | 8.33e-4 |
| “Toulouse” | 0.0486 | 0.0916 | 0.2388 | 7.1484 | 4 | $\{f_1, f_4, f_6, f_4\}$ | 0.0075 | 0.0195 | 5 | 0.0025 | 0.0107 |
| “Foulum” | 0.1356 | 0.6857 | 0.2268 | 5.6822 | 4 | $\{f_4, f_4, f_2, f_1\}$ | 0.0047 | 0.0167 | 6 | 0.0018 | 0.0076 |

particular distribution to empirical data. As far as the qualitative evaluation is concerned, we visually compare the plots of the estimates with those of histograms.

For comparison purpose, the results are compared with those obtained by three following state-of-the-art models: the GFD [5], the two-component Nakagami mixture model (2NMM) [39] and the enhanced dictionary-based SEM (EDSEM) approach [13]. As a single component case of GFMM, the GFD can achieve better goodness of fit than the existing parametric PDFs in most cases with the MoLC parameter estimates [5]. For the 2NMM, the number of mixture components is fixed to 2 in advance, and its noniterative MoLC estimates of the underlying parameters have closed-form expressions. In view of the existence of two roots for mixture weight, we choose the one which generates the better fit accuracy. The EDSEM adopts an enhanced dictionary of eight distinct PDFs for mixture components including log-normal, Weibull, Fisher, GFD, Nakagami, \mathcal{K} , GGR, and $S\alpha SGR$ (respectively, denoted by f_1, f_2, \dots, f_8 , and listed in Table I of [13]). In EDSEM, component annihilation along with the MoLC estimation is integrated into the SEM procedure for selecting the number of mixture components and providing the corresponding estimates of each component parameters. In the experiments, the initial maximum number of components for the EDSEM is set to 7, and further increase does not affect the results. As far as the HECM-FJ-GFMM is concerned, we initialize $\nu_m = 2$ and $\kappa_m = 1$ for $\forall m \in [1, M_{max}]$ (i.e., Rayleigh case) as well as equal weight $\pi_m = 1/M_{max}$, and uniformly choose M_{max} data points from available gray levels of a given SAR image as the mode of Rayleigh components to determine σ_m 's. M_{max} and ϵ are set as follows: 1) $M_{max} = 20$ for the “India_Port” and “Sanchagang-VV” images, $M_{max} = 30$ for the “Toulouse” and “Foulum” ones, all with $\epsilon = 3 \times 10^{-2}$; 2) $M_{max} = 40$, $\epsilon = 1 \times 10^{-2}$ for the “Sanchagang-HH” image, and $M_{max} = 40$, $\epsilon = 1 \times 10^{-3}$ for the “Livorno” image.

Table I lists the values of both quantitative results of metrics D_{KS} and D_{sKL} obtained by the GFD, 2NMM, EDSEM and HECM-FJ-GFMM approaches. Moreover, the normalized histograms and the estimated PDFs of GFD, EDSEM and HECM-FJ-GFMM for six test high-resolution SAR images are shown in Fig. 4, together with the best two cases of 2NMM in Fig. 5, for a visual comparison. Here, it should be noted that, for the “Toulouse” image, the second- and third-order sample second-kind cumulants don't satisfy the applicability condition of MoLC [40]: $\hat{k}_2 \geq 0.63|\hat{k}_3|^{2/3}$, so we make use of Song's SISE estimator [41] to obtain the estimates of GFD parameters.² We stress at this point, that due to the MoLC-applicability restrictions for this distribution, the use of EDSEM [13] for the estimation of GFD mixtures, i.e. restricting the dictionary to solely this PDF, is not feasible. Indeed, at some point of the iterative process the applicability condition is violated and the process stops prematurely.

Increasing the resolution causes a reduction of the number of scatterers within each resolution cell. Thus, the SAR images may contain more high intensity pixels, which lead to strong heavy-tailed effects of the underlying PDFs. Furthermore, more complex land-cover topologies are present in the high-resolution imagery, whose different

²This, in fact, means that the set of MoLC equations based on the GFD distribution assumption doesn't have a solution for the observed values of log-cumulants $(\hat{k}_1, \hat{k}_2, \hat{k}_3)$. Thus, the observed data present significant deviations from the GFD distribution. Nevertheless, the parameter estimates can still be obtained with other estimation methods, such as, SISE.

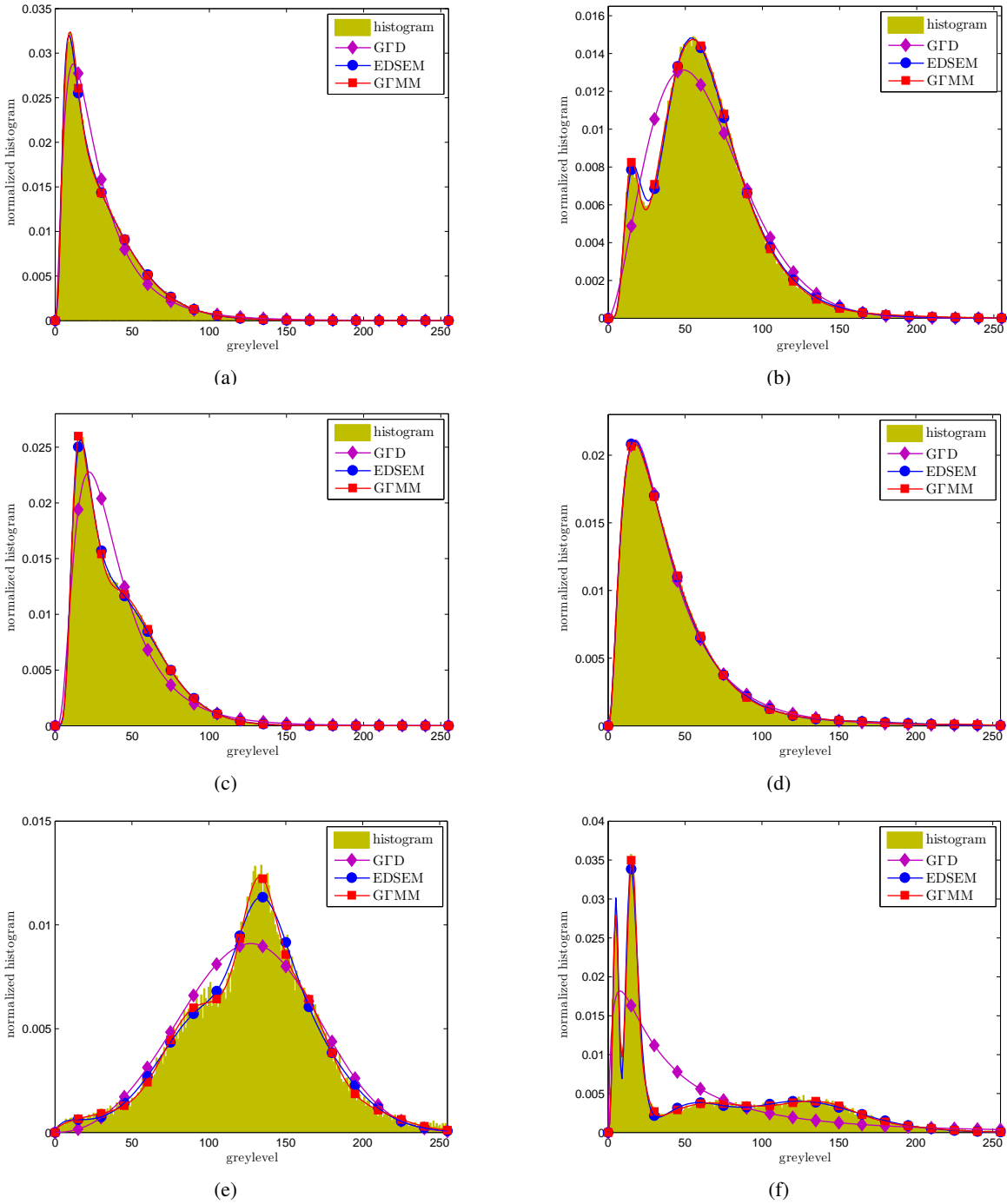


Fig. 4. Histograms and the estimated PDFs: (a) “India-Port” TerraSAR-X image, (b) “Sanchagang-HH” TerraSAR-X image, (c) “Sanchagang-VV” TerraSAR-X image, (d) “Livorno” COSMO-SkyMed image, (e) “Toulouse” RAMSES image, and (f) “Foulum” EMISAR image.

backscattering properties result in the complicated bi/multimodal statistics from the imaging mechanism of the SAR system. Fig. 4 provides evidence that the normalized histograms of the considered SAR images vary from unimodal to multimodal, and some of them exhibit heavy-tails. Since the GFD is intrinsically monomodal, it provides a good fit of the “Livorno” histogram as expected. However, it yields poor estimates for other images, whose values of D_{KS} and D_{sKL} are on average one order of magnitude larger than that of EDSEM and HECM-FJ-GFMM (see Table I). At the same time, the results of 2NMM are less accurate on the considered images, even for the “India-Port” and “Livorno” images with relatively simple histograms, as shown in Fig. 5. The reason for this behavior is the fixed number of components (equal to two) and the consequent lower number of degrees of freedom in the

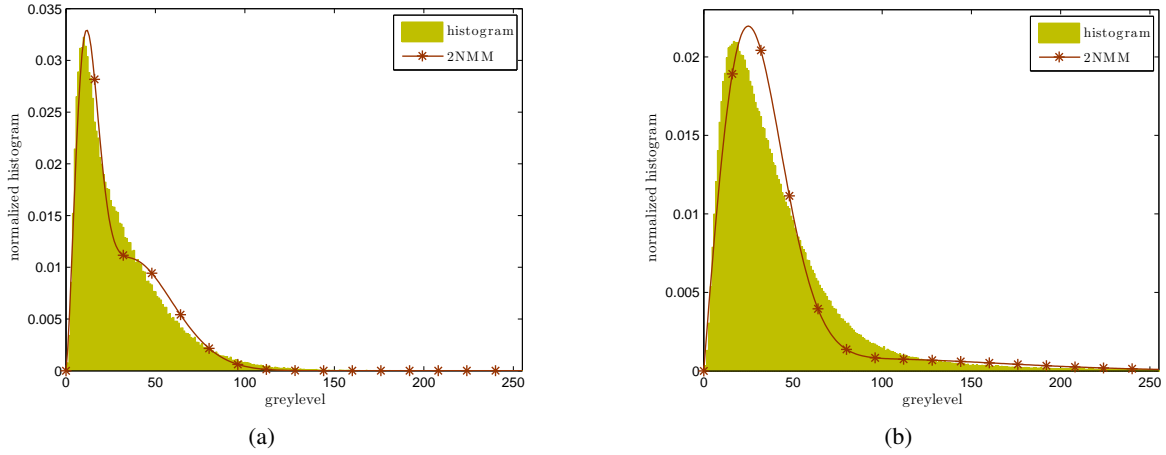


Fig. 5. Histograms and the estimated PDFs of 2NMM: (a) “India-Port” TerraSAR-X image, and (b) “Livorno” COSMO-SkyMed image.

model (three parameters). This underlines the usefulness of the adaptive semi-parametric approach adopted. The EDSEM can effectively detect the uni/bi/multimodal structures of the involving histograms and allows achieving high quantitative results. However, some obvious estimation biases around the peaks and/or valleys are presented for the relatively complex histograms, such as in the cases of the “Sanchagang-HH”, “Toulouse”, and “Foulum” images (see Fig. 4 (b), (e), and (f)). Especially for the latter two images, the histogram fit is more challenging. A comparison between the quantitative measures listed in Table I together with the visual analysis of PDF estimate plots demonstrates that the proposed HECM-FJ-G Γ MM performs best and can readily fit all of the SAR images. Only for the “Sanchagang-VV” image, the result of HECM-FJ-G Γ MM is inferior to that of EDSEM in terms of D_{sKL} , but its D_{KS} value is slightly better than that of EDSEM. Therefore, we conclude that the proposed HECM-FJ-G Γ MM technique provides very accurate and competitive PDF estimates, and is a powerful technique for modeling the high-resolution SAR images.

To highlight the merit of HECM-FJ-G Γ MM, we point out the relevance of automatically selecting an optimal number M^* of mixture components of the proposed method. Since we lack the ground truth for the experimental datasets, the visual inspection of the optical remote sensing imagery in Google Earth is the only method to assess whether the proposed method has generated a more appropriate G Γ MM. To this end, we make use of the classification map, which is generated according to the Bayesian decision rule, i.e., by maximizing the conditional posterior probability. In other words, the pixels related to gray level r are labeled to one of M^* classes, according to

$$\begin{aligned} C(r) &= \arg \max_{m \in \{1, \dots, M^*\}} p(m|r, \hat{\Theta}) \\ &= \arg \max_{m \in \{1, \dots, M^*\}} \hat{\pi}_m p(r|\hat{\Theta}_m). \end{aligned} \quad (42)$$

The EDSEM only separates water bodies and land roughly for the first three SAR images. The proposed HECM-FJ-G Γ MM can distinguish more scattering components than the EDSEM, besides both correctly identifying two classes for the “India-Port” image. For example, Fig. 6(d) and (j) show the resulting classification maps for EDSEM and HECM-FJ-G Γ MM, respectively. The land meta-class in Fig. 6(d) is further divided into three classes in Fig. 6(f): bare land, vegetation-cover land, and man-made building. Moreover, we illustrate the EDSEM and HECM-FJ-G Γ MM estimated mixtures with the plots of components for the “Toulouse” and “Foulum” images, and their corresponding classification maps in the second and third columns of Fig. 6. The visual comparison validates the better class-discriminative capability of HECM-FJ-G Γ MM. Specifically, as shown in Fig. 6(e), the under-segmentation occurs between the classes trees and buildings. By contrast, the Bayesian classification based on HECM-FJ-G Γ MM PDF estimates gives more accurate and complete classification results, see five classes with different backscattering levels: shadow, road, land, tree, and building (respectively corresponding to the purple, yellow, green, blue, red colors in Fig. 6(k)). Similarly, it is interesting to see from Fig. 6(l) that the land class together with the vegetation class detected by EDSEM in Fig. 6(f) is split into two classes due to the wide range of intensities covered within this class (i.e., high variance). The visual interpretation suggests that two types of crops are identified: rye (purple)

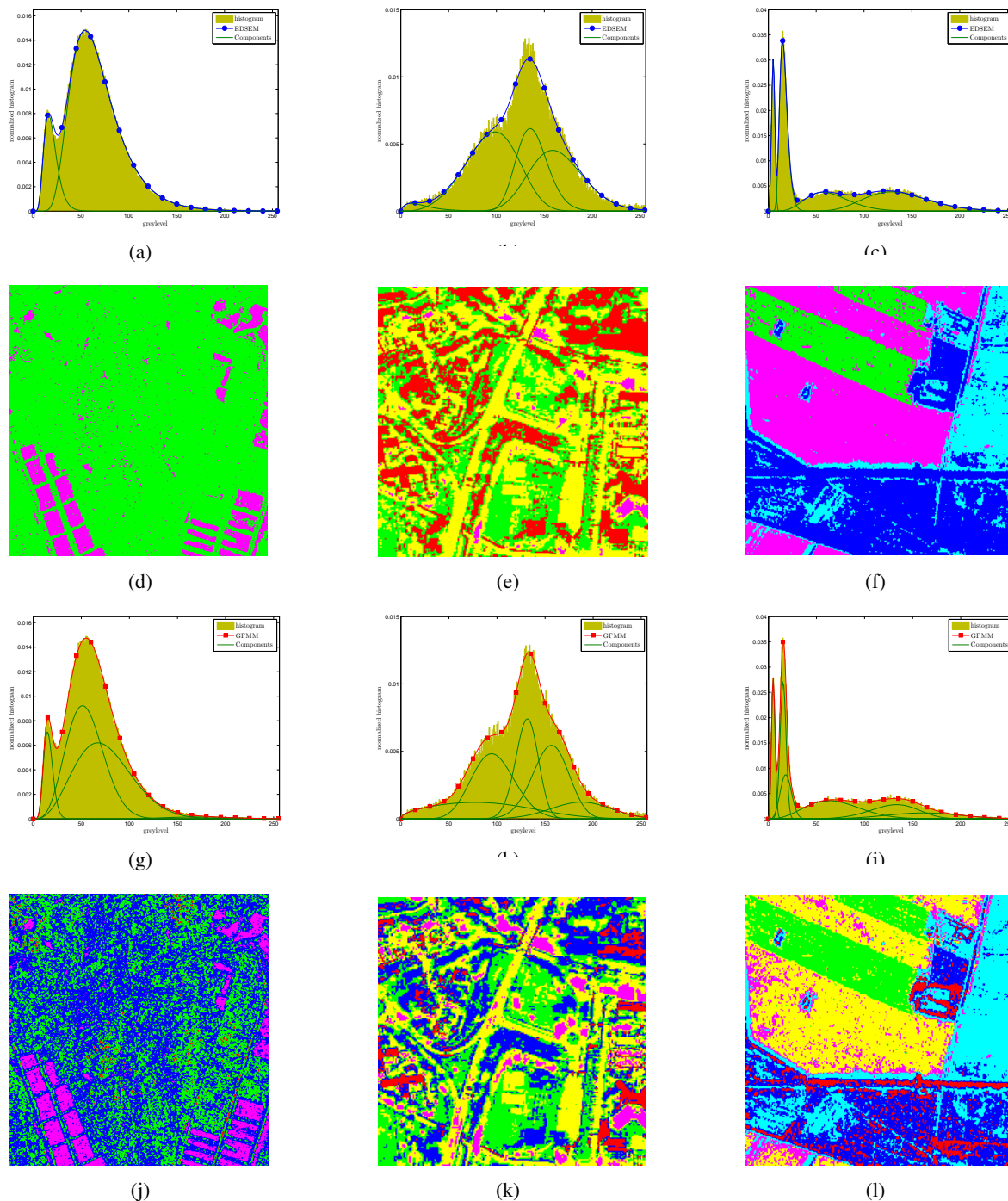


Fig. 6. The estimated mixtures with the plots of components and their classification maps according to Bayesian decision rule: (Left to Right) corresponding to the “Sanchagang-HH” TerraSAR-X image, “Toulouse” RAMSES image, and “Foulum” EMISAR image, respectively; (Top to Bottom) EDSEM estimate with the plots of components, EDSEM classification map, HECM-FJ-GMM estimate with the plots of components, and HECM-FJ-GMM classification map.



Fig. 7. 16-bit “Port-au-Prince” COSMO-SkyMed image (size: 512×512).

TABLE II
QUANTITATIVE MEASURES OBTAINED BY EDSEM AND HECM-FJ-G Γ MM WITH THE OPTIMAL NUMBER OF MIXTURE COMPONENTS FOR TWO 16-BIT AMPLITUDE AND A 8-BIT INTENSITY SAR IMAGES

| Image | EDSEM | | | | HECM-FJ-G Γ MM | | |
|--------------------|-------|---------------------|----------|-----------|-----------------------|---------------|---------------|
| | M^* | Selected Models | D_{KS} | D_{sKL} | M^* | D_{KS} | D_{sKL} |
| “India-Port16” | 2 | $\{f_4, f_1\}$ | 0.0056 | 0.0091 | 2 | 0.0018 | 0.0053 |
| “Port-au-Prince16” | 3 | $\{f_1, f_8, f_4\}$ | 0.0081 | 0.0289 | 5 | 0.0012 | 0.0169 |
| “Ottawa” | – | – | – | – | 4 | 0.0026 | 0.0081 |

with its lower intensity responses and wheat (yellow) with higher intensities, as well as low vegetation (blue) and high vegetation (red). We stress that the classification and modeling results obtained with the proposed method and EDSEM may differ appreciably for some landcover classes due to the different PDF families exploited in the mixture models. Certainly, further classes can be discriminated, if the polarimetric information is available [42].

C. Further Experimental Analysis

The SAR image data are quantized on a discrete scale, for example, 8 bits or 16 bits per pixel. A large-scale quantization yields the wide dynamic range of pixel values. To test the applicability of HECM-FJ-G Γ MM, we conduct further experimental evaluation on two 16-bit SAR images: 1) the original 16-bit counterpart to the “India-Port” image shown in Fig. 2(a); and 2) a subscene (see Fig. 7) extracted from a 16-bit quantized StripMap COSMO-SkyMed amplitude image over Port-au-Prince (Haiti, HH polarization, geocoded ellipsoid corrected, 5m (ground range) \times 5m (azimuth) resolution). For brevity, they will be denoted as “India-Port16” and “Port-au-Prince16”. The same parameter settings are employed as before except $M_{max} = 30$ and $\epsilon = 1 \times 10^{-2}$ for the “India-Port16” image and $M_{max} = 40$ and $\epsilon = 1 \times 10^{-2}$ for the “Port-au-Prince16” image. In this test, a 2-component G Γ MM and a 5-component G Γ MM are obtained, respectively, for the “India-Port16” and “Port-au-Prince16” images. Test results with corresponding measurements are shown in Fig. 8 and Table II. From these two examples, it can be seen that the estimated PDFs match well the observed heavy-tailed histograms, and the obtained quantitative results confirm the accuracy and effectiveness of the proposed method. In addition, we provide the obtained results with the EDSEM model (see Fig. 8 and Table II) for comparison, which correspondingly yields the PDF estimates with 2 and 3 components. The proposed HECM-FJ-G Γ MM outperforms the EDSEM for both indexes.

We proceed by experimentally validating the robustness of the proposed estimator to the radiometric resolution. This is done by comparing the estimates obtained on the above initial 16 bit-per-pixel images and their compressed 8-bit versions. For 8-bit data of the “Port-au-Prince16” image, HECM-FJ-G Γ MM also yields the 5-component G Γ MM, whose plots are shown in Fig. 9. To perform the quantitative comparison, we stretch the PDF estimate $p(x|\Theta)$ of G Γ MM in 8 bits to the original dynamic range by scaling the argument (with a factor inverse to the compression ratio) and renormalizing the PDF. The comparison of the stretched 8-bit estimates and those obtained

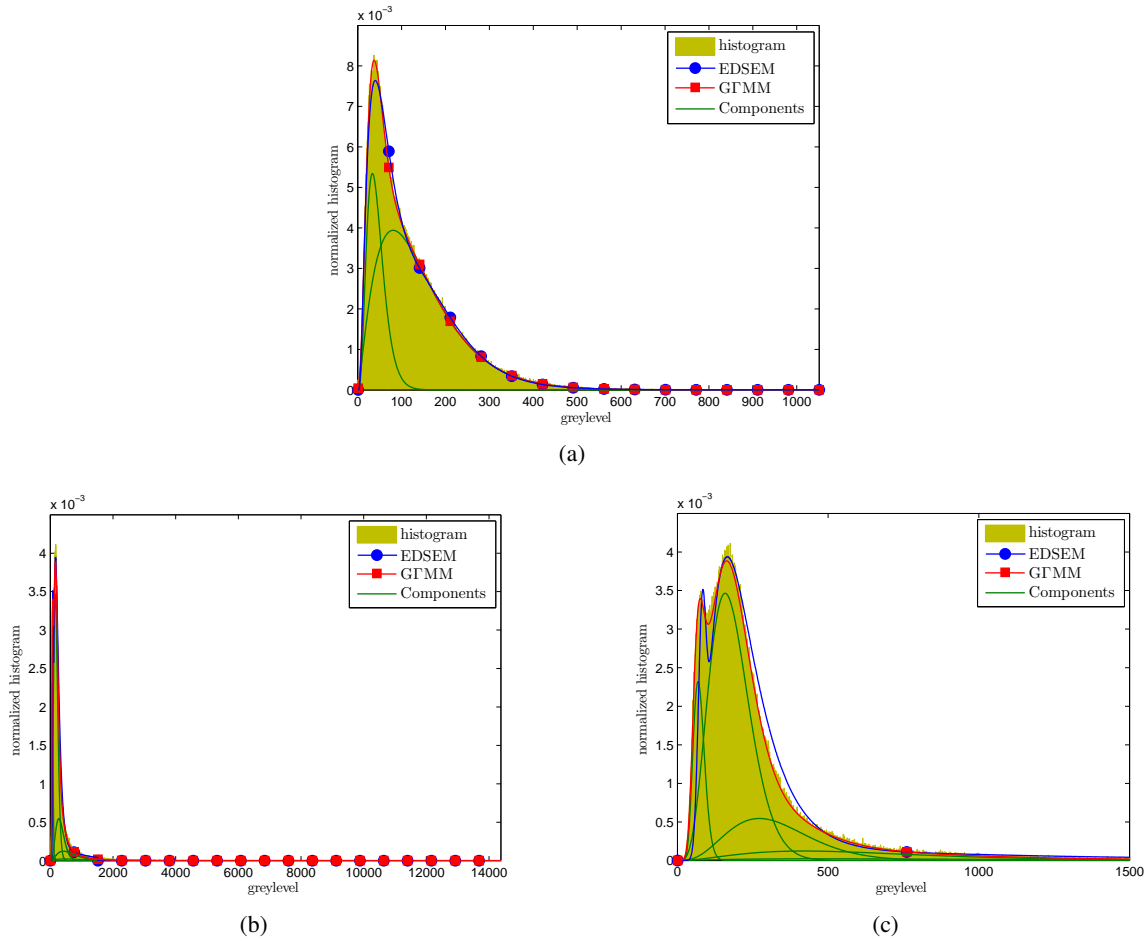


Fig. 8. Histogram and the estimated PDFs of EDSEM and HECM-FJ-G Γ MM with the plots of components: (a) “India-Port16” TerraSAR-X image, (b) “Port-au-Prince16” COSMO-SkyMed image, and (c) Zoomed-in result for (b).

on the initial 16-bit data results in $D_{KS} = 0.0020$, $D_{sKL} = 0.0024$ and $D_{KS} = 0.0136$, $D_{sKL} = 0.0101$, respectively, for the “India-Port16” and “Port-au-Prince16” images. On the one hand, the difference obtained on “India-Port16” is almost negligible, which is consistent with the visual interpretation of Fig. 4(a) and Fig. 8(a). On the other hand, the relatively large D_{KS} and D_{sKL} are observed in the case of “Port-au-Prince16”. This is due to the fact that most pixels cluster in a very narrow range of gray levels such that uniform scaling causes the histogram change from bimodality toward unimodality, see Fig. 9(a). Although the resulting histogram preserves the heavy tails of the initial 16-bit data, the pixels corresponding to the central part of histogram lose a great deal of details, which hinders the follow-up image processing and interpretation. In general, tail truncation is preferable to address this issue before scaling. From Fig. 9(b), we can observe that HECM-FJ-G Γ MM is still effective, and the scaling histogram with truncating threshold³ equal to $T = 5101$ shows the visual bimodal behavior like that of original “Port-au-Prince16” image, exhibiting substantial differences from the uniformly-compressed histogram depicted in Fig. 9(a).

Next the performance of the developed model on the intensity image is investigated. Note that even though the performance of GFD has been experimentally validated only on amplitude data [5], this model is a generalization of the Gamma distribution – a widely used tool for intensity SAR data modeling. As such, application of the designed HECM-FJ-G Γ MM to intensity data is potentially of high interest. In contrast, the EDSEM has to update the underlying dictionary with available PDFs, otherwise the resulting estimates are of little physical interest. For this experiment, we employ another SAR image: a VV polarimetric CONVAIR-580 SAR intensity image over Ottawa (Canada, 8-bit quantized, a 10x azimuth multilooking), denoted as “Ottawa”. In initialization process, we still set equal weights $\pi_m = 1/M_{max}$, but let $\nu_m = 1$ and $\kappa_m = 1$ for $\forall m \in [1, M_{max}]$ (i.e., exponential case)

³The specific value of this threshold is chosen in order to preserve the 99.98% of data.

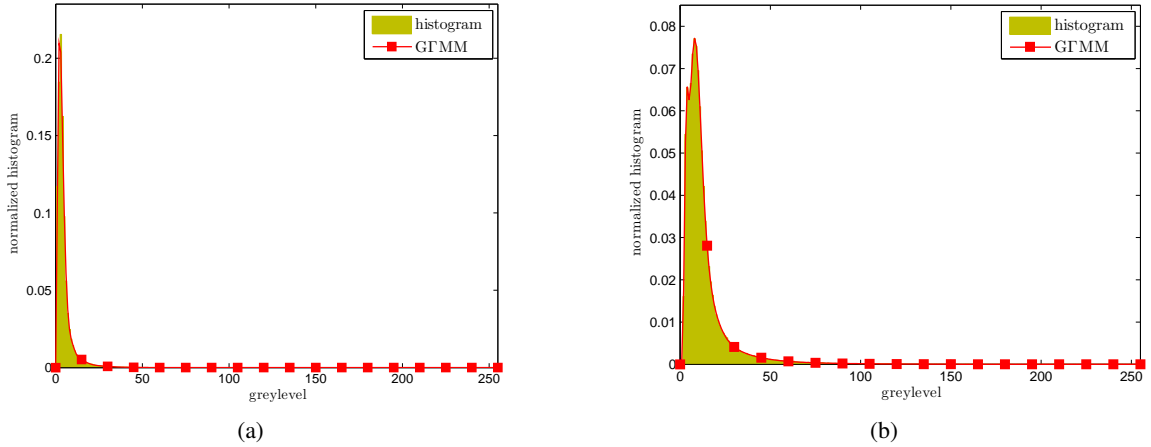


Fig. 9. Histogram and the estimated PDFs of HECM-FJ-GMMM on the “Port-au-Prince16” COSMO-SkyMed image: (a) 8-bit uniformly compressed version, and (b) 8-bit compressed statistics after truncation.

and uniformly choose M_{max} data points from available gray levels of a given image as the mean of exponential components to determine σ_m 's. Initial values for M_{max} and ϵ are selected to be $M_{max} = 30$ and $\epsilon = 1 \times 10^{-2}$. Fig. 10 shows the “Ottawa” image and its corresponding PDF estimate of HECM-FJ-GMMM with the plots of components. It allows achieving the 0.0026 KS distance and the 0.0081 sKL divergence. As expected, the proposed method demonstrates good results in fitting the statistics of intensity SAR image.

Finally, we consider the “India-Port” image to analyze the effect of downsampling. In the following experiments, the downsampling in 2×2 and 4×4 blocks by averaging is employed to create subsampled images. Note that since the pixel spacing is roughly half the value of the spatial resolution for this dataset, the original dataset is oversampled and the 2×2 averaging yields the data at its actual physical resolution. With the same initial conditions as presented in Section B (i.e., $M_{max} = 20$, $\epsilon = 3 \times 10^2$, and $\pi_m = 1/M_{max}$), the proposed HECM-FJ-GMMM algorithm has been applied to these two downsampled images. The resulting PDF estimates both are 2-component GM, which, together with the corresponding normalized histograms, are illustrated in Fig. 11. The quantitative results are as follows: $D_{KS} = 0.0043$, $D_{sKL} = 0.0074$, and $D_{KS} = 0.0040$, $D_{sKL} = 0.0131$, after 2×2 and 4×4 downsampling, respectively. In both cases, HECM-FJ-GMMM performs well in fitting the histograms of the downsampled images. It is well known that downsampling is equivalent to spatial multilooking. Higher values of the equivalent number of looks (ENL) result in the lower the presence of speckle noise in the images. With the increase of ENL, two regions become more distinguishable such that the histogram exhibits obvious bimodal structure, which can be readily observed in Fig. 11. The “Toulouse” image is an example of high ENL in test dataset, for which we further do the experiment to compare with standard Gaussian mixture model (GMM). To this end, 5-component GMM (denoted as 5GMM) fitting estimate is presented in Fig. 12(a), reporting $D_{KS} = 0.0078$ and $D_{sKL} = 0.0169$ from the data. The quantitative comparison with the values in Table I for GM suggests that 5GMM yields relatively worse result, and therefore less efficient for this dataset. In addition, from Fig. 12(a), we can see that a visually unsatisfactory accuracy of fit is presented by 5GMM. On the contrary, due to the flexibility of GFD, GM can improve the estimation performance, especially in the peak and the left-hand tail of histogram, see Fig. 12(b).

VI. CONCLUSION

This paper has explored the statistical analysis of high-resolution SAR images from the semi-parametric perspective. To accomplish this goal, we have introduced a generalized Gamma mixture model (GM) due to the good properties of the underlying GFD, such as its very flexible and compact form. The identifiability of this kind of mixtures has been demonstrated to ensure that the statistical estimation problem is well-posed. Further, the unsupervised learning algorithm for GM, namely HECM-FJ-GM, has been developed by integrating the proposed histogram-based ECM algorithm into the Figueiredo-Jain algorithm, that allows us to simultaneously estimate the parameters of the model and determine the optimal number of mixture components at a relatively low computational cost. The experimental results, obtained from the considered wide range of space/airborne SAR

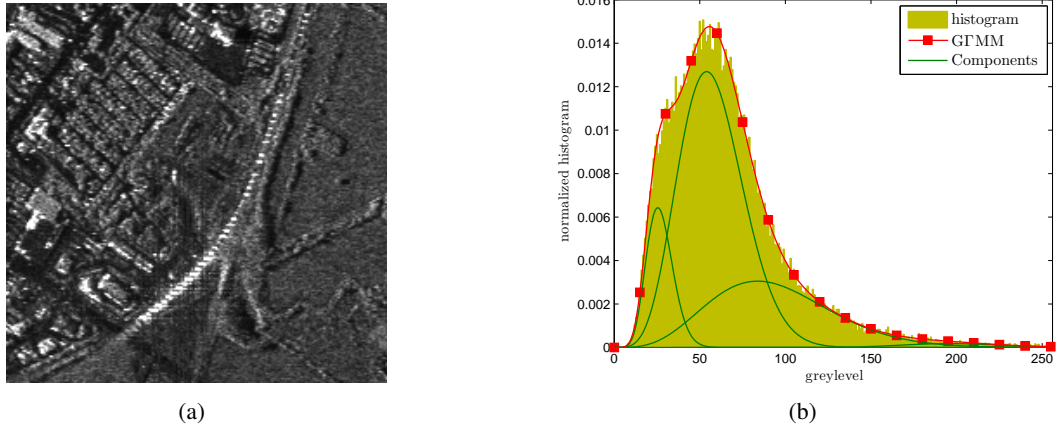


Fig. 10. (a) “Ottawa” CONVAIR-580 SAR image (size: 220×220). (b) Histogram and the estimated PDF of HECM-FJ-GM with the plots of components.

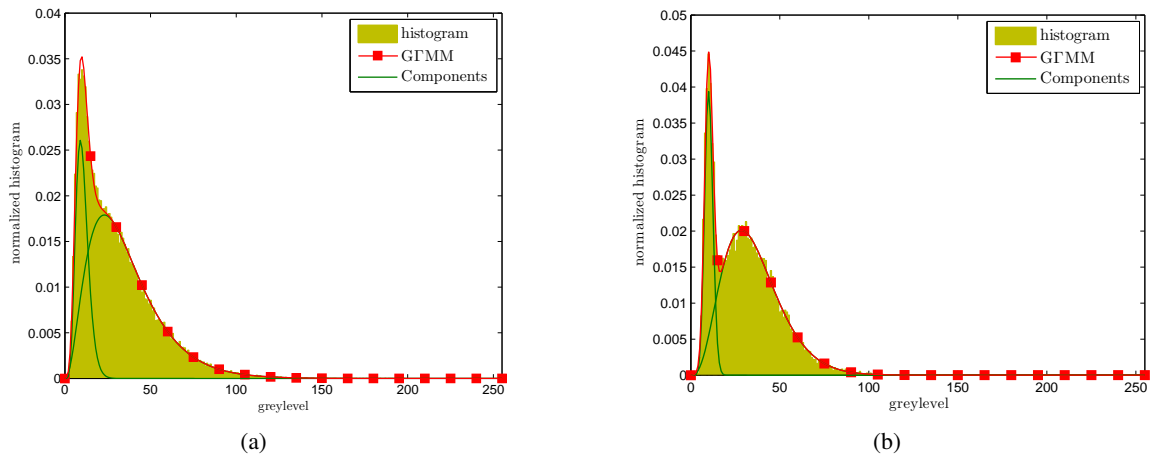


Fig. 11. Histogram and the estimated PDF of HECM-FJ-GM with the plots of components: (a) 2×2 downsampled image of “India-Port”, and (b) 4×4 downsampled image of “India-Port”.

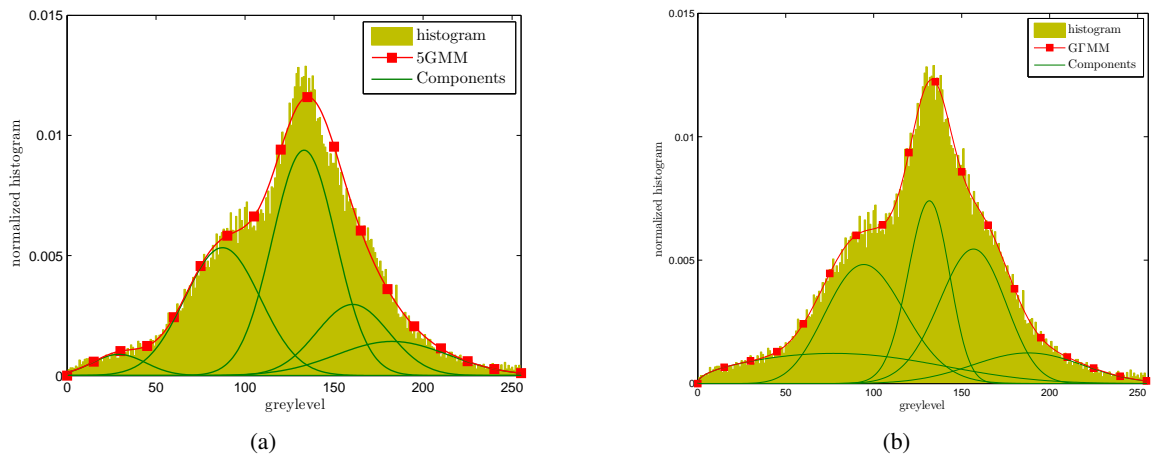


Fig. 12. Histogram and the estimated PDF of 5GMM and GM with the plots of components for the “Toulouse” image: (a) 5GMM, and (b) GM.

images with various scenes, prove the proposed GFMM model and the designed estimator to be efficient and flexible in modeling the challenging bi/multimodal, heavy tailed SAR image data. Future work will be devoted to investigating the unsupervised contextual segmentation of high-resolution SAR images by combining HECM-FJ-GFMM and spatial context within a Markov random field (MRF) framework.

ACKNOWLEDGMENT

The authors would like to thank European Space Agency (ESA) and Infoterra GmbH for providing the “Foulum”, “Ottawa”, and “India-Port”, “Sanchagang” SAR images, which are available at <http://earth.esa.int/polsarpro/> and <http://www.infoterra.de/>, respectively, as well as Italian Space Agency (ASI) for providing the “Livorno” and “Port-au-Prince16” COSMO-SkyMed images. The experiments on the “Toulouse” RAMSES image, given by the French space agency (CNES), have been performed at INRIA by the second and fourth authors. The authors would also like to thank the Editor-in-Chief, the anonymous Associate Editor, and the reviewers for their insightful comments and suggestions, which have greatly improved the paper.

REFERENCES

- [1] C. J. Oliver and S. Quegan, *Understanding Synthetic Aperture Images*. Norwood, MA: Artech House, 1998.
- [2] H. C. Li, W. Hong, Y. R. Wu and P. Z. Fan, “An efficient and flexible statistical model based on generalized Gamma distribution for amplitude SAR images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2711-2722, Jun. 2010.
- [3] R. O. Duta, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, USA: John Wiley & Sons, 2001.
- [4] C. Tison, J. M. Nicolas, F. Tupin, and H. Maitre, “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.
- [5] H. C. Li, W. Hong, Y. R. Wu, and P. Z. Fan “On the empirical-statistical modeling of SAR images with generalized Gamma distribution,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 386-397, Jun. 2011.
- [6] J. W. Goodman, “Statistical properties of laser speckle patterns,” in *Laser Speckle and Related Phenomena*. Heidelberg, Germany: Springer-Verlag, 1975, pp. 9-75.
- [7] E. E. Kuruoglu and J. Zerubia, “Modeling SAR images with a generalization of the Rayleigh distribution,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 527-533, Apr. 2004.
- [8] G. Moser, J. Zerubia, and S. B. Serpico, “SAR amplitude probability density function estimation based on a generalized Gaussian model,” *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1429-1442, Jun. 2006.
- [9] C. J. Oliver, “A model for non-Rayleigh scattering statistics,” *Opt. Acta*, vol. 31, no. 6, pp. 701-722, Jun. 1984.
- [10] 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.
- [11] G. J. McLachlan and D. Peel, *Finite Mixture Models*. New York: Wiley, 2000.
- [12] G. Moser, J. Zerubia, and S. B. Serpico, “Dictionary-based stochastic expectation-maximization for SAR amplitude probability density function estimation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 1, pp. 118-200, Jan. 2006.
- [13] V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, “Enhanced dictionary-based SAR amplitude distribution estimation and its validation with very high-resolution data,” *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 148-152, Jan. 2011.
- [14] J. M. Nicolas, “Introduction to second kind statistic: application of log-moments and log-cumulants to SAR image law analysis,” *Trait. Signal*, vol. 19, no. 3, pp. 139-167, Mar. 2002.
- [15] G. Celeux and J. Diebolt, “The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem,” *Comput. Stat. Quart.*, vol. 2, no. 1, pp. 73-82, Jan. 1985.
- [16] 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.
- [17] E. W. Stacy, “A generalization of the Gamma distribution,” *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1187-1192, Mar. 1962.
- [18] H. Teicher, “Identifiability of finite mixtures,” *Ann. Math. Statist.*, vol. 34, no. 4, pp. 1265-1269, Dec. 1963.
- [19] N. Atienza, J. G. Heras, and J. M. M. Pichardo, “A new condition for identifiability of finite mixture distributions,” *Metrika*, vol. 63, no. 2, pp. 215-221, Apr. 2006.
- [20] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via EM algorithm,” *J. Royal Statist. Soc., Ser. B*, vol. 39, no. 1, pp. 1-38, 1977.
- [21] J. Bilmes, “A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models,” University of Berkely, Tech. Rep. ICSI-TR-97-021, 1997.
- [22] M. Abramowitz and L. A. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1972.
- [23] X. L. Meng and D. Rubin, “Maximum likelihood estimation via the ECM algorithm: a general framework,” *Biometrika*, vol. 80, no. 2, pp. 267-278, 1993.
- [24] X. L. Meng, “On the rate of convergence of the ECM algorithm,” *Ann. Statist.*, vol. 22, no. 1, pp. 326-339, 1994.
- [25] T. P. Minka, “Beyond Newton’s method,” [Online]. Available: <http://research.microsoft.com/minka/papers/newton.html>, 2000.
- [26] W. R. Press, S. A. Teukolsky, W. T. Wetterling, and B. P. Flannery, *Numerical Recipes in C*. London, U.K.: Cambridge Univ. Press, 2002.
- [27] C. F. J. Wu, “On the convergence properties of the EM algorithm,” *Ann. Statist.*, vol. 11, no. 1, pp. 95-103, 1983.
- [28] G. Schwarz, “Estimating the dimension of a model,” *Ann. Statist.*, vol. 6, no. 2, pp. 461-464, 1978.

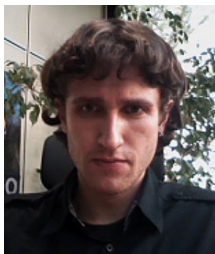
- [29] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control.*, vol. AC-19, no. 6, pp. 716-723, Dec. 1974.
- [30] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465-471, Sept. 1978.
- [31] C. S. Wallace and D. M. Boulton, "An information measure for classification," *Comput. J.*, vol. 11, no. 2, pp. 185-194, Aug. 1968.
- [32] R. A. Baxter and J. J. Oliver, "Finding overlapping components with MML," *Statistics and Computing*, vol. 10, no. 1, pp. 5-16, 2000.
- [33] C. Biernacki, G. Celeux, and G. Govaert, "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.
- [34] D. Titterton, A. Smith, and U. Makov, *Statistical Analysis of Finite Mixture Distributions*. Chichester, U.K.: John Wiley & Sons, 1985.
- [35] J. Bernardo and A. Smith, *Bayesian Theory*. Chichester, U.K.: John Wiley & Sons, 1994.
- [36] A. R. Barron, C. Huang, J. Q. Li, and X. Luo, "MDL, penalized likelihood, and statistical risk," In Proc. ITW'08, Porto, 2008, pp. 247-257.
- [37] G. Celeux, S. Chretien, F. Forbes, and A. Mkhadri, "A component-wise EM algorithm for mixtures," *J. Comput. Graphical Statist.*, vol. 10, no. 4, pp. 699-712, Dec. 2001.
- [38] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.
- [39] J. M. Nicolas and F. Tupin, "Gamma mixture modeled with second kind statistics: application to SAR image processing," in Proc. IGARSS'02, Toronto, Canada, 2002, pp. 2489-2491.
- [40] V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, "On the method of logarithmic cumulants for parametric probability density function estimation," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3791-3806, Oct. 2013.
- [41] K. S. Song, "Globally convergent algorithms for estimating generalized Gamma distributions in fast signal and image processing," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1233-1250, Aug. 2008.
- [42] V. Akbari, A. P. Doulgeris, G. Moser, T. Eltoft, S. N. Anfinsen, and S. B. Serpico, "A textural-contextual model for unsupervised segmentation of multipolarization synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2442-2453, Apr. 2013.



Heng-Chao Li (S'06–M'08–SM'14) received the B.Sc. and M.Sc. degrees from the Southwest Jiaotong University, Chengdu, China, in 2001 and 2004, respectively, and the Ph.D. degree from the Graduate University of the Chinese Academy of Sciences, Beijing, China, in 2008, all in information and communication engineering.

He is currently a Professor with Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University, Chengdu, China. Since November 2013, he is a Visiting Scholar working with Prof. William J. Emery at the University of Colorado Boulder, Boulder, CO, USA. His research interests include statistical analysis and processing of synthetic aperture radar (SAR) images, and signal processing in communications.

Dr. Li received several scholarships or awards, especially including the Special Grade of the Financial Support from China Postdoctoral Science Foundation in 2009 and the New Century Excellent Talents in University from the Ministry of Education of China in 2011. In addition, he also has been a Reviewer for several international journals and conferences, such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, *IET Radar, Sonar and Navigation*, and *Canadian Journal of Remote Sensing*.



Vladimir A. Krylov received the "specialist" (M.Sc.) degree in applied mathematics and computer science, and the "candidate of physico-mathematical sciences" (Ph.D.) degree in statistics both from the Lomonosov Moscow State University, Moscow, Russia, in 2007 and 2011, respectively.

In 2011–2012 he was a Postdoctoral Fellow with Ariana and Ayin research teams at INRIA, Sophia Antipolis, France; in 2012–2013 he was a research associate with the Department of Statistical Science at the University College London, London, UK. Since 2014 he is with the Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture (DITEN) at the University of Genoa, Italy.

His research interests are in the field of statistical signal processing and pattern recognition applied to medical and remote sensing imagery.



Ping-Zhi Fan (M'93–SM'99–F'15) received his PhD degree in Electronic Engineering from the Hull University, UK.

He is currently a Professor and Director of the Institute of Mobile Communications, Southwest Jiaotong University, China. He is a recipient of the UK ORS Award, the Outstanding Young Scientist Award by NSFC, and the Chief Scientist of a National 973 research project. He served as General Chair or TPC Chair of a number of international conferences, and is the Guest Editor-in-Chief, Guest Editor or Editorial member of several international journals. He is the Founding Chair of IEEE VTS BJ Chapter and ComSoc CD Chapter, the Founding Chair of IEEE Chengdu Section. He also served as a board member of IEEE Region 10, IET(IEE) Council and IET Asia-Pacific Region. He has over 200 research papers published in various academic English journals (IEEE/IEE/IEICE, etc), and 8 books (incl. edited), and is the inventor of 20 granted patents. His research interests include high mobility wireless communications, 5G technologies, wireless networks for big data, signal design & coding, etc. He is an IEEE VTS Distinguished Lecturer (2015–2017), and a fellow of IEEE, IET, CIE and CIC.



Josiane 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 has been a permanent research scientist at INRIA since 1989, and director of research since July 1995. She was head of the PASTIS remote sensing laboratory (INRIA Sophia-Antipolis) from mid-1995 to 1997 and of the Ariana research group (INRIA/CNRS/University of Nice), which worked on inverse problems in remote sensing, from 1998 to 2011. Since January 2012, she has been head of Ayin research group (INRIA-SAM) dedicated to hierarchical and stochastic models for remote sensing and skincare imaging. She has been Professor at SUPAERO (ISAE) in Toulouse since 1999. Before that, she was with the Signal and Image Processing Institute of the University of Southern California (USC) in Los-Angeles as a postdoc. She also worked as a researcher for the LASSY (University of Nice/CNRS) from 1984 to 1988 and in the Research Laboratory of Hewlett Packard in France and in Palo-Alto (CA) from 1982 to 1984.

Dr. Zerubia was a member of the IEEE IDMSPT TC (SP Society) from 1997 till 2003, of the IEEE BISP TC (SP Society) from 2004 till 2012 and of the IVMSPT TC (SP Society) from 2008 till 2013. She was Associate Editor of IEEE Transactions on Image Processing from 1998 to 2002; Area Editor of Transactions on Image Processing from 2003 to 2006; Guest Co-editor of a special issue of IEEE Transactions on Pattern Analysis and Machine Intelligence in 2003; and member-at-large of the Board of Governors of the IEEE SP Society from 2002 to 2004. She has also been a member of the editorial board of the French Society for Photogrammetry and Remote Sensing (SFPT) since 1998, of the International Journal of Computer Vision since 2004, and of the Foundation and Trends in Signal Processing since 2007. She has been Associate Editor of the on-line resource Earthzine (IEEE CEO and GEOSS). 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); Technical Program Chair of a workshop on Photogrammetry and Remote Sensing for Urban Areas, Marne La Vallée, France, 2003; Co-chair of the special sessions at IEEE ICASSP 2006 (Toulouse, France) and IEEE ISBI 2008 (Paris, France); and Publicity Chair of IEEE ICIP 2011 (Brussels, Belgium), Tutorial Co-chair of IEEE ICIP 2014 (Paris, France), General Co-chair of the workshop Earthvision at IEEE CVPR 2015 (Boston, USA) and a member of the organizing committee and plenary talk co-chair of IEEE-EURASIP EUSIPCO 2015 (Nice, France).

Her main research interest is in image processing using probabilistic models. She also works on parameter estimation, statistical learning and optimization techniques.



Willaim J. Emery (M'90–SM'01–F'03) received the Ph.D. degree in physical oceanography from the University of Hawaii, Honolulu, HI, USA, in 1975.

After working at Texas A&M University, College Station, TX, USA, he moved to the University of British Columbia, Vancouver, BC, Canada, in 1978, where he created a Satellite Oceanography facility/education/research program. He was appointed Professor of Aerospace Engineering Sciences at the University of Colorado, Boulder, CO, USA, in 1987. He is an Adjunct Professor of informatics at Tor Vergata University, Rome, Italy. He has authored over 182 refereed publications on both ocean and land remote sensing and two textbooks.

Prof. Emery is a Fellow of the American Meteorological Society (2010), the American Astronautical Society (2011), and the American Geophysical Union (2012). He is the Vice President for Publications of the IEEE Geoscience and Remote Sensing Society (GRSS) and a member of the IEEE Periodicals Committee. He was the recipient of the GRSS Educational Award in 2004 and its Outstanding Service Award in 2009.