Dictionary-based Stochastic Expectation-Maximization for SAR amplitude probability density function estimation

Gabriele Moser, Josiane Zerubia, Sebastiano B. Serpico

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Dictionary-based Stochastic Expectation-Maximization for SAR amplitude probability density function estimation

Gabriele Moser — Josiane Zerubia — Sebastiano B. Serpico

N° 5154

March 2004

THÈME 3

ISSN 0249-6399 ISRN INRIA/RR--5154--FR+ENG

INRIA
SOPHIA ANTIPOLIS

Dictionary-based Stochastic Expectation-Maximization for SAR amplitude probability density function estimation

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

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Ariana

Rapport de recherche n° 5154 — March 2004 — 49 pages

Abstract: In the context of remotely sensed data analysis, a crucial problem is represented by the need to develop accurate models for the statistics of the pixel intensities. In the current research report, we address the problem of parametric probability density function (PDF) estimation in the context of Synthetic Aperture Radar (SAR) amplitude data analysis. Specifically, several theoretical and heuristic models for the PDFs of SAR data have been proposed in the literature, and have been proved to be effective for different land-cover typologies, thus making the choice of a single optimal SAR parametric PDF a hard task. In thia report, an innovative estimation algorithm is proposed, which addresses this problem by adopting a finite mixture model (FMM) 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 (SEM) iterative methodology and the recently proposed "method-of-log-cumulants" (MoLC) for parametric PDF estimation for non-negative random variables. Experimental results on several real SAR images are presented, showing the proposed method is accurately modelling the statistics of SAR amplitude data.

Key-words: Synthetic Aperture Radar (SAR) images, probability density function estimation, parametric estimation, finite mixture models, Stochastic Expectation Maximization (SEM).

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

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

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

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

Estimation de la densité de probabilité de l'amplitude d'une image radar RSO via un algorithme SEM en utilisant un dictionnaire

Résumé : En télédetection, un problème vital est le besoin de développer des modèles précis pour représenter les statistiques des intensités des images. Dans ce rapport de recherche, nous traitons le problème de l'estimation de la densité de probabilité de l'amplitude d'une image de type Radar à Synthèse d'Ouverture (RSO). Plusieurs modèles théoriques ou heuristiques, ultilisés pour représenter l'amplitude d'un signal du type RSO, ont été proposés dans la littérature et ce sont révelés être efficaces pour différentes types de classes dans le contexte des cartes d'occupation des sols, rendant ainsi difficile le choix d'une seule densité de probabilité paramétrique.

Dans ce rapport de recherche, un algorithme d'estimation innovant est proposé, se fondant sur un modèle de mélange fini pour la densité de probabilité de l'amplitude, les diverses composantes du mélange appartenant à un dictionnaire specifique. La mèthode proposée dans ce rapport intégre, de façon automatique, les procédures de sélection d'un modèle optimal pour chaque composante, d'estimation de paramètres et d'optimisation du nombre de composantes, en combinant un algorithme EM stochastique et la méthode des logs-cumulants pour l'estimation de la densité de probabilité paramètrique. Des resultats expérimentaux sur plusieurs images RSO réelles sont présentés, montrant ainsi que la mèthode proposée est suffisamment précise pour modéliser les statistiques du signal d'amplitude radar de type RSO.

Mots-clés : image Radar à Synthèse d'Ouverture (RSO), densité de probabilité, estimation paramétrique, modèles de mélange fini, EM Stochastique (SEM).

Contents

1 Introduction

In the context of remotely sensed data analysis, a crucial problem is represented by the need to develop accurate models for the statistics of the pixel intensities. Focusing on Synthetic Aperture Radar (SAR) [11][12][45][39] data, this modelling process turns out to be a crucial task, for instance, for classification [17] or for denoising [39] purposes.

From a methodological viewpoint, either parametric or non-parametric estimation strategies can be used for this task [17][18]. Specifically, the 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, such as the Maximum Likelihood (ML) methodology [17] or the "Method of Moments" (MoM) [27][37][43][51]. On the contrary, non-parametric PDF estimation approaches do not assume any specific analytical model for the unknown PDF, thus providing a higher flexibility, although usually presenting internal architecture parameters to be set by the user [17]. In particular, several non-parametric kernel-based estimation and regression architectures have been proposed in the literature, proving to be effective estimation tools, such as standard Parzen window estimators [17][41], artificial neural networks [6], or Support Vector Machines (SVMs) [29][53][54].

In the current research report, we address the problem of PDF estimation in the context of SAR amplitude data analysis. Specifically, several different theoretic and heuristic models for the PDFs of SAR data have been proposed in the literature, and have proved to be effective for specific land-cover typologies. For instance, the Rayleigh [39] distribution has been proposed as a theoretical amplitude PDF, stemming from a Gaussian model for the backscattering phenomena [39] involved in the generation of a single-look SAR image. The Nakagami-Gamma [37][39][51] distribution generalizes the Rayleigh one to multilook data and is a usually accepted model for non-textured image areas [36][39]. The K distribution [39] for SAR intensity data is obtained from a different model for the statistics of the backscattering phenomena [15][21][22] and is known in the literature to correctly describe the statistics of highly textured image areas [36][39]. The log-normal and the Weibull distributions [39] have been introduced as heuristic models of the amplitude or of the intensity statistics, and successfully applied to image of urban areas and of ocean, land, sea-ice, respectively [39]. The Fisher distribution has been adopted as an empirical model for the SAR statistics over high-resolution urban regions [51], and the Pearson system of parametric families has been applied for SAR image segmentation purposes [14]. In [27], symmetric α-stable (SαS) distributions [26][28] are employed to describe the SAR backscattering phenomena and the resulting amplitude PDF is proved to model correctly the amplitude statistics over an urban area.

Hence, different parametric families among the above-mentioned ones turn out to be effective models for different land cover typologies [39], which makes the choice of a single optimal SAR amplitude parametric PDF a hard task. In addition, a remotely sensed image, in general, can depict a varied scene, jointly presenting several distinct land cover typologies. In this report, an innovative SAR amplitude parametric estimation algorithm is proposed,

which addresses these problems by adopting a Finite Mixture Model (FMM) [44][43] for the amplitude PDF, i.e., by postulating the unknown amplitude PDF to be a linear combination of parametric components, each one corresponding to a specific land cover type [15][42].

In addition, in order to take explicitly into account the possible differences in the statistics of the mixture components, we avoid choosing *a priori* a specific parametric family for each component, but we allow each of them to belong to a given dictionary of SAR-specific PDFs, namely, the well-known log-normal, Nakagami-Gamma, Weibull, and K distributions, together with the SαS generalized Rayleigh PDF [27] and with the recently proposed generalized Gaussian Rayleigh (GGR) model [34].

Specifically, the proposed algorithm automatically integrates the procedures of selection of the optimal model for each component and of parameter estimation, by combining the Stochastic Expectation Maximization (SEM) algorithm [5][9][30] and the "method-of-log-cumulants" (MoLC) [37][38]. The former is a stochastic iterative parametric estimation methodology, dealing with problems of data incompleteness and developed as an improvement of the standard Expectation-Maximization (EM) algorithm [16][43], in order to increase its capability to compute Maximum Likelihood (ML) estimates [52]. MoLC [37] is a recently proposed estimation approach stemming from the adoption of the Mellin transform [49] (instead of the usual Fourier transform) in the computation of characteristic functions, and from the corresponding generalization of the concepts of moment and of cumulant [40]. We adopt this method both for its good estimation properties [36][37][51] and because it turns out to be feasible and fast for all the parametric families in the dictionary [34][36]. On the contrary, the well-known Maximum-Likelihood (ML) and "method-of-moments" (MoM) estimation strategies [27][37] involves numerical difficulties for several of these families [34][37][39]. In addition, the developed method automatically performs an optimal choice of the number K of mixture components, by computing, for each values 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 proposed parametric approach is validated using several real ERS-1, XSAR, ESAR, and airborne SAR images. The experimental results show the method to accurately model the amplitude distribution of all the considered images, both from a qualitative viewpoint (i.e., visual comparison between the data histogram and the estimated PDF) and from a quantitative viewpoint (i.e., correlation coefficient between the data histogram and the estimated PDF), thus showing its effectiveness and flexibility.

In Section 2, the proposed FMM-based estimation scheme is presented and the SEM and the MoLC methods are described. Section 3 reports the results of the application of the proposed approach to the statistical modelling of the grey-level of several real SAR images, proving the method to fit the amplitude distribution better than several previously proposed parametric models for SAR amplitude data and also comparing the performances of the method with the ones of a benchmark non-parametric approach. Finally, conclusions are drawn in Section 4.

2 A dictionary-based finite mixture model for SAR data PDF estimation

2.1 Finite mixture models for SAR amplitude data

In order to take explicitly into account the possible presence in a given SAR amplitude [39] image $\mathcal{I}$ of several distinct land-cover typologies, yielding different contributions to the statistics of the pixel intensity, we assume a finite mixture model (FMM) [43][44] for the grey level PDF. Specifically, we model the SAR image as a set $\mathcal{I} = \{r_1, r_2, \ldots, r_N\}$ of independent and identically distributed (i.i.d.) samples drawn according to the following probability density function (PDF)[1]:

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

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

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

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

$$\theta = \mathrm{col}[\theta_i, P_i : i = 1, 2, \ldots, K]. \tag{3}$$

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

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

the problem of FMM parametric estimation, i.e., the computation of a parameter vector $\theta^* \in \Theta$ optimally representing the observed image data $\mathcal{I}$, has been addressed according to several different approaches [43][44]. Specifically, the i.i.d. assumption allows obtaining the following expression for the log-likelihood function [52] of the image data $\mathcal{I}$ [43][44]:

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

[1] This approach is widely accepted in the context of estimation theory [17][18][52] and operatively corresponds to discard in the estimation process the contextual information associated to the correlation between neighboring pixels in the image, thus exploiting only the greylevel information.

The computation of Maximum Likelihood (ML) estimates would involve the maximization of this function, but the solution of this maximization problem is not feasible analytically [44] and also involves several numerical difficulties, due, for instance, to the usual presence of several local maxima [43]. In order to circumvent this issues in the computation of ML estimates, the use of the Expectation-Maximization (EM) algorithm has been proposed [16][32][43], which formalizes the problem of the estimation of the parameters of a FMM as an incomplete data problem and introduces a sequence $\{\theta^t\}_{t=0}^{\infty}$ of parameter estimates, iteratively maximizing a Pseudo-Likelihood function, i.e. [16][43]:

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

where

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

In the context of mixture densities EM has been proved to converge to a stationary point of the log-likelihood function $L_{\mathcal{I}}(\cdot)$ [16][43][56], although not converging, in general, to a global maximum point, and exhibiting sometimes long convergence times [9][44]. In addition, the maximization problem reported in Eq. (6), although analytically tractable in several applications [32] (e.g.: mixtures of exponential families [43]), does not always yield a closed form solution, and moreover the convergence point can belong to the boundary of the parameter space Θ, thus possibly involving analytical singularities [44]. Several variants of EM have been introduced in order to address these difficulties. In [20], [25], and in [50] modified versions of EM, as well as regularized covariance estimators, are proposed in order to optimize the robustness of the estimation process in the context of Gaussian Mixtures Models (GMMs) for hyperspectral data classification. A simplified version of EM, named Classification EM (CEM) [5], has been developed, which converges in a finite number of iterations, but yields, in general, biased parameter estimates. A sequential version of EM, namely, the Component-wise EM (CEM^2), is proposed in [10], which aims at a reduction in the computation time and also at avoiding analytical singularities [44]. The Stochastic EM (SEM) [9] avoids the computation of the Pseudo-Likelihood function $Q_{\mathcal{I}}(\cdot|\cdot)$ and the related analytical maximization issues, by integrating a stochastic sampling procedure in the estimation process. Hence, the sequence of parameter estimates generated by SEM is a discrete time random process, which does not converge pointwise, but has been proved to be an ergodic and homogeneous Markov chain, converging to a unique stationary distribution, which is expected to be concentrated around the global maxima of the log-likelihood function [9]. Simulated Annealing EM (SAEM) [9] is a combination of EM and SEM allowing SEM to converge also almost surely (a.s.) [40][55], although at a local maximum and requiring the preliminary definition of a suitable annealing schedule. Monte Carlo EM (MCEM) and Simulated Annealing Monte Carlo EM (SA-MCEM) are further stochastic variants of EM substituting a Monte Carlo sampling procedure to the computation of the $Q_{\mathcal{I}}(\cdot|\cdot)$ function and a.s. converging, under mild assumptions, to a local maximum [9]. In [14] the Iterated

Conditional Expectation (ICE) method is used, as an alternative to EM and SEM, to the FMM estimation for image segmentation and clustering purposes and in [44] the Minimum Message Length (MML) principle is applied in order to formulate a FMM unsupervised parametric estimator, aiming at avoiding convergence to the boundary of the parameter space, as well as to improve, with respect to EM, robustness with respect to initialization, but still involving the maximization problem (see Eq. 6).

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

2.2 SEM for FMM parametric density estimation

A general parametric density estimation problem assumes the availability of an observation random vector $x \in X \subset \mathbb{R}^n$, whose density function $p_x(\cdot|\cdot) : X \times \Theta \to [0, +\infty)$ depends on a parameter vector taking values in a given set $\Theta \subset \mathbb{R}^\ell$ (namely, the parameter space) [2].

This general problem is said to present data incompleteness when the data vector x cannot be directly observed, for instance, due to presence of lacking or corrupted data [32]. Such incompleteness issues is formalized by assuming the "complete" data vector x not to be available, but to be observed only through an "incomplete" data vector $y = \Phi(x)$, obtained through a many-to-one mapping $\Phi : X \to Y \subset \mathbb{R}^m$ [16]. Hence, a given realization $y \in Y$ of the incomplete data may have been generated by any realization $x \in \Phi^{-1}(y) \subset X$ in the inverse image $\Phi^{-1}(y)$ of y, thus not allowing, for instance, a direct feasible computation of an ML estimate. SEM tries to avoid these difficulties by iteratively random sampling a complete data set and by using it to compute a ML standard estimate.

Specifically, we denote as $p_y(\cdot|\cdot) : Y \times \Theta \to [0, +\infty)$ the parametric incomplete data density function[3] and as $p_{x|y}(\cdot|y, \cdot) : \Phi^{-1}(y) \times \Theta \to [0, +\infty)$ the complete data parametric density conditioned to an incomplete data realization $y \in Y$, and we state that [16][9][33]:

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

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

[2] We explicitly avoid here to focus on the specific cases of a probability density function (PDF) for a continuous random variable or of a probability mass function (PMF) for a discrete random variable and we generically refer to a "density function". More technically, x is a measurable mapping between the underlying probability space $(S, \mathcal{F}, P)$ and a given σ-finite measure space $(X, \mathcal{X}, \xi)$, such that the image measure $D \in \mathcal{X} \longmapsto P\{x \in D\}$ is absolutely continuous with respect to ξ [47]. Hence, the Radon-Nicodym theorem [47] assures the existence of an $\mathcal{X}$-measurable density function on X for the random vector x.

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

where ζ_y is a suitable measure[4] on $\Phi^{-1}(y)$.

Given an observed incomplete data realization $y \in Y$, SEM computes a random sequence $\{\theta^t\}_{t=0}^{\infty}$, by performing at the t-th iteration ($t = 0, 1, \ldots$) the following three processing steps [9]:

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

As previously stated, the resulting discrete time random process $\{\theta^t\}_{t=0}^{\infty}$ is not pointwise nor a.s. convergent, but has been proved to be an homogeneous Markov chain. If this sequence turns out to be also ergodic, it converges in distribution to the unique stationary distribution of the Markov chain and is expected to concentrate around the global maxima of the likelihood function [9].

The FMM case satisfies this ergodicity assumption [9], thus suggesting SEM as a promising estimation tool. The specific FMM framework can be viewed as affected by data incompleteness problems, since it is not known from which of the K available statistical populations involved in Eq. (1) a given image sample is drawn. This implicitly means that no training information about the possible thematic classes in the SAR image is exploited in the estimation process, i.e., the SAR amplitude PDF estimation problem is addressed in an unsupervised context. Thus, denoting as $\Sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_K\}$ the set of the K different populations, we assume the population label $s_k \in \Sigma$ of the k-th image pixel not to be known, thus suggesting the following definition of the complete and of the incomplete data vectors, respectively:

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

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

[4]More specifically, a σ-finite measure space $(\Phi^{-1}(y), \mathcal{Z}_y, \zeta_y)$ is introduced on the inverse image $\Phi^{-1}(y)$ of $y \in Y$, such that the restriction of $p_x(\cdot|\theta)$ to $\Phi^{-1}(y)$ is $\mathcal{Z}_y$-measurable for all $\theta \in \Theta$. This allows introducing the model described by Eq. (8) for the incomplete data density, which yields the expression (9) for the conditional complete data one [16][33][43].

density[5] of (r_k, s_k), respectively $(k = 1, 2, \ldots, N)$. Hence [33][43]:

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

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

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

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

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

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

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

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

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

- **M-step:** update the parameter estimates, by computing, according to the partition generated by the S-step, a standard supervised ML estimate $\theta^{t+1} \in \Theta$, i.e.[6]:

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

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

[5] The couple (r_k, s_k) of random variables cannot be adequately described by a PDF or by a PMF, since r_k is a continuous random variable, whereas s_k is a discrete one. Specifically, denoting as $\mathcal{B}$ the collection of the Borel subsets of $[0, +\infty)$ [46], as $\mathcal{S}$ the power set of Σ (i.e., the collection of all the subsets of Σ), as m the Lebesgue measure on $\mathbb{R}$ [46], and as c the counting measure on Σ [43], the PDFs p_r and $p_{r|s}$ are defined with respect to the measure space $([0, +\infty), \mathcal{B}, m)$, the PMFs P_s and $P_{s|r}$ with respect to $(\Sigma, \mathcal{S}, c)$, and the joint density p_{rs} has to be referred to the product space $([0, +\infty) \times \Sigma, \mathcal{B} \otimes \mathcal{S}, m \otimes c)$ [47]. In particular, p_{rs} can be proved to satisfy Eq. (12) [33].

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

A feasible initialization procedure for the described iterative process consists in setting initially a uniform posterior distribution for all image pixels, and by using this distribution in the random sampling process performed by the S-step. Furthermore, since SEM is not pointwise or a.s. convergent, a specific termination procedure has to be defined in order to extract from the stationary steady distribution of the random sequence $\{\theta^t\}_{t=0}^{\infty}$ a single parameter estimate θ^*. Several strategies have been proposed to this purpose, for instance, in [5] and in [9].

2.3 The "dictionary" approach to SAR amplitude PDF estimation

The application of the FMM model described by Eq. (1) requires the definition of suitable parametric models for the mixture components. However, as described in Section 1, in the present SAR-specific context several parametric models have been proposed and proved to be effective descriptions of the statistics of the pixel intensities corresponding to different land cover typologies. In order to overcome the intrinsic difficulty of an *a priori* choice of a single suitable model and to improve the flexibility of the adopted FMM approach, we avoid selecting a specific parametric family for each component $p_i(\cdot|\cdot)$ $(i = 1, 2, \ldots, K)$, and we adopt instead a finite dictionary $\mathcal{D} = \{f_1, f_2, \ldots, f_M\}$ of M SAR-specific distinct parametric PDFs $f_j(\cdot|\cdot) : [0, +\infty) \times \Xi_j \rightarrow [0, +\infty)$ $(j = 1, 2, \ldots, M)$ with parameter spaces Ξ_j $(j = 1, 2, \ldots, M)$. In [14] a similar approach was proposed, which applied EM, SEM, and ICE to a "generalized mixture model" with mixture components not restricted to belong to a specific parametric family, but selected inside the Pearson system of distributions. Here, dealing with SAR amplitude data PDF estimation, we adopt a dictionary $\mathcal{D}$ consisting of the following six parametric PDFs:

- the empirical log-normal distribution [39]:

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

- the Nakagami distribution, proposed as an amplitude model for multi-look SAR data [37][51]:

$$f_2(r|L, \mu) = \frac{2}{\Gamma(L)} \left(\frac{L}{\mu}\right)^L r^{2L-1} \exp\left(-\frac{Lr^2}{\mu}\right), \quad r \geq 0, \tag{18}$$

  where $\Gamma(\cdot)$ is the Gamma function [46];

- the generalized Gaussian Rayleigh (GGR) distribution, based on a generalized Gaussian model for the backscattered SAR signal [34]:

$$f_3(r|\lambda, \gamma) = \frac{\gamma^2 r}{\lambda^2 \Gamma^2(\lambda)} \int_0^{\pi/2} \exp[-(\gamma r)^{1/\lambda}(|\cos\theta|^{1/\lambda} + |\sin\theta|^{1/\lambda})] d\theta, \quad r \geq 0; \tag{19}$$

- the SαS generalized Rayleigh distribution (hereafter simply denoted as SαSGR), based on a SαS model for the SAR backscattered signal [27]:

$$f_4(r|\alpha,\gamma) = r\int_0^{+\infty} \rho \exp(-\gamma\rho^\alpha) J_0(r\rho) d\rho, \quad r \geq 0, \tag{20}$$

  where $J_0(\cdot)$ is the zero-th order Bessel function of the first kind [49];

- the empirical Weibull distribution [36][39]:

$$f_5(r|\eta,\mu) = \frac{\eta}{\mu^\eta} r^{\eta-1} \exp\left[-\left(\frac{r}{\mu}\right)^\eta\right], \quad r \geq 0; \tag{21}$$

- the amplitude distribution corresponding to a K-distributed intensity [39] (hereafter denoted simply as "K-root"):

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

  where $K_\nu(\cdot)$ ($\nu > 0$) is the ν-th order modified Bessel function of the second kind [49].

Hence, with this specific choice, $M = 6$ distinct parametric families are involved in the estimation process with 2-parameter families (i.e., $\Xi_j = (0,+\infty)^2$) for $j \neq 6$ and a 3-parameter family for $j = 6$ (i.e., $\Xi_6 = (0,+\infty)^3$). We do not include the Rayleigh distribution in the dictionary, since this PDF is a particular case of almost all the PDFs above [27][34][39].

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

$$L_{ij}^t = \sum_{k \in Q_{it}} \ln f_j(r_k|\xi_{ij}^t), \quad i = 1,2,\ldots,K,\ j = 1,2,\ldots,M. \tag{23}$$

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

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

$$\xi_{ij}^t = \arg\max_{\xi \in \Xi_j} \sum_{k \in Q_{it}} \ln f_j(r_k|\xi) \tag{24}$$

This approach turns out not to be feasible for several SAR-specific PDFs, such as the K distribution [39]. Hence, we avoid using ML estimates and we adopt in the M-step the "method of log-cumulants" (MoLC) [37] instead, which has been proved to be a feasible and effective estimation tool for all usual SAR parametric models [36][37].

2.4 Parameter estimation with the Mellin transform and MoLC

The method of log-cumulants (MoLC) has been recently proposed as a parametric PDF estimation technique feasible for distributions defined on $[0, +\infty)$, and has been explicitly applied in the context of the usual parametric families employed for SAR amplitude and intensity data modelling (e.g.: the Nakagami-Gamma and the K distributions) [36][37][38][51]. MoLC is based on the generalization of the usual moment-based statistics, by using the Mellin transform in the computation of characteristic functions and moment generating functions, instead of the usual Fourier and Laplace transforms.

Given a generic random variable u, the moment generating function (MGF) Φ_u of u is defined as the bilateral Laplace transform of the PDF of u [40], i.e.:

$$\Phi_u(s) = \mathcal{L}(p_u)(s) = \int_{-\infty}^{+\infty} p_u(u) \exp(su) du, \quad s \in \mathbb{C}, \tag{25}$$

where $\mathcal{L}$ is the bilateral Laplace transform operator[7] on the Lebesgue space $L^1(\mathbb{R})$ [19]. The MGF is known to converge and to be analytical at least in a vertical strip of the complex plane and it turns out to be implicitly related to the MoM estimation approach. In fact, if the interior of the convergence strip contains a neighborhood of the origin, then the ν-th order moment ($\nu = 1, 2, \ldots$) can be expressed as

$$m_\nu = E\{u^\nu\} = \Phi_u^{(\nu)}(0), \tag{26}$$

where the superscript denotes a differentiation operator [40]. Related quantities are the characteristic function of u, defined as the Fourier transform of the PDF, the second moment generating function Ψ_u, defined as the complex logarithm of the MGF, and the ν-th order cumulant k_ν, defined as the ν-th order derivative of the second MGF computed in the origin of the complex plane:

$$\Psi_u(s) = \ln \Phi_u(s), \qquad k_\nu = \Psi_u^{(\nu)}(0). \tag{27}$$

In particular, the first and second order cumulants turn out to be equal to the distribution mean and variance, respectively [40].

The MoM estimates are actually computed by analytically expressing the moments (or the cumulants) of the parametric PDF under investigation as functions of the unknown parameters, and by estimating the moments as sample-moments, thus formulating the parameter estimation problem as the solution of a (typically non-linear) system of equations.

[7] Actually, the bilateral Laplace operator would involve the exponential $\exp(-su)$ [19], but, in the context of statistics, the MGF is usually defined with the exponential $\exp(su)$ as reported in Eq. (25). However, this slight modification has no significant impact on the analytical properties of the resulting transform. Hence, hereafter we will refer to the bilateral Laplace transform as defined by Eq. (25).

In [37][38] this approach is specialized to non-negative random variables (such as SAR amplitude and intensity), corresponding to PDF defined on $[0,+\infty)$, by re-defining MGFs and characteristic functions as Mellin transforms and resulting in a more feasible estimation.

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

$$\phi_u(s) = \mathcal{M}(p_u)(s) = \int_0^{+\infty} p_u(u) u^{s-1} du, \quad s \in \mathbb{C}, \tag{28}$$

where $\mathcal{M}$ is the Mellin transform on $L^1(0,+\infty)$. Also ϕ_u is known to converge and to be analytical in a vertical strip $\mathcal{S}$ of the complex plane [37][49]. If the interior of the convergence strip contains a neighborhood of 1, then the following definitions are formulated by analogy with the Laplace-based case [37]:

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

The expressions "log-moments" and "log-cumulants" are also employed for the second kind moments and cumulants, thanks to their relation with the moments of the logarithm of u, i.e. [37][51]:

$$\mu_\nu = E\{(\ln u)^\nu\}, \quad \kappa_1 = \mu_1 = E\{\ln u\}, \quad \kappa_2 = \mathrm{Var}\{\ln u\}, \quad \kappa_3 = E\{(\ln u - \kappa_1)^3\}. \tag{29}$$

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

Table 2.1 shows the resulting non-linear equations for the parametric families (see Eqs. (17)-(22)) adopted in the proposed dictionary-based method [34][36][37]. The application of MoLC to the models $f_1, f_2, \ldots, f_5$ requires the computation of a sample-mean estimate $\hat{\kappa}_1$ of κ_1 and of a sample-variance estimate $\hat{\kappa}_2$ of κ_2, respectively, whereas f_6 (namely, K-root) also involves a sample estimate $\hat{\kappa}_3$ of κ_3.

The solution of the resulting equations turns out to be feasible and fast for all the considered distributions. Specifically, log-normal does not require a real solution process, since the parameters of this distributions are exactly the first two log-cumulants. SαSGR and Weibull allow an analytical solution of the corresponding system of two equations. Nakagami, GGR, and K-root require a numerical solution procedure, but, thanks to the strict monotonicity properties of the functions involved, this procedure has been proved to be simple and fast for all the three parametric families [34][36][37]. In addition, for several of the considered models, good estimation properties have been proved theoretically for the

Parametric family		MoLC equation
f_1	Log-Normal	$\kappa_1 = \mu$ $\kappa_2 = \sigma^2$
f_2	Nakagami	$2\kappa_1 = 2\ln\mu + \Psi(L) - \ln L$ $4\kappa_2 = \Psi(1, L)$
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	$2\kappa_1 = \ln\mu + \Psi(L) - \ln L + \Psi(M) - \ln M$ $4\kappa_2 = \Psi(1, L) + \Psi(1, M)$ $8\kappa_3 = \Psi(2, L) + \Psi(2, M)$

Table 2.1: MoLC equations for all the parametric families included in the adopted dictionary ($\Psi(\cdot)$ is the digamma function [8][37], $\Psi(\nu, \cdot)$ is the ν-th order polygamma function [8][37], and $G_\nu(\cdot)$ is the integral function introduced in [34] for GGR parametric estimation.)

MoLC approach. In particular, the MoLC estimates exhibit a lower variance with respect to the MoM ones for the Nakagami distribution [37], and are consistent for the GGR one [34].

However, as pointed out in [34], the MoLC equations for GGR and K-root can yield no solutions for specific values of $\hat{\kappa}_2$ and $\hat{\kappa}_3$. In such situations, these parametric families are not compatible with the empyrical data distributions and are not considered in the selection of the optimal model.

2.5 Overall architecture of the proposed method

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

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

$$\tau_i^t(z) = \frac{P_i^t p_i^t(z)}{p^t(z)}, \quad \text{where} \quad p^t(z) = \sum_{i=1}^{K} P_i^t p_i^t(z); \tag{30}$$

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

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

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

$$\kappa_{2i}^t = \frac{\sum_{z\in Q_{it}} h(z)(\ln z - \kappa_{1i}^t)^2}{\sum_{z\in Q_{it}} h(z)}, \quad \kappa_{3i}^t = \frac{\sum_{z\in Q_{it}} h(z)(\ln z - \kappa_{1i}^t)^3}{\sum_{z\in Q_{it}} h(z)}, \tag{31}$$

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

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

$$L_{ij}^t = \sum_{z\in Q_{it}} h(z)\ln f_j(z|\xi_{ij}^t) \tag{32}$$

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

Thus, the proposed selection of an optimal SAR-specific model for each mixture component operatively requires the substitution of the usual ML-based M-step with the described MoLC-step, and the integration in the iterative procedure of a further MS-step, devoted to the model selection process. The resulting "dictionary-based" SEM method will be denoted hereafter simply as DSEM.

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

As in the general SEM framework, the resulting sequence of PDF estimates is expected not to converge pointwise nor a.s., but to reach a stationary behavior, thus requiring the

definition of a specific procedure to extract a single optimal PDF estimate from the sequence itself. We adopt here the approach suggested in [5], which computes, at any iteration t ($t = 0, 1, \ldots$), the log-likelihood of the current PDF estimate $p_r^t(\cdot)$ over the whole image data set $\mathcal{I}$, i.e.:

$$L_{\mathcal{I}}^t = \sum_{z=0}^{Z-1} h(z) \ln p_r^t(z), \tag{33}$$

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

On the other hand, the log-likelihood cannot be adopted as a criterion to choose the optimal number K of components, due to the monotonic relation between these two quantities [44]. Several different validation functionals have been proposed in the literature as selection criteria for the parameter K, for instance, according to a Bayesian model-based framework [3][4], to discriminant analysis [1][24][31], or to the MML approach [44]. In the context of the proposed method, we adopt the correlation coefficient ρ_K [40] 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 selection of a suitable number of mixture components is automatized by predefining a maximum number K_{max} of components and by computing the K-component PDF estimate for all $K = 1, 2, \ldots, K_{max}$, thus choosing the number K^* of components yielding the highest value ρ^* of ρ_K.

We stress, in particular, that the proposed PDF estimation method for SAR amplitude data turns out to be completely automatic. In fact, the selection of the number of mixture components, the choice of the optimal model for each component and the estimation of the parameters of the model are jointly and automatically performed by the algorithm without any need for user intervention. Only the maximum number K_{max} of components and the maximum number t_{max} of iterations have to be defined prior to the application of the method, but the choice of K_{max} is not critical, since K_{max} is only an upper bound on the number of components [44][31], and t_{max} only has to be large enough to let the iterative SEM process reach stationarity.

3 Experimental results

3.1 Data sets for experiments

The proposed DSEM algorithm for PDF estimation has been tested on 11 real SAR images, and compared with several usual SAR-specific parametric PDF estimation strategies and with a non-parametric approach. The first six images used for experiments are single bands acquired in August 1989 on the agricultural region of Feltwell (UK) by a fully polarimetric PLC-band NASA/JPL airborne sensor [48]. Specifically, all the three polarizations (HH, HV, VV) acquired at band C, the HV and VV polarizations acquired at band L and the HH polarization acquired at band P have been used. The remaining channels (namely, L-HH,

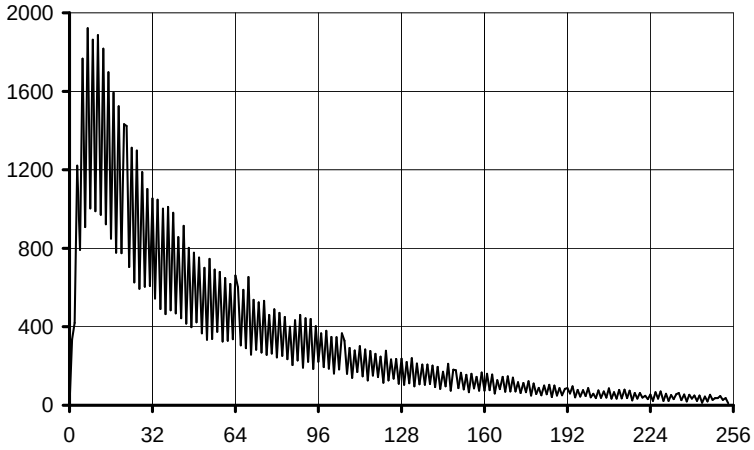

Figure 3.1: Plot of the histogram of the HH polarization acquired at band L for the "Feltwell" data set, highlighting the presence of an irregular "zig-zag" behavior of the histogram itself.

P-HV, and P-VV) have been discarded, since their histograms exhibit strong irregularities (see, for instance, Fig. 3.1), probably due to the underlying calibration and registration processes. Hereafter, the adopted Feltwell bands will be denoted synthetically as "Feltwell-CHH", "Feltwell-CHV", ..., "Feltwell-PHH".

The other five employed images are:

- a single-look ERS-1 image, acquired in April 1993 on the urban and agricultural region around Bourges (France);
- an ERS-1 image of the agricultural region of Flevoland (Netherlands);
- a 3-look XSAR scene (hereafter denoted as "Suisse-Lake") of a portion of the Swiss territory, comprising a mountain zone, a lake, and a urban area [13];
- a 3-look XSAR image (hereafter denoted as "Suisse-Mountain") of a mountain area in the Swiss territory;
- a 3-look ESAR image of the area of Oberpfaffenhofen near Munich (Germany).

We stress, in particular, that "Suisse-Mountain" exhibits a bimodal histogram, whereas all the other employed images have a unimodal statistics. All images are shown in Appendix A, after histogram stretching and/or equalization, and all the corresponding histograms are reported in Appendix B.

3.2 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

Image	ρ^*	K^*
Feltwell-CHH	99,42%	2
Feltwell-CHV	99,41%	5
Feltwell-CVV	99,62%	3
Feltwell-LHV	99,41%	3
Feltwell-LVV	99,70%	6
Feltwell-PHH	99,44%	2
Bourges	99,81%	3
Flevoland	99,88%	6
Suisse-Lake	99,70%	4
Suisse-Mountain	99,88%	4
Oberpfaffenhofen	99,86%	1

Table 3.1: Results for the DSEM algorithm applied to all the employed SAR images: correlation coefficient ρ^* between the estimated PDF and the image histogram and optimal number K^* of mixture components.

coefficients with the image histograms, and qualitatively, by visually comparing the plots of the estimates and of the histograms. Appendix B shows all the plotted comparisons for all the 11 images.

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 the 11 considered images (see Table 3.1), thus assessing the effectiveness of the proposed method from the viewpoint of the estimation accuracy. The visual comparison between the PDF estimates and the corresponding image histograms confirms this conclusion, as shown, for example, in Fig. 3.2.

In order to further assess the capabilities of the method, a comparison has also been performed with several other standard parametric models for SAR amplitude data. Specifically, all the six (either theoretical or empirical) models included in the dictionary have been employed in this comparison, and the corresponding parameters have been estimated by using MoLC. The resulting correlation coefficients are listed in Table 3.2. A comparison between Table 3.1 and Table 3.2 shows that the proposed DSEM algorithm yields the PDF estimate with the highest correlation coefficients with the image histograms of 10 of the 11 images, the exception being "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% provided by K-root and by DSEM. In particular, "Oberpfaffenhofen" is the only data set for which the automatic method for the selection of an optimal number of clusters chooses $K^* = 1$ (see Table 3.1). In fact, "Oberpfaffenhofen" exhibits a very small dynamics range, thus being accurately modelled by a single-component distribution (Fig. B.11).

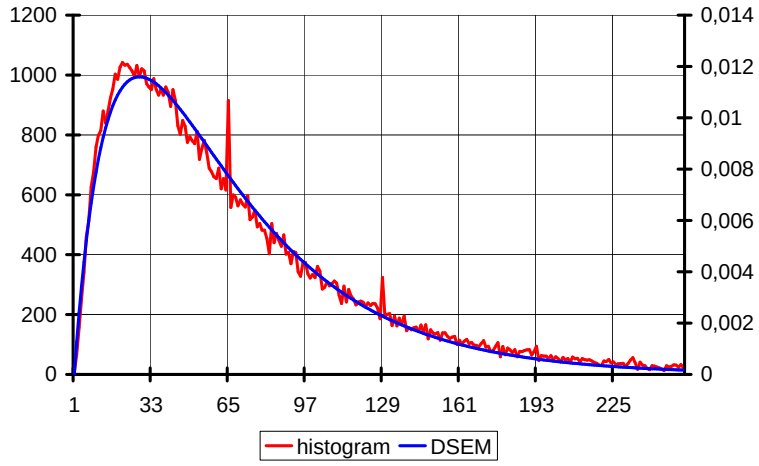

Figure 3.2: Plot of the image histogram and of the DSEM PDF estimate for the "Feltwell-CHH" data set.

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%

Table 3.2: Correlation coefficients between the estimated PDFs and the image histograms for all considered parametric families and all SAR images.

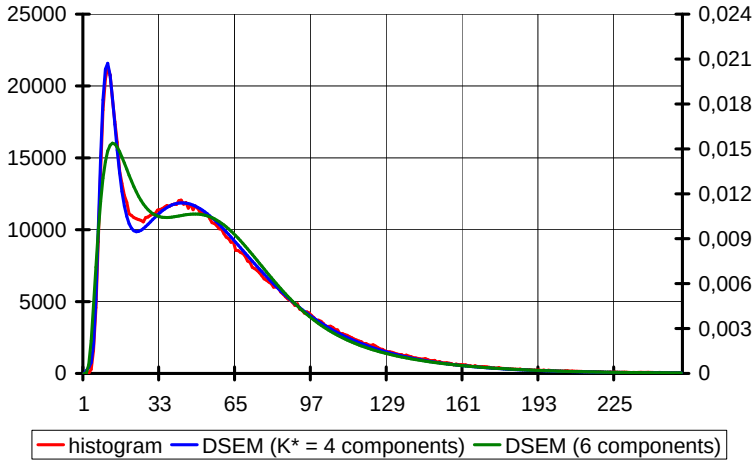

Figure 3.3: Plot of the image histogram and of the DSEM PDF estimates with 4 ($= K^*$) and 6 components, respectively, for the "Suisse-Mountain" data set.

In particular, the good results achieved with "Suisse-Mountain", which exhibits a bimodal histogram, stress the usefulness of the adopted FMM approach. As shown in Fig. 3.3, DSEM effectively describes the bimodal statistics; on the contrary, all the parametric models considered in Table 3.2 are intrinsically monomodal and provide poor estimates. We stress, in particular, the relevance, in this case, of the automatic selection of an optimal number K^* of components. DSEM selects $K^* = 4$, which yields a high correlation coefficient (Table 3.1) and a visually accurate identification of both modes of the distribution (Fig. 3.3), but a different number of components would result in a worse estimate. In Fig. 3.3 we show, for instance, the six-components PDF estimate, which detects the bimodal structure of the histogram but strongly underestimates the height of the left mode and exhibits a bias in the identification of the position of the right one.

On the other hand, we also note from the comparison between Table 3.1 and Table 3.2 that for most of the remaining images, presenting a unimodal histogram, at least one of the single-model parametric distributions listed in Table 3.2 achieves results similar to the ones provided by DSEM. In particular, at least one among the GGR, K-root, log-normal and Weibull distributions allows to obtain very accurate estimation results. In these cases, thanks to the unimodal empirical distribution, a FMM-based approach is not mandatory and a single-component PDF estimate turns out to be effective. A multi-component model allows obtaining also in such situations a slightly better result than a single-component one, but only with a small improvement. For instance, in Fig. 3.4 we plot for the "Feltwell-CHV" image the PDF estimates provided by DSEM and by the best performing single parametric model for this image (namely, GGR).

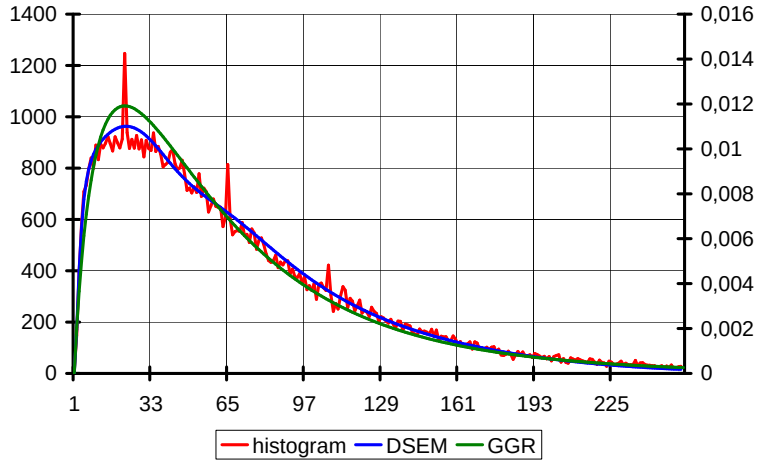

Figure 3.4: Plot of the image histogram and of the DSEM and GGR PDF estimates for the "Feltwell-CHV" data set.

3.3 Comparison with a non-parametric estimator

The results provided by the developed DSEM algorithm have also been compared with the ones achieved by the well-known kernel-based Parzen estimation method [17][18], adopted as a benchmark non-parametric strategy. Specifically, the Parzen estimate is a linear combination of kernel functions centered on the available data samples [17][41], and Gaussian kernels have been adopted in the present comparative experiment. However, a critical issue for Parzen PDF estimation is the selection of a suitable value for the kernel width, providing a good trade-off between correctly modelling the image histogram and avoiding overfitting the histogram itself.[8]

In order to avoid the usual "trial-and-error" approach to the optimization of the kernel-width, the well-known cross validation (CV) selection method has been adopted, which computes, in a predefined search range, the kernel width value minimizing an estimate of the L^2 distance between the estimated and the true PDFs [2][7][23]. In addition, a SAR-specific selection criterion has also been developed, which computes the asymptotically optimal kernel width corresponding to a reference Rayleigh model for the unknown SAR amplitude PDF. Further details about the kernel width optimization procedures are provided in Appendix C. Hereafter, we will refer to the CV approach and to the developed SAR-specific one as "ParzenCV" and as "ParzenSAR", respectively.

ParzenCV has been applied by selecting, for each considered SAR image, the best kernel width inside the range $[0, 25]$. For all the 11 images, the CV functional exhibits a monomodal

[8] Note that the correlation coefficient between the PDF estimate and the histogram is not a good selection criterion here, since it takes on its maximum value of 100% in the case of zero width (i.e., with perfect histogram overfitting).

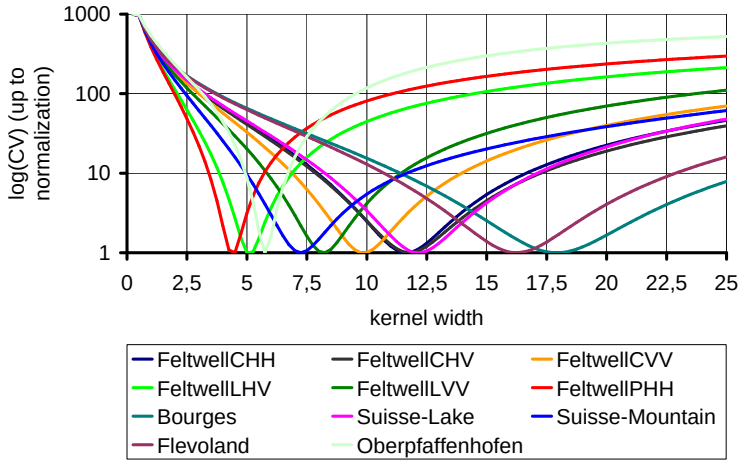

Figure 3.5: Logarithmic plot of the CV functional for all the considered 11 images (the 11 functionals have been rescaled in the range $[1, 1000]$ only for visualization purposes).

shape, with a single global minimum inside the range $[0, 25]$ (see Fig. 3.5), thus not involving critical minimization issues.

The correlation coefficients provided by ParzenCV and by ParzenSAR are listed in Table 3.3 and all the corresponding PDF estimates are plotted in Appendix C. A comparison between these two strategies suggests that ParzenSAR provides more accurate estimates in the present context than ParzenCV, and a comparison between Table 3.3 and Table 3.1 highlights ParzenSAR and DSEM to exhibit very similar accuracy performances, with sometimes slightly better results for DSEM. However, a higher quality for the ParzenSAR estimates than for the ParzenCV ones is an expected result, since the ParzenSAR kernel width optimization strategy is explicitly tailored to the context of SAR amplitude data, through the adoption of a Rayleigh reference distribution for the image noise. We stress, in particular, that the proposed parametric DSEM algorithm performs very similarly to this SAR-specific approach to non-parametric Parzen estimation and better than the CV-based approach, thus further assessing the flexibility of DSEM and highlighting the importance of a correct definition of the underlying image model.

A visual analysis of the corresponding PDF estimates confirms the previous conclusions and stresses, in particular, the ParzenCV estimates to be sometimes oversmoothed (see Figs. 3.6-3.7). In particular, ParzenCV strongly overestimates the kernel width for the "Suisse-Mountain" image, thus poorly modelling the left mode in the corresponding histogram, whereas ParzenSAR performs well, with a result visually very similar to the one produced by DSEM (see Fig. 3.7).

Image	ParzenCV	ParzenSAR
Feltwell-CHH	98,47%	99,56%
Feltwell-CHV	97,97%	99,11%
Feltwell-CVV	98,23%	99,51%
Feltwell-LHV	96,08%	97,41%
Feltwell-LVV	97,94%	99,38%
Feltwell-PHH	96,41%	98,64%
Bourges	99,26%	99,82%
Flevoland	99,41%	99,93%
Suisse-Lake	98,85%	99,89%
Suisse-Mountain	95,38%	99,62%
Oberpfaffenhofen	92,92%	99,9986%

Table 3.3: Correlation coefficients between the histograms and the corresponding Parzen PDF estimates, computed both selecting the optimal kernel width according to the CV criterion (ParzenCV) and according to the developed SAR-specific algorithm (ParzenSAR).

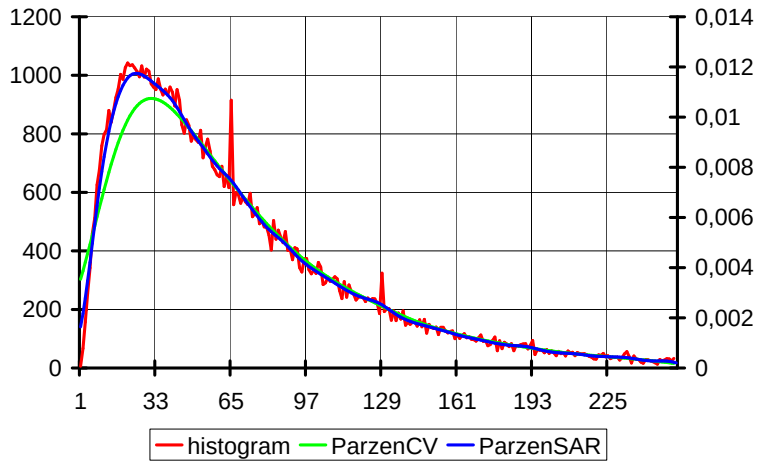

Figure 3.6: Plot of the image histogram and of the Parzen PDF estimates for the "Feltwell-CHH" data set.

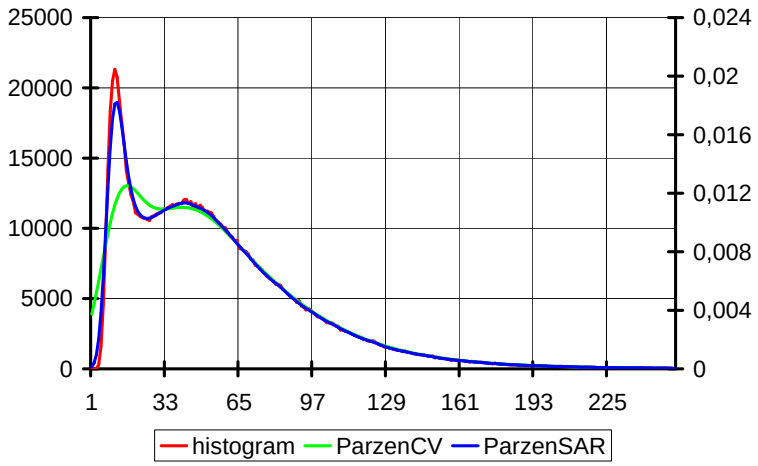

Figure 3.7: Plot of the image histogram and of the Parzen PDF estimates for the "Suisse-Mountain" data set.

4 Conclusions

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

The numerical results of the application of the method to several real SAR images acquired by XSAR, ESAR, ERS-1, and airborne SAR sensors prove the proposed DSEM algorithm to provide very accurate PDF estimates, both from the viewpoint of a visual comparison between the estimates and the corresponding image histograms, and from the viewpoint of the quantitative correlation coefficient between them. We stress, in particular, that the method proves to be effective on all the considered images, despite of their different statistics (i.e., histogram unimodality or multimodality), dynamics range or multilooking characteristics. Correlation coefficients higher that 99% are obtained, in fact, in all cases, thus proving the flexibility of the method.

Specifically, the use of a mixture model is mandatory when dealing with multimodal statistics. Applied to the "Suisse-Mountain" image, which exhibits a bimodal histogram, the developed DSEM algorithm correctly detects the position and the heights of both statistical modes. On the other hand, the results provided by DSEM in case of unimodal histograms are usually comparable with the ones yielded by at least one of the single-component parametric

models included in the dictionary. In case of a unimodal distribution, the estimation results provided by the single-component models are already very good, and the multi-component DSEM estimator can yield only a minor improvement. In particular, the experiments have suggested the generalized Gaussian Rayleigh and the K models as the most effective single-component parametric families, achieving a correlation coefficient higher than 98% in all experiments. However, as discussed in [34], the estimation process for both models can provide no solutions for specific combinations of the values of the sample-log-cumulants. On the other hand, good results are also provided by using a log-normal or a Weibull distribution, which can be feasibly fitted to any image histogram without restrictions and which gives correlation coefficients often higher than 98%.

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

A comparison with the non-parametric Parzen PDF estimation method confirms these conclusions. Specifically, DSEM produces more accurate estimates than the Parzen algorithm, applied adopting the well-known cross-validation criterion in order to optimize the value of the kernel width. On the other hand, a SAR-specific version of the Parzen method is also developed here, which adopts a Rayleigh PDF as a reference SAR amplitude distribution and analytically derives the corresponding asymptotically optimal kernel width. The resulting non-parametric estimator turns out to be very effective, providing results similar to the ones produced by DSEM.

The results of these two proposed estimation approaches (namely, the parametric DSEM one and the non-parametric Parzen one with SAR-specific optimization of the kernel width) further stress the importance of the adoption of a correct model for the greylevel statistics in the context of SAR image analysis.

Acknowledgments

This research was carried out within the framework of the IMAVIS project which was funded by the European Union (EU). The support is gratefully acknowledged. The authors would also like to thank the French Space Agency CNES and the French Research Center CESBIO which provided the "Bourges", "Flevoland", "Suisse-Lake", "Suisse-Mountain", and "Oberpfaffenhofen" data sets, available on the CD-ROM "Speckle filters compararative tests" (©CNES, 2001).

Appendix

A SAR images employed for experiments

The present Appendix shows the 11 images used in the experiments (Figs. A.1-A.11). Only for visualization purposes, their histograms have been stretched and/or equalized.

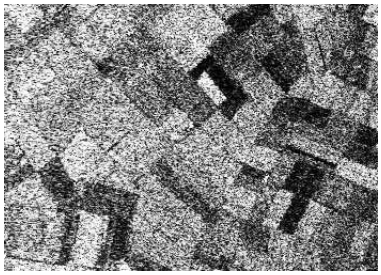

Figure A.1: "Feltwell-CHH" image.

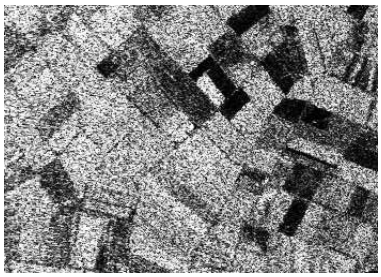

Figure A.2: "Feltwell-CHV" image.

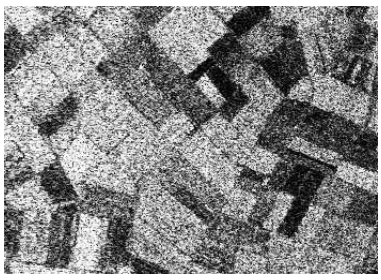

Figure A.3: "Feltwell-CVV" image.

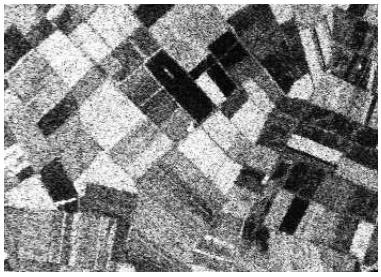

Figure A.4: "Feltwell-LHV" image.

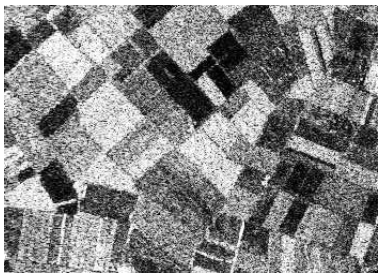

Figure A.5: "Feltwell-LVV" image.

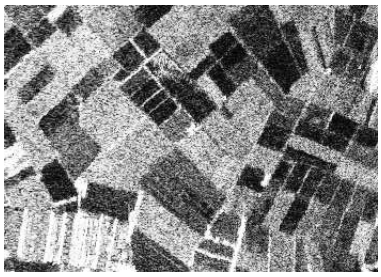

Figure A.6: "Feltwell-PHH" image.

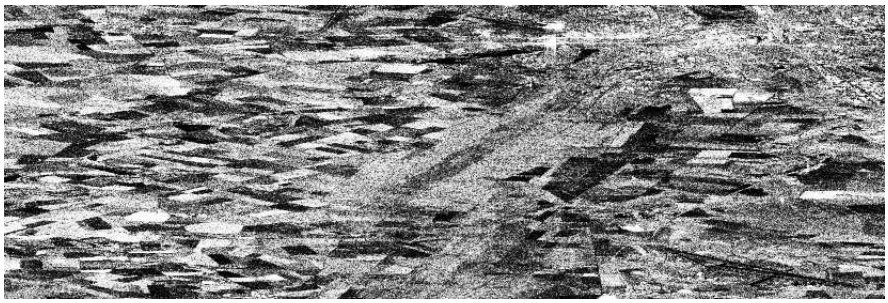

Figure A.7: "Bourges" image.

Figure A.8: "Flevoland" image.

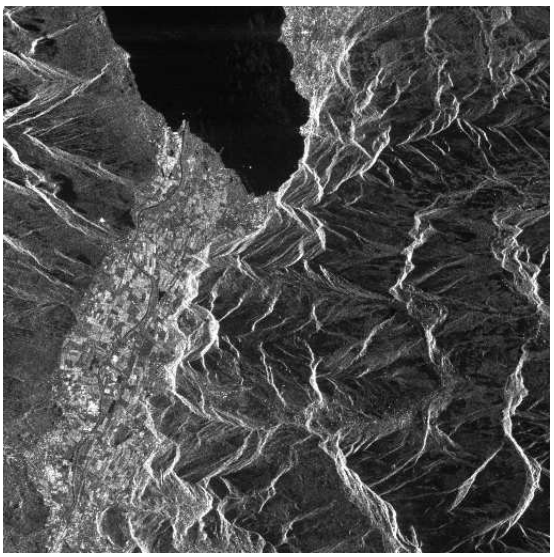

Figure A.9: "Suisse-Lake" image.

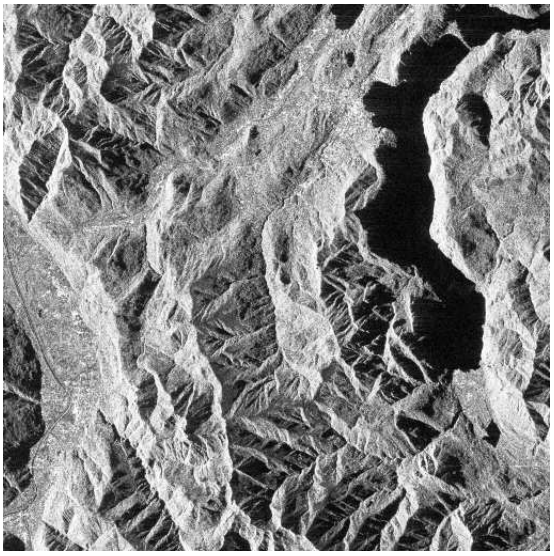

Figure A.10: "Suisse-Mountain" image.

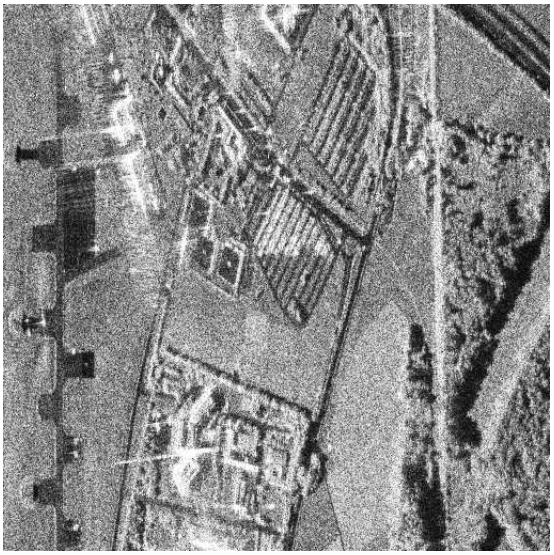

Figure A.11: "Oberpfaffenhofen" image.

B Plots of the estimated PDFs

In Appendix B we report the estimated PDFs provided by the proposed DSEM algorithm (Figs. B.1-B.11), together with the histograms of the corresponding images.

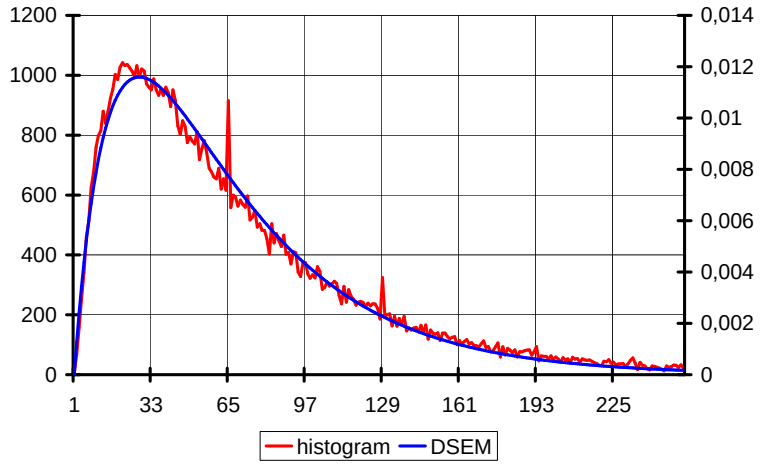

Figure B.1: Plot of the image histogram and of the DSEM PDF estimate for the "Feltwell-CHH" data set.

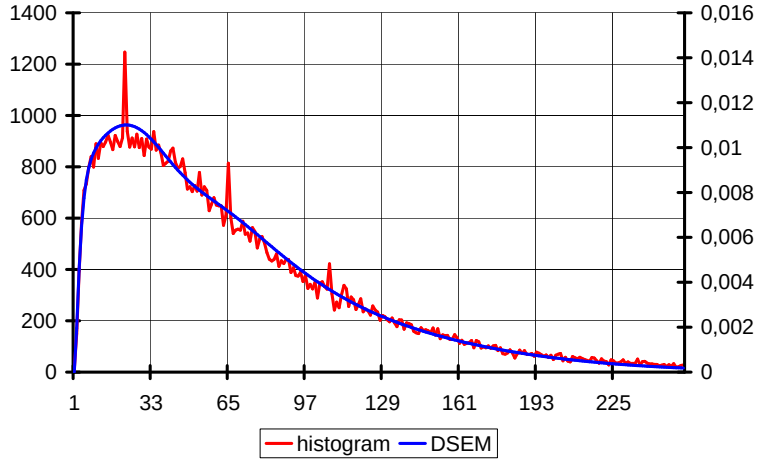

Figure B.2: Plot of the image histogram and of the DSEM PDF estimate for the "Feltwell-CHV" data set.

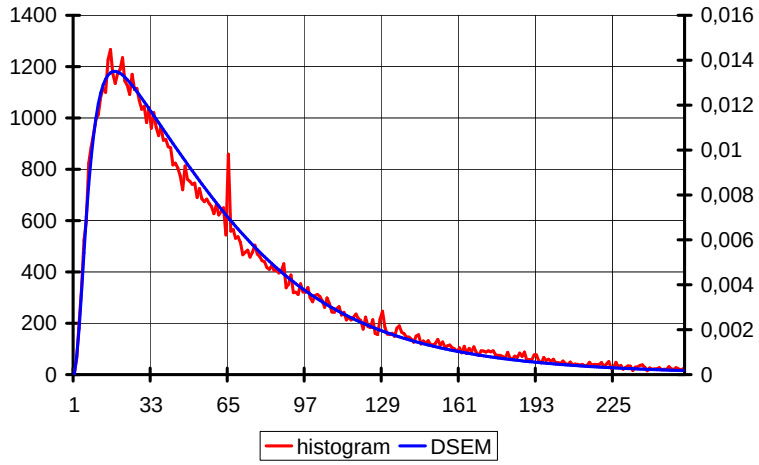

Figure B.3: Plot of the image histogram and of the DSEM PDF estimate for the "Feltwell-CVV" data set.

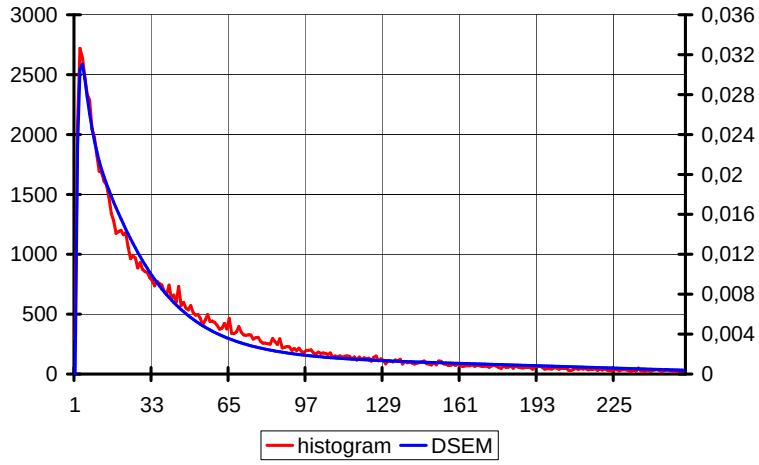

Figure B.4: Plot of the image histogram and of the DSEM PDF estimate for the "Feltwell-LHV" data set.

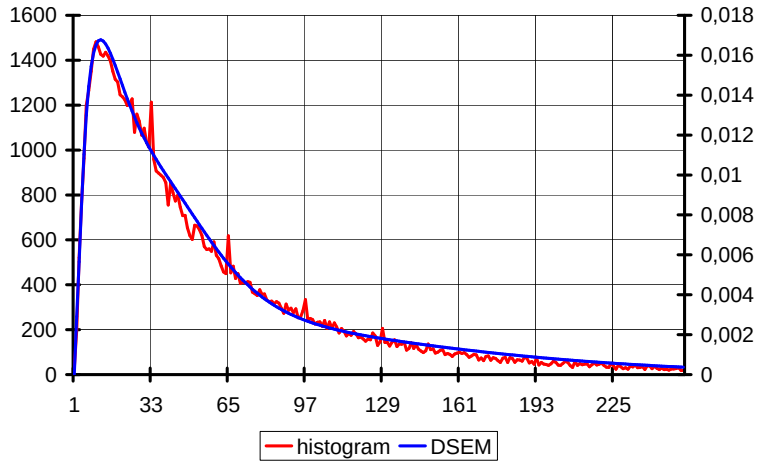

Figure B.5: Plot of the image histogram and of the DSEM PDF estimate for the "Feltwell-LVV" data set.

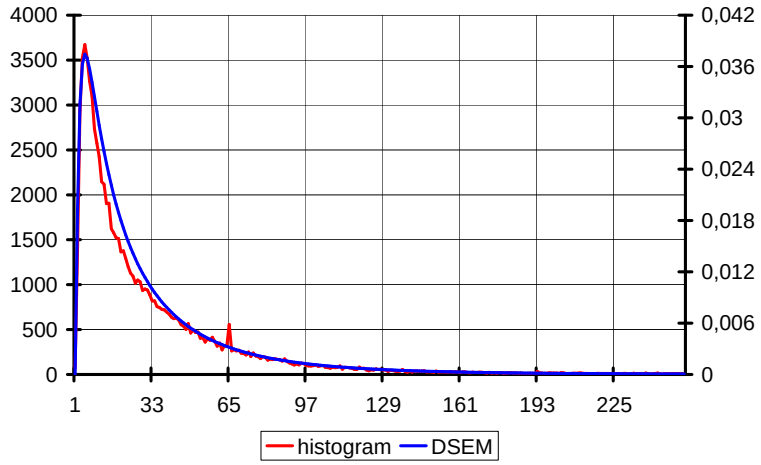

Figure B.6: Plot of the image histogram and of the DSEM PDF estimate for the "Feltwell-PHH" data set.

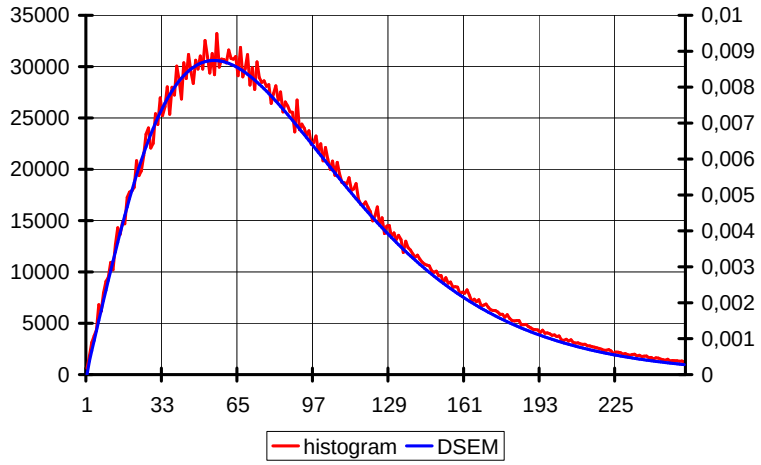

Figure B.7: Plot of the image histogram and of the DSEM PDF estimate for the "Bourges" data set.

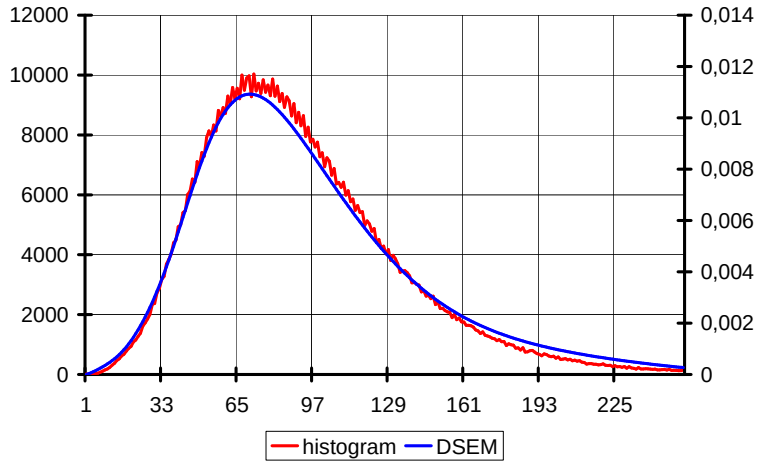

Figure B.8: Plot of the image histogram and of the DSEM PDF estimate for the "Flevoland" data set.

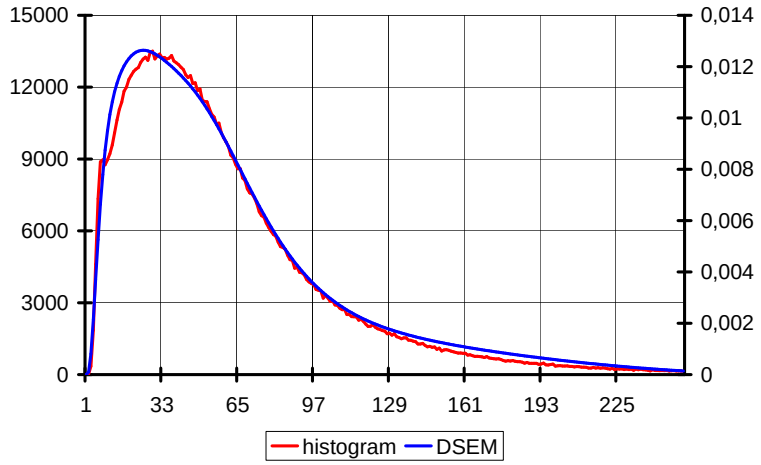

Figure B.9: Plot of the image histogram and of the DSEM PDF estimate for the "Suisse-Lake" data set.

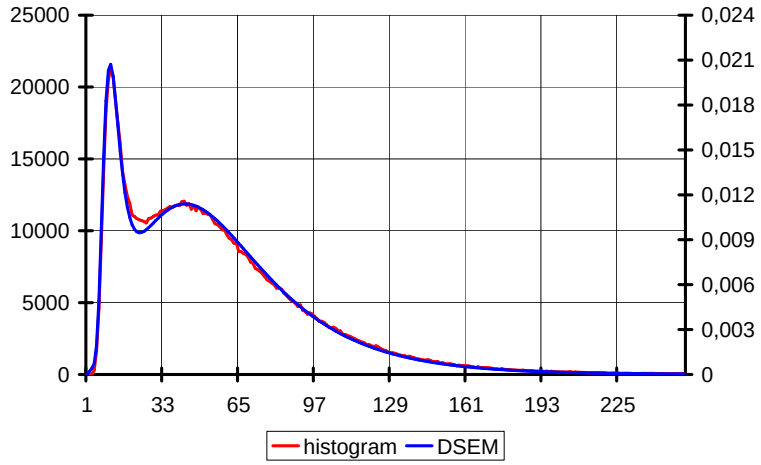

Figure B.10: Plot of the image histogram and of the DSEM PDF estimate for the "Suisse-Mountain" data set.

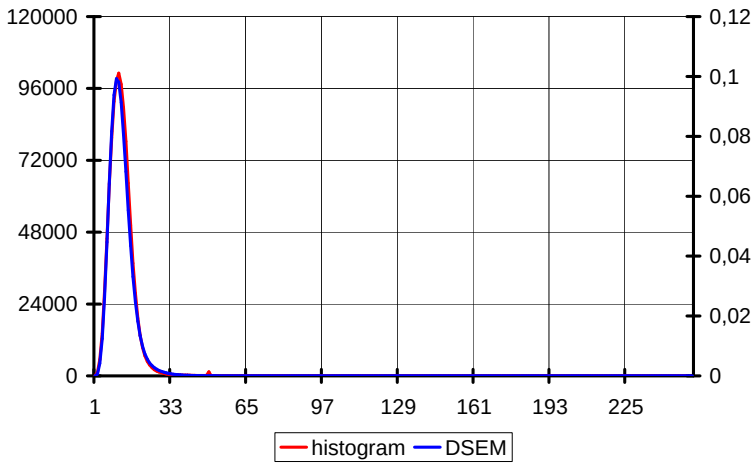

Figure B.11: Plot of the image histogram and of the DSEM PDF estimate for the "Oberpfaffenhofen" data set.

C SAR-specific optimal kernel width selection

Given a SAR amplitude image $\mathcal{I} = \{r_1, r_2, \ldots, r_N\}$ with histogram $\{h(z) : z = 0, 1, \ldots, Z-1\}$, the non-parametric Parzen estimate for the greylevel PDF $p_r(\cdot)$ is defined as a linear combination of kernel functions centered on the image samples [17][18][41], i.e.:

$$\hat{p}_\sigma(r) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{\sigma}K\left(\frac{r-r_k}{\sigma}\right) = \frac{1}{N}\sum_{z=0}^{Z-1}\frac{h(z)}{\sigma}K\left(\frac{r-z}{\sigma}\right), \tag{C.1}$$

where $K(\cdot) : \mathbb{R} \to [0, +\infty)$ is a non-negative, symmetrical and Borel-measurable function with $\int_{\mathbb{R}} K(x)dx = 1$ (namely, the kernel function) and σ is a positive "kernel width" parameter, to be set in order to avoid both overfitting the image histogram (for too small values of σ) and oversmoothing it (for too large values of σ). "Trial-and-error" approaches are often adopted to select a suitable value for σ for a given data set. On the other hand, several automatic selection algorithms have been proposed to address this selection task, such as L^2 cross validation, L^1 cross validation, plug-in methods, or double kernel approaches: we refer to [2] for a comprehensive review of the main developed strategies.

Here, as a first criterion, assuming $p_r, K \in L^2(\mathbb{R})$, we adopt the L^2 cross-validation (CV) approach, which aims at minimizing the L^2 squared distance between the true PDF p_r and the Parzen estimate $\hat{p}_\sigma$, i.e.[9]:

$$||\hat{p}_\sigma - p_r||^2 = ||\hat{p}_\sigma||^2 - 2(\hat{p}_\sigma, p_r) + ||p_r||^2 =$$

$$= ||\hat{p}_\sigma||^2 - 2\int_0^{+\infty}\hat{p}_\sigma(r)p_r(r)dr + ||p_r||^2 = ||\hat{p}_\sigma||^2 - 2E\{\hat{p}_\sigma(r)\} + ||p_r||^2. \tag{C.2}$$

Dropping $||p_r||^2$ (which does not depend on σ) and introducing an asymptotically unbiased estimate of the $E\{\hat{p}_\sigma(r)\}$, the following cross-validation functional is obtained [2][8]:

$$CV(\sigma) = ||\hat{p}_\sigma||^2 - \frac{2}{N^2}\sum_{k\neq h}\frac{1}{\sigma}K\left(\frac{r_k - r_h}{\sigma}\right), \tag{C.3}$$

where $||\hat{p}_\sigma||^2$ can be computed analytically according to Eq. (C.1).

Then, as a second kernel width selection approach, we specialize to the present SAR amplitude context a general result proved in [35] about asymptotically optimal mean square non-parametric estimation. Specifically, it has been proved [2][35] that, under mild smoothness assumptions on p_r (in particular, p_r has to be twice differentiable with $p_r'' \in L^2(\mathbb{R})$), the kernel width value asymptotically minimizing the Mean Integrated Square Error (MISE) $E\{||\hat{p}_\sigma - p_r||^2\}$ between the true and the estimated PDFs is given by:

$$\sigma^* = \left[\frac{||K||^2}{N m_2^2(K)||p_r''||^2}\right]^{1/5}, \quad \text{where} \quad m_2(K) = \int_{-\infty}^{+\infty} x^2 K(x)dx. \tag{C.4}$$

[9] All the norms and the inner products in the present Appendix are implicitly referred to the Hilbert space $L^2(\mathbb{R})$ [19][46].

Adopting a Gaussian $\mathcal{N}(0,1)$ kernel we have:

$$||K||^2 = \frac{1}{2\sqrt{\pi}}, \quad m_2(K) = 1 \Longrightarrow \sigma^* = (2N\sqrt{\pi})^{-1/5}||p_r''||^{-2/5}. \tag{C.5}$$

Operatively applying this result to compute an optimal kernel width requires the adoption of a reference model for the PDF p_r, allowing a feasible calculation of the L^2 norm $||p_r||$. This reference density is usually assumed to be Gaussian, but SAR amplitude data are known to be strongly non-Gaussian, thus suggesting a different choice. For the present problem of SAR amplitude PDF estimation we have selected a Rayleigh reference PDF, i.e.:

$$p_r(r) = 2\lambda r \exp(-\lambda r^2), \quad r \geq 0, \tag{C.6}$$

both for its simplicity and because it is the basic theoretic model for single-look SAR amplitude data. The resulting second derivative turns out to be:

$$p_r''(r) = 4\lambda^2(2\lambda r^3 - 3r)\exp(-\lambda r^2), \quad r \geq 0. \tag{C.7}$$

which yields, after standard integral calculations:

$$||p_r''||^2 = \int_0^{+\infty} p_r''(r)^2 dr = \frac{15\sqrt{2\pi}}{4}\lambda^{5/2} \tag{C.8}$$

In particular, the mean square of the Rayleigh distribution is known to be $E\{r^2\} = 1/\lambda$ [39], which finally yields:

$$||p_r''||^2 = \frac{15\sqrt{2\pi}}{4}E\{r^2\}^{-5/2} \Longrightarrow \sigma^* = \left(\frac{\sqrt{2}}{15\pi N}\right)^{1/5} E\{r^2\}^{1/2}. \tag{C.9}$$

Hence the MISE asymptotically optimal kernel width for a Rayleigh reference distribution and a Parzen estimator with Gaussian kernels is directly proportional to the root mean square of the greylevel. Estimating this expected value as a sample-quadratic-moment, we finally obtain an estimate of the MISE asymptotically optimal width:

$$\hat{\sigma}^* = \left(\frac{\sqrt{2}}{15\pi N}\right)^{1/5}\left(\frac{1}{N}\sum_{k=1}^{N} r_k^2\right)^{1/2} = \left(\frac{\sqrt{2}}{15\pi N}\right)^{1/5}\left[\frac{1}{N}\sum_{z=0}^{Z-1} h(z)z^2\right]^{1/2}. \tag{C.10}$$

The Parzen estimation method has been applied to the 11 images used in the experiments with both kernel width selection algorithms. Figs. C.1-C.11 show the estimated PDFs provided by both approaches, together with the histograms of the corresponding images. Note that the CV strategy (namely, "ParzenCV") involves the computation of the functional (C.3) for several values of the kernel width, by scanning the adopted search range (here 100 equally spaced values in $[0, 25]$ have been employed), whereas the SAR-specific approach (namely, "ParzenSAR") is a "one-shot" procedure, only requiring the computation of the sample root mean square of the image greylevels.

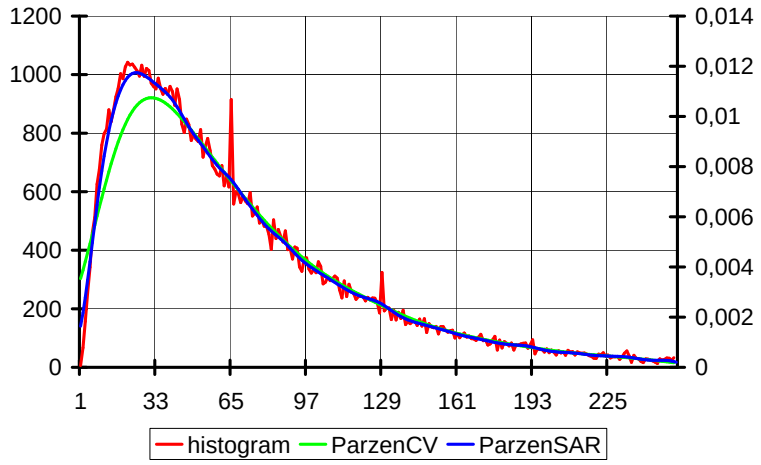

Figure C.1: Plot of the image histogram and of the Parzen PDF estimates for the "Feltwell-CHH" data set.

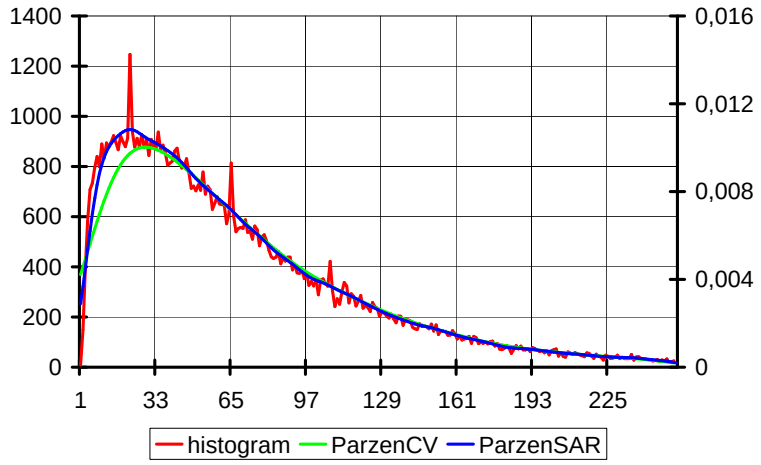

Figure C.2: Plot of the image histogram and of the Parzen PDF estimates for the "Feltwell-CHV" data set.

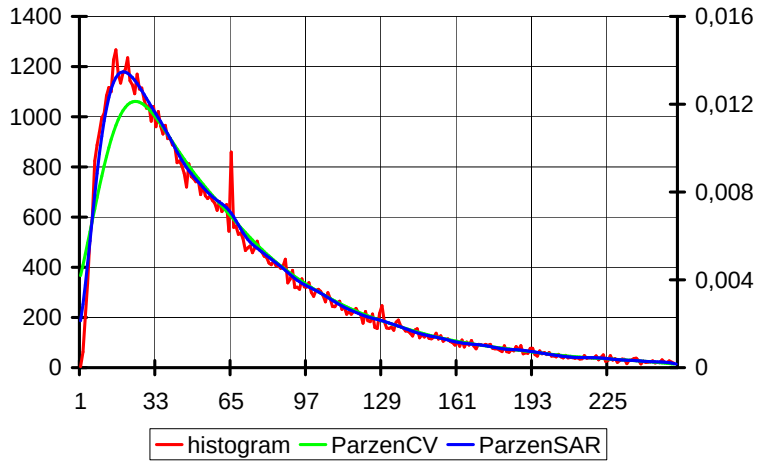

Figure C.3: Plot of the image histogram and of the Parzen PDF estimates for the "Feltwell-CVV" data set.

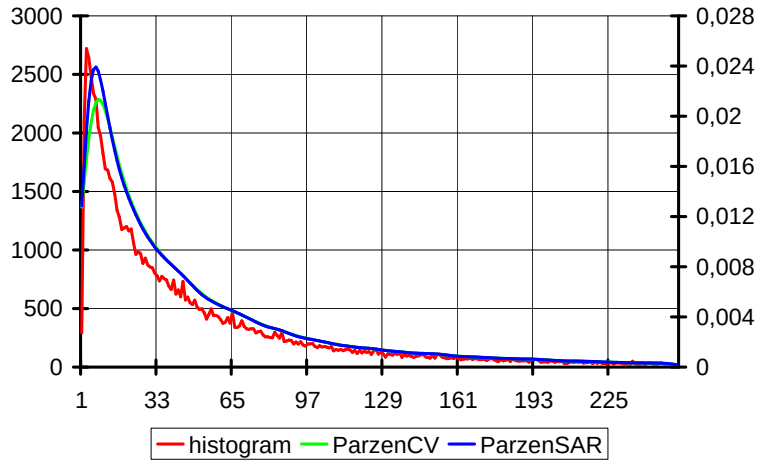

Figure C.4: Plot of the image histogram and of the Parzen PDF estimates for the "Feltwell-LHV" data set.

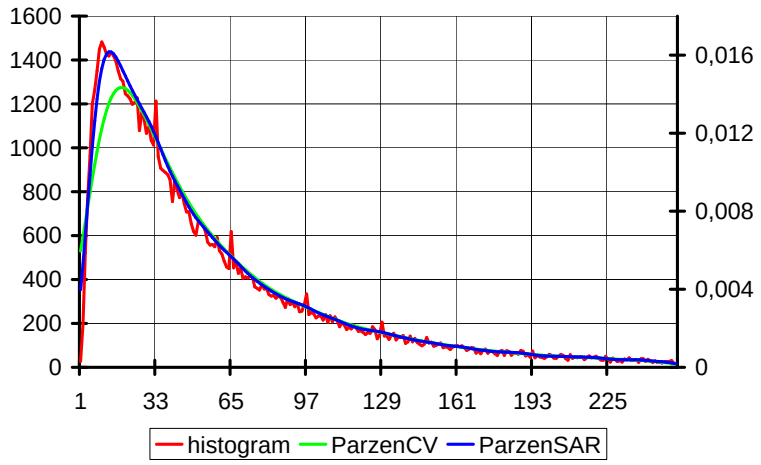

Figure C.5: Plot of the image histogram and of the Parzen PDF estimates for the "Feltwell-LVV" data set.

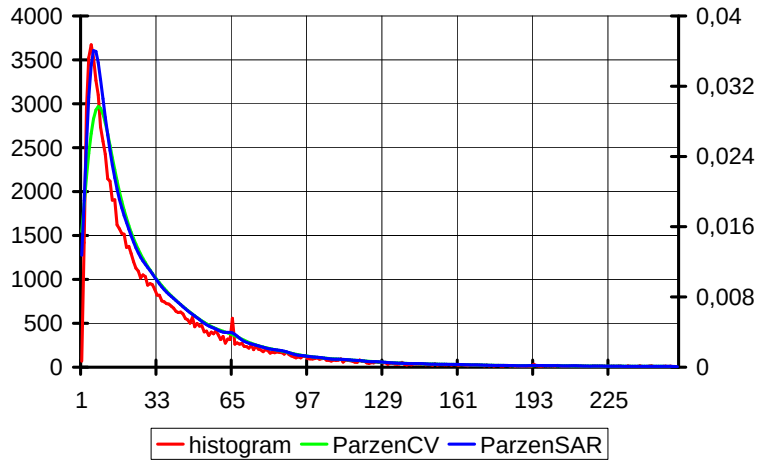

Figure C.6: Plot of the image histogram and of the Parzen PDF estimates for the "Feltwell-PHH" data set.

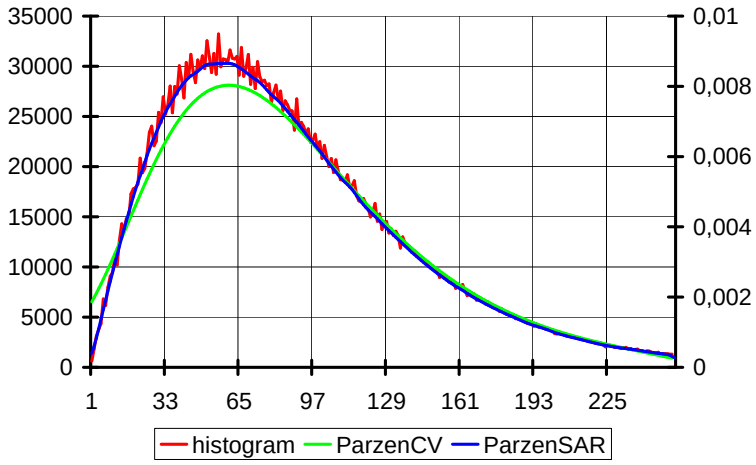

Figure C.7: Plot of the image histogram and of the Parzen PDF estimates for the "Bourges" data set.

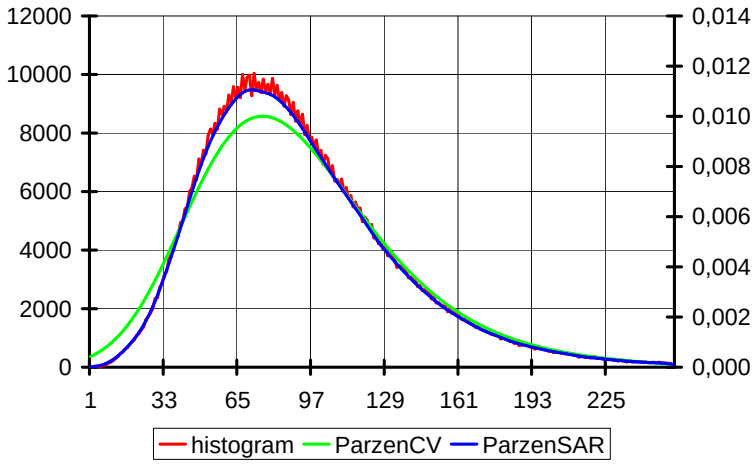

Figure C.8: Plot of the image histogram and of the Parzen PDF estimates for the "Flevoland" data set.

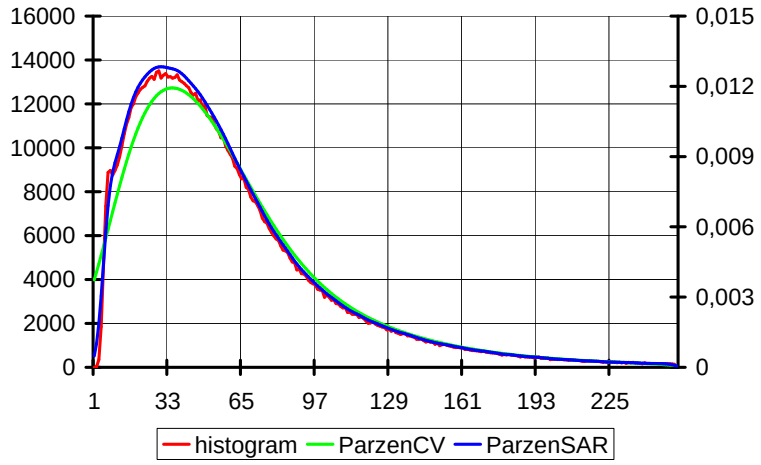

Figure C.9: Plot of the image histogram and of the Parzen PDF estimates for the "Suisse-Lake" data set.

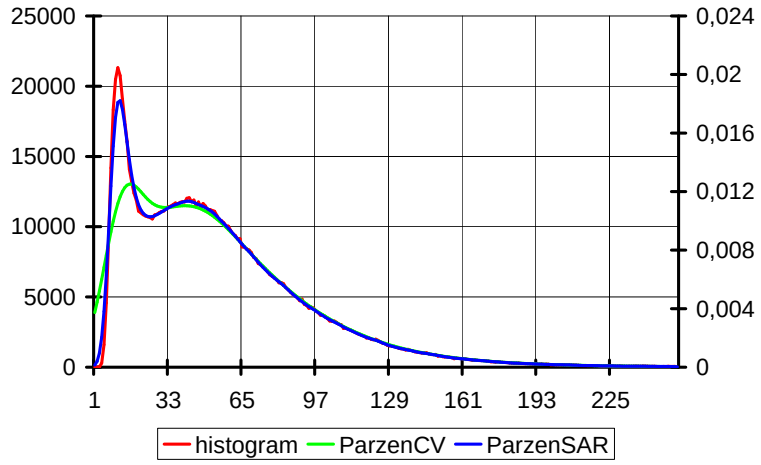

Figure C.10: Plot of the image histogram and of the Parzen PDF estimates for the "Suisse-Mountain" data set.

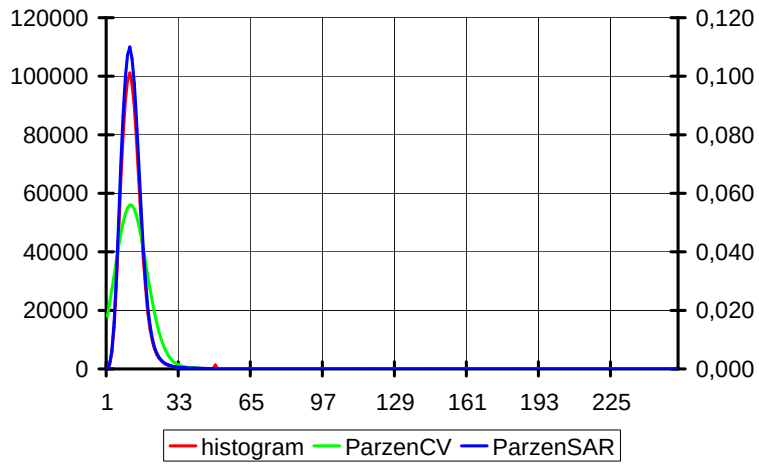

Figure C.11: Plot of the image histogram and of the Parzen PDF estimates for the "Oberpfaffenhofen" data set.

References

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

[2] A. Berlinet and L. Devroye, *A comparison of kernel density estimates*, Publications de l'Institut de Statistique de l'Université de Paris **38** (1994), no. 3, 3–59.

[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 Recognition Letters **20** (1999), 267–272.

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

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

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

[7] A. Bowman, *A comparative study of some kernel-based nonparametric density estimators*, J. Stat. Comp. Simul. **21** (1985), 313–327.

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

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

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

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

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

[13] M. Datcu, K. Seidel, and M. Walessa, *Spatial information retrieval from remote sensing images: Part I. information theoretical perspective*, IEEE Transactions on Geoscience and Remote Sensing **36** (1998), 1431–1445.

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

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

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

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

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

[19] G. Gilardi, *Analisi tre*, McGraw-Hill Italia, Milano (in italian), 1994.

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

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

[22] ———, *Significance of K distributions in scattering experiments*, Phys. Rev. Lett. **40** (1978), 546–550.

[23] K.-D. Kim and J.-H. Heo, *Comparative study of flood quantiles estimation by nonparametric models*, Journal of hydrology **260** (2002), 176–193.

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

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

[26] E. E. Kuruoglu, *Density parameter estimation of skewed alpha-stable distributions*, IEEE Transactions on Signal Processing **49** (2001), no. 10, 2192–2201.

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

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

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

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

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

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

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

[34] G. Moser, J. Zerubia, and S. B. Serpico, *SAR amplitude probability density function estimation based on a Generalized Gaussian scattering model*, Research report, INRIA, March 2004.

[35] M.Woodroofe, *On choosing a delta sequence*, Annals of Mathematical Statistics **41** (1970), 1665–1671.

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

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

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

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

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

[41] E. Parzen, *On estimation of probability density function and mode*, Signal Processing **33** (1962), 267–281.

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

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

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

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

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

[47] ———, *Real and complex analysis*, 3rd edition, McGraw-Hill, 1987.

[48] S. B. Serpico, L. Bruzzone, and F. Roli, *An experimental comparison of neural and statistical non-parametric algorithms for supervised classification of remote sensing images*, Pattern Recognition Letters **17** (1996), 1331–1341.

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

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

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

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

[53] V. N. Vapnik, *Statistical learning theory*, John Wiley and Sons Inc., 1998.

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

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

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

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

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

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