# A Textural-Contextual Model for Unsupervised Segmentation of Multi-Polarization Synthetic Aperture Radar Images

Vahid Akbari, *Student Member, IEEE*, Anthony P. Doulgeris, *Member, IEEE*, Gabriele Moser, *Member, IEEE*,

Torbjørn Eltoft, *Member, IEEE*, Stian N. Anfinsen, *Member, IEEE*, and Sebastiano B. Serpico, *Fellow, IEEE*

*Abstract*—This paper proposes a novel unsupervised, non-Gaussian, contextual segmentation method that combines an advanced statistical distribution with spatial contextual information for multi-look polarimetric SAR (PolSAR) data. This extends on previous studies that have shown the added value of both non-Gaussian modeling and contextual smoothing individually, or for intensity channels only. The method is based on a Markov random field (MRF) model that integrates a $\mathcal{K}$-Wishart distribution for the PolSAR data statistics conditioned to each image cluster and a Potts model for the spatial context. Specifically, the proposed algorithm is constructed based upon the stochastic expectation maximization (SEM) algorithm. A new formulation of SEM is developed to jointly perform clustering of the data, and parameter estimation of the $\mathcal{K}$-Wishart distribution and the MRF model. Experiments on simulated and real PolSAR data demonstrate the added value of using an appropriate statistical representation, in combination with contextual smoothing.

*Index Terms*—Polarimetric synthetic aperture radar (PolSAR); stochastic expectation-maximization (SEM); $\mathcal{K}$-Wishart distribution; Markov random field (MRF); unsupervised segmentation.

## I. INTRODUCTION

CLASSIFICATION techniques play an important role in automatic analysis of remote sensing data. This paper addresses image segmentation of polarimetric synthetic aperture SAR (PolSAR) data using an unsupervised classification approach. PolSAR data are multi-channel complex data sets that, when multi-looked, are given in terms of complex sample covariance matrices. This data representation poses challenges with regards to finding appropriate statistical models, as well as creating efficient classification strategies. Many studies have reported various methods with greater classification accuracies using polarimetric radar data instead of conventional single-polarization SAR data [1], [2].

Polarimetric SAR image segmentation is commonly performed with the Gaussian-based Wishart clustering algorithm operating on a pixel-by-pixel basis [3]. Analysis of PolSAR images often reveals that non-Gaussian models give better representation of the scattering vec-tor statistics, compared to complex Gaussian distributions, implying that processing algorithms based on non-Gaussian statistics should improve performance. The doubly stochastic product model has been widely used in non-Gaussian modeling, processing, and analysis of single- and multi-PolSAR images [4]. The model states that, under certain conditions, the complex-valued scattering vector results as the product of two independent random variables: a circular complex multinormal speckle noise component and a real scalar texture component. Several distributions could be used to model SAR image texture with different spatial correlation properties and various degrees of inhomogeneity [5], [6]. For the multi-look covariance matrix data, the product model produces models that deviate from the Wishart model. In the context of this paper, the term "non-Gaussian" will refer to models that deviate from the Wishart distribution. Statistical properties are widely used for feature extraction, image segmentation and land cover classification of PolSAR data, and several supervised and unsupervised classification schemes have been proposed in the literature [7]–[9]. In this study, we use the non-Gaussian $\mathcal{K}$-Wishart clustering algorithm [10], that accounts for potential textural differences in the classes, to represent the individual pixelwise statistical properties.

The potential of Markov random field (MRF) models to retrieve spatial contextual information is desired to improve the accuracy and reliability of the image clustering. Previous studies have shown the added values of both non-Gaussian modeling and contextual smoothing individually. This paper addresses the problem of unsupervised contextual polarimetric SAR image segmentation by combining advanced statistical modeling and spatial context within an MRF framework. MRF models have been used in remote sensing to address many image analysis problems, including (supervised and unsupervised) classification, segmentation, texture extraction, denoising, and change detection (see e.g., [8], [11]–[18]).

The proposed contextual clustering method uses a specific Markovian energy function for integrating the $\mathcal{K}$-Wishart distribution for the PolSAR data statistics conditioned to each image cluster and a Potts model for the spatial context. This algorithm thereby combines the benefits of a flexible non-Gaussian model for the covariance matrix data classes and an MRF for contextual smoothing. The algorithm is implemented as a stochastic expectation maximization (SEM) algorithm [19]. SEM is an iterative parameter estimation technique developed for parametric

V. Akbari, A. P. Doulgeris, T. Eltoft, and S. N. Anfinsen are with the Department of Physics and Technology, University of Tromsø, 9037 Tromsø, Norway (e-mail:vahid.akbari@uit.no;anthony.p.doulgeris @uit.notorbjorn.eltoft@uit.no;stian.normann.anfinsen@uit.no).

G. Moser and S. B. Serpico are with the Department of Telecommunications, Electronic, Electrical, and Naval Engineering, University of Genoa, 16145 Genoa, Italy (e-mail:gabriele.moser@unige.it; sebastiano.serpico@unige.it).

modeling problems characterized by data incompleteness. In each iteration, the current cluster parameter estimates are used to segment the image, and the new segments are used to recompute the cluster parameters. Here, a novel formulation of SEM is developed by formulating the maximum a posteriori (MAP) decision rule such that it jointly addresses data clustering and parameter estimation [20], [21].

The SEM is an improvement of the classical expectation maximization (EM) algorithm that incorporates a stochastic sampling procedure on every iteration. It is therefore more likely to avoid local maxima in the log-likelihood function during iterations. The SEM algorithm was used by *Moser et al.* [22] to estimate the distribution of single-polarization SAR data. In the context of multi-look PolSAR data, the EM and SEM algorithms have been applied by several authors to the problem of unsupervised segmentation. *Kersten et al.* [23] first applied the EM algorithm to this problem under the assumption of complex Wishart distributed data. *Reigber et al.* [14] augmented the Wishart mixture model, which was again resolved by the EM algorithm, with spatial context implemented in terms of relaxation labeling. *Doulgeris et al.* [10] used the SEM algorithm to solve a mixture model with $\mathcal{K}$-Wisharted-distributed components. The approach proposed in this paper uses SEM to address parameter estimation when the $\mathcal{K}$-Wishart distribution is combined with a spatial MRF model. Thus, it represents a new step in the evolution which joins an advanced model of pixel statistics with contextual information.

This paper is organized as follows. In Section II, we describe the pixelwise data format, followed by a description of the product model and the $\mathcal{K}$-Wishart distribution. In Section III, the general concepts of Markov random fields, the chosen MRF model, and the estimation method are discussed. In Section IV, the scheme of the textural-contextual classifier is explained. Section V demonstrates the performance of the method with simulated and real SAR data and discusses the results. Section VI is dedicated to a summary and conclusions.

## II. Pixelwise model

Assuming that the reciprocity principle is satisfied, the backscattering of a monostatic polarimetric SAR system is characterized by the complex scattering vector

$$\mathbf{k} = [S_{HH}, \sqrt{2}S_{HV}, S_{VV}]^T, \qquad (1)$$

where the elements represent the three complex backscattering coefficients in horizontal transmit horizontal receive (HH), horizontal transmit vertical receive (HV) and vertical transmit vertical receive (VV) polarization, respectively, and the superscript $T$ denotes the matrix transpose. Usually, polarimetric data are transformed into the form of multi-looked sample covariance matrices in order to reduce speckle noise, i.e.,

$$\mathbf{C} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{k}_l\mathbf{k}_l^H, \qquad (2)$$

where $L$ is the nominal number of looks used for averaging, and the superscript $H$ means complex conjugate transpose. Hence, after multi-looking, each pixel $s$ in the image is a realization of the $d \times d$ stochastic matrix variable denoted $\mathbf{C}$, and the image is referred to as the multi-look complex (MLC) covariance image.

### A. Multi-look Product Model

The non-Gaussian product model describes $\mathbf{C}$ as the product of a *texture* term and a *speckle* term [6]. Assuming that the texture has spatial correlation lengths larger than the local neighborhood size [10], the doubly stochastic product model for multi-looked PolSAR data is given by [4] as

$$\mathbf{C} = Z\mathbf{W}, \qquad (3)$$

where the strictly positive, unit mean scalar variable $Z$ models texture, and represents the backscatter variability due to heterogeneity of the radar cross section. The texture term is scalar because of the assumption of equal textural variations for all polarimetric channels. The second contribution, the speckle noise term $\mathbf{W}$, follows a *scaled complex Wishart* distribution [24], denoted $s\mathcal{W}(L, \mathbf{\Sigma})$, with parameters $L$, the nominal number of looks, and $\mathbf{\Sigma}$, the mean sample covariance matrix $E\{\mathbf{W}\}$. The probability density function (PDF) of $\mathbf{W}$ is given as

$$p_{\mathbf{W}}(\mathbf{W}) = \frac{L^{Ld}|\mathbf{W}|^{L-d}}{\Gamma_d(L)|\mathbf{\Sigma}|^L}\exp\left(-L\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{W})\right), \qquad (4)$$

where $tr(\cdot)$ and $|\cdot|$ denote the trace and determinant, respectively, and $\Gamma_d(L)$ is a normalization constant

$$\Gamma_d(L) = \pi^{\frac{d(d-1)}{2}}\prod_{i=1}^{d}\Gamma(L-i+1). \qquad (5)$$

named the multivariate gamma function in [24], while $\Gamma(\cdot)$ is the standard Euler gamma fuction.

The non-Gaussian nature of the product model depends on the specific model for the scalar texture variable $Z$ [5], [6], [10].

### B. $\mathcal{K}$-Wishart Distribution

Assume that the texture term of the product model follows the gamma distribution with PDF given by

$$p_Z(z; \alpha) = \alpha^\alpha \frac{z^{\alpha-1}}{\Gamma(\alpha)}\exp\left(-\alpha z\right). \qquad (6)$$

$\alpha > 0$ is called the shape parameter, and we impose unit mean, i.e., $E\{Z\} = 1$. Then the marginal distribution for $\mathbf{C}$ may be obtained by integrating the conditional PDF over the prior distribution of $Z$, that is

$$p_C(\mathbf{C}) = \int_0^\infty p_{C|Z}(\mathbf{C}|z)p_Z(z)dz, \qquad \mathbf{C}|z \sim s\mathcal{W}(L, z\mathbf{\Sigma}). \qquad (7)$$

The resulting distribution for $\mathbf{C}$ was called the $\mathcal{K}-$ distribution in the seminal paper [5]. We refer to it as the $\mathcal{K}$-Wishart distribution (as did [10]), denoted $\mathcal{K}\mathcal{W}(L, \alpha, \mathbf{\Sigma})$,

to distinguish the matrix-variate form from other members of the $\mathcal{K}$ distribution family. It is given in closed form as

$$
\begin{aligned}
p_{\mathbf{C}}(\mathbf{C}) = {} & \frac{2|\mathbf{C}|^{L-d}(L\alpha)^{\frac{\alpha+Ld}{2}}}{\Gamma_d(L)\Gamma(\alpha)|\mathbf{\Sigma}|^L}\left(\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})\right)^{\frac{\alpha-Ld}{2}} \\
& \times K_{\alpha-Ld}\left(2\sqrt{L\alpha\,\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})}\right),
\end{aligned}
\tag{8}
$$

where $K_\rho(\cdot)$ is the modified Bessel function of the second kind with order $\rho$. Note that the PDF is parameterized by the shape parameter $\alpha$, the number of looks $L$, and the scale matrix $\mathbf{\Sigma}$. As $\alpha \to \infty$, the texture variable tends to a constant, and the $\mathcal{K}$-Wishart model converges in distribution to the standard Wishart model. The flexibility of the $\mathcal{K}$-Wishart model is indicated in Fig. 1, by showing the PDF of intensity variations for increasing values of $\alpha$.
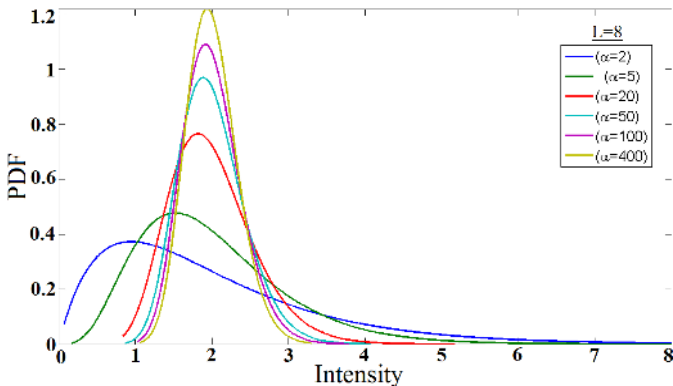


Fig. 1. Non-Gaussian flexibility of $\mathcal{K}$-Wishart model with different shape parameter values. Lower values of $\alpha$ represent extreme texture and higher values represent low texture with the $\mathcal{KW}$ asymptotically tending toward the Wishart distribution. Therefore, the flexible $\mathcal{KW}$ distribution better captures real data variations. Plotted is the 1-dimensional representation of intensity variations.

### C. Parameter Estimation of the $\mathcal{K}$-Wishart Model

The shape parameter $\alpha$ in the $\mathcal{K}$-Wishart PDF is estimated with the method of matrix log-cumulants (MoMLC) [24]. This method is founded on Mellin transform theory [25], [27], has been proved to be a feasible and effective estimation technique associated with the multi-look polarimetric product model. In [24] it was shown that the $\nu^{\mathrm{th}}$ order log-cumulant equations for $\mathbf{C}$ separate the texture and speckle contributions under the multi-look product model, such that

$$
\kappa_\nu\{\mathbf{C}\} = \kappa_\nu\{\mathbf{W}\} + d^\nu \kappa_\nu\{Z\}
\tag{9}
$$

The matrix log-cumulant equations of the $\mathcal{K}$-Wishart distribution are given as

$$
\begin{aligned}
\kappa_1\{\mathbf{C}\} &= \ln|\mathbf{\Sigma}| + \psi_d^0(L) + d\left(\psi^0(\alpha) - \ln(\alpha L)\right), \\
\kappa_\nu\{\mathbf{C}\} &= \psi_d^{\nu-1}(L) + d^\nu \psi^{\nu-1}(\alpha), \qquad \nu > 1,
\end{aligned}
\tag{10}
$$

where $\psi_d^\nu(\cdot)$ and $\psi^\nu(\cdot)$ are the $\nu^{\mathrm{th}}$ order multivariate and the ordinary polygamma functions, respectively [24], and the relation between log-moments and log-cumulants is

given by

$$
\kappa_\nu\{\cdot\} = \mu_\nu\{\cdot\} - \sum_{i=1}^{\nu-1}\binom{\nu-1}{i-1}\kappa_i\{\cdot\}\mu_{\nu-i}\{\cdot\}.
\tag{11}
$$

where $\binom{\cdot}{\cdot}$ represents the binomial coefficient. Several sample matrix log-cumulants $\langle\kappa_\nu\rangle$ can hence be obtained from the sample matrix log-moments, which are calculated from the data as

$$
\langle\mu_\nu\rangle = \frac{1}{n}\sum_{i=1}^n(\log|\mathbf{C}_i|)^\nu.
\tag{12}
$$

The weighted least squares method is subsequently used to estimate the K-Wishart parameters from an overdetermined system of matrix log-cumulant equations. The method is explained in detail in [28], where it is referred to as the maximum asymptotic likelihood (MAL) estimator.
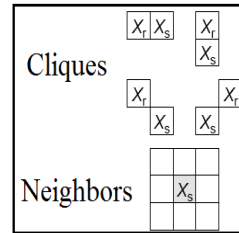


Fig. 2. Second-order neighborhood system and pairwise cliques.

### III. MARKOV RANDOM FIELD (MRF) MODEL

Let $S = \{s_{i,j}; 1 \le i \le R, 1 \le j \le M\}$ be regarded as a 2-D pixel lattice, where $s_{i,j}$ is site $(i,j)$, $R$ and $M$ are the number of rows and columns of the image, respectively, and let $\mathcal{L} = \{1, 2, ..., J\}$ denote the set of all possible labels in the clustering map. A label random field $\mathbf{X} = \{X_s; X_s \in \mathcal{L}, s \in S\}$ defined on $S$ is an MRF with respect to a given neighborhood system if, and only if, the following two conditions hold:
1) The positivity condition: $P(\mathbf{X}) > 0$,
2) Markovianity condition: $P(X_s|\mathbf{X}_{S\setminus s}) = P(X_s|\mathbf{X}_{\mathcal{N}(s)})$, where $S\setminus s$ denotes the set containing all sites in $S$ except $s$, and $\mathcal{N}(s)$ is the neighborhood of $s$ [20], [29].
According to the Hammersley-Clifford theorem [30], the joint probability distribution of an MRF is a Gibbs distribution [29], which enables the field to be characterized by its local, instead of its global, properties. Therefore, the PDF of $\mathbf{X}$ has the form

$$
P(\mathbf{X}) = W^{-1}\exp\left(-U(\mathbf{X})\right),
\tag{13}
$$

where $U$ is an *energy function*

$$
U(\mathbf{X}) = \sum_{c\in\mathcal{C}}V_c(\mathbf{X}_c).
\tag{14}
$$

$c$ is a clique and $\mathcal{C}$ is the collection of all cliques with respect to the adopted neighborhood system, $\mathbf{X}_c$ is the set of samples $\mathbf{X}_s$ such that $s$ belongs to $c$, and $V_c$ is the potential associated with clique $c$. $W = \sum_{\mathbf{X}}\exp\left(-U(\mathbf{X})\right)$ is a normalization constant called the *partition function* [20].

Consequently, the local spatial correlations of $X_s$ with respect to its neighbors are modeled by defining suitable potential functions $V_c$. This actually is an essential problem in all MRF models. For simplicity of computations, the current work confines the neighborhood system to an isotropic second-order system, with the related set of pairwise cliques depicted in Fig 2.

### A. Potts MRF Model

We will use the classical homogeneous Potts model to model the spatial correlation between pixels. According to this model, a single global parameter $\beta > 0$, which is known as the spatial interaction parameter, regulates the pairwise pixel interactions.[1] With cliques consisting of up to two sites, the second-order energy function of the homogeneous MRF model is given as [20], [29]

$$U(\mathbf{X}) = \sum_{s \in S} V_1(X_s) + \beta \sum_{s \in S} \sum_{r \in \mathcal{N}(s)} V_2(X_s, X_r), \quad (15)$$

where $X_s$ is the label of the central pixel $s$, $X_r$ is the label of a neighboring pixel in the neighborhood system of $s$. The first term of the energy function, $\sum_{s \in S} V_1(X_s)$, is determined by the pixelwise PDF, and the second term conveys the contextual information. For a single pixel, $s$, the contextual energy term is the sum over the second-order cliques of the neighborhood, i.e.,

$$U(X_s|\mathbf{X}_{\mathcal{N}(s)}) = \beta \sum_{r \in \mathcal{N}(s)} V_2(X_s, X_r), \quad (16)$$

which for the global Potts model is given as

$$V_2(X_s, X_r) = -\delta(X_s, X_r) = \begin{cases} -1 & \text{if } X_s = X_r \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

This potential function results in the following conditional probability mass function (PMF) of $X_s$, given its neighbors:

$$
\begin{aligned}
P(X_s|\mathbf{X}_{\mathcal{N}(s)}; \beta) &= \frac{\exp\left(-\beta \sum_{r \in \mathcal{N}(s)} V_2(X_s, X_r)\right)}{\sum_{X_s \in \mathcal{L}} \exp\left(-\beta \sum_{r \in \mathcal{N}(s)} V_2(X_s, X_r)\right)} \\
&= \frac{\exp(\beta m_{X_s}(s))}{\sum_{l \in \mathcal{L}} \exp(\beta m_l(s))},
\end{aligned}
\quad (18)
$$

where $m_{X_s}(s)$ is the number of neighbors of pixel $s$ with label equal to $X_s$.

### B. MRF Parameter Estimation

The main difficulty of MRF parameter estimation is that the maximum likelihood (ML) method is computationally intractable for most MRF models. An alternative is to adopt the maximum pseudo-likelihood (MPL) technique, which is computationally feasible and simple to implement.

[1] Spatially inhomogeneous MRF models allow the $\beta$-parameter to change over the field, in which case the models are defined by a set of local conditional density functions [31].

The pseudo-likelihood (PL) approach consists in approximating the likelihood in Eq. (13) as follows [32], [33]:

$$PL(\mathbf{X}; \beta) = \prod_{s \in S} P(X_s|\mathbf{X}_{\mathcal{N}(s)}; \beta), \quad (19)$$

By substituting the local conditional probabilities from Eq. (18) into (19), an approximation of $P(\mathbf{X})$ in Eq. (13) is obtained by

$$
\begin{aligned}
PL(\mathbf{X}; \beta) &= \prod_{s \in S} P(X_s|\mathbf{X}_{\mathcal{N}(s)}; \beta) \\
&= \prod_{s \in S} \frac{\exp(\beta m_{X_s}(s))}{\sum_{l \in \mathcal{L}} \exp(\beta m_l(s))},
\end{aligned}
\quad (20)
$$

Taking the logarithm, the above equation leads to the maximization of

$$\Phi(\beta) = \sum_{s \in S} \left[ \beta m_{X_s}(s) - \log\left(\sum_{l \in \mathcal{L}} \exp(\beta m_l(s))\right) \right]. \quad (21)$$

This function is optimized by a simulated annealing algorithm to estimate the MRF parameter $\beta$, a computationally intensive global minimization approach [33], [34].

### IV. Textural – Contextual Classifier

Let $\mathcal{C} = \{\mathbf{C}_s; s \in S\}$ be an MLC image, and let $\mathbf{X} = \{X_s; X_s \in \mathcal{L}, s \in S\}$ be the class labels of $\mathcal{C}$. The unobserved class labels, $\mathbf{X}$, are now considered as a discrete Potts MRF with the energy function $U$ depending on the parameter $\beta$ [29]. The MLC image is statistically modeled as a mixture of $\mathcal{K}$-Wishart distributions, where the pixels are assumed to be conditionally independent given the label field $\mathbf{X}$, i.e.,

$$p_{\mathcal{C}|\mathbf{X}}(\mathcal{C}|\mathbf{X}; \boldsymbol{\theta}) = \prod_{s \in S} p_{\mathbf{C}_s|X_s}(\mathbf{C}_s|X_s; L, \boldsymbol{\Sigma}, \alpha), \quad (22)$$

and $\boldsymbol{\theta} = \{L, \alpha, \boldsymbol{\Sigma}\}$ is the vector of the parameters of the $\mathcal{K}$-Wishart distribution. In order to develop the $\mathcal{K}$-Wishart MRF classifier using the MAP criterion, a feasible and computationally affordable approach is the iterated conditional mode (ICM) technique for MRF energy minimization [20], [35]. Accordingly, we need to obtain the conditional probability mass function of the label random field $\mathbf{X}$ given an observation of the MLC image $\mathcal{C}$. Pixelwise, this conditional PMF can be formulated using the Bayes rule as

$$P_{X_s|\mathbf{C}_s}(X_s|\mathbf{C}_s) \propto p_{\mathbf{C}_s|X_s}(\mathbf{C}_s|X_s; \boldsymbol{\theta}) P_{X_s}(X_s), \quad (23)$$

where $P_{X_s}$ is the prior PMF and $P_{X_s|\mathbf{C}_s}$ is the posterior PMF of the class label, i.e., the latter is the probability that pixel $s$ belongs to class $X_s$, given the observation $\mathbf{C}_s$. In the homogeneous Potts MRF model $X_s$ is then estimated by

$$
\begin{aligned}
\hat{X}_s &= \arg\max_{X_s \in \mathcal{L}} \{ P_{X_s|\mathbf{C}_s}(X_s|\mathbf{C}_s, \mathbf{X}_{\mathcal{N}(s)}; \boldsymbol{\Theta}) \} \\
&= \arg\max_{X_s \in \mathcal{L}} \{ p_{\mathbf{C}_s|X_s}(\mathbf{C}_s|X_s; \boldsymbol{\theta}) P_{X_s|\mathbf{X}_{\mathcal{N}(s)}}(X_s|\mathbf{X}_{\mathcal{N}(s)}; \beta) \},
\end{aligned}
\quad (24)
$$

and $\mathbf{\Theta} = (\boldsymbol{\theta}, \beta)$ has to be estimated in the iterative segmentation process. The posterior probability of $X_s$ given $\mathbf{C}_s$ becomes

$$P(X_s|\mathbf{C}_s, \mathbf{X}_{\mathcal{N}(s)}; \mathbf{\Theta}) = \frac{\exp\left(-U(X_s|\mathbf{C}_s, \mathbf{X}_{\mathcal{N}(s)}; \mathbf{\Theta})\right)}{\sum_{X_s \in \mathcal{L}} \exp\left(-U(X_s|\mathbf{C}_s, \mathbf{X}_{\mathcal{N}(s)}; \mathbf{\Theta})\right)}, \quad (25)$$

and the associated local posterior energy function [35]

$$
\begin{aligned}
U(X_s|\mathbf{C}_s, \mathbf{X}_{\mathcal{N}(s)}; \mathbf{\Theta}) &= U(X_s|\mathbf{X}_{\mathcal{N}(s)}; \beta) + U(\mathbf{C}_s|X_s; \boldsymbol{\theta}) \\
&= -\beta m_{X_s}(s) + \log\left(\sum_{l \in \mathcal{L}} \exp(\beta m_l(s))\right) \\
&\quad - \log p_{\mathbf{C}_s|X_s}(\mathbf{C}_s|X_s; \boldsymbol{\theta}).
\end{aligned}
\quad (26)
$$

## A. SEM Clustering Process

The proposed contextual MRF-based classifier consists of two processing stages: a non-contextual stage and an MRF stage.

**Non-contextual stage**: In this stage, an initial non-contextual segmentation map is to be estimated. We assume that the MLC image is a realization drawn from a mixture of $\mathcal{K}$-Wishart distributions, i.e., at each site the probability density function of the sample covariance matrix is given as

$$p_{\mathbf{C}_s}(\mathbf{C}_s) = \sum_{l \in \mathcal{L}} \pi_l p_l(\mathbf{C}_s; \boldsymbol{\theta}_l), \quad (27)$$

where the non-negative mixture proportions $\pi_l$ satisfy the relation $\sum_{l \in \mathcal{L}} \pi_l = 1$, $p_l(\mathbf{C}_s; \boldsymbol{\theta}_l)$ are densities for each individual class $l$, and $\boldsymbol{\theta}_l$ is the parameter vector of the $l^{\text{th}}$ $\mathcal{K}$-Wishart mixture component. Since both the class labels and the parameters are unknown and must be estimated from the observations, we have a so-called *incomplete* data problem [29], which can be solved by applying the SEM algorithm. For a given number of $J$ components, the task is to estimate the vector of parameters $\mathbf{\Theta}_{NC} = (\pi_1, \pi_2, \cdots, \pi_J, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_J)$ that maximizes the log-likelihood function

$$LL(\mathbf{\Theta}_{NC}) = \sum_{s \in S} \left(\log\left(\sum_{l \in \mathcal{L}} \pi_l p_l(\mathbf{C}_s; \boldsymbol{\theta}_l)\right)\right). \quad (28)$$

The SEM algorithm starts with an initial segmentation with $J$ classes and parameter vector $\mathbf{\Theta}_{NC}^{(0)}$, and works in an iterative manner where, in each iteration, the current cluster parameter estimates are used to segment the image, and the new segments are used to recompute the cluster parameters. The SEM algorithm has three steps [19]:

1. An *expectation (E) step*, in which, for each given observation, the posterior probabilities associated with the clusters are estimated using the current class-conditional PDF components and mixture proportions.

2. A *stochastic (S) step*, which randomly samples the label of each sample covariance matrix according to the current estimated posterior PMF from the *E-step*.

3. A *maximization (M) step*, which updates all class parameters, i.e., the $\mathcal{K}$-Wishart parameters using the log-cumulant estimators discussed in Section II-C, and $\pi_l$ using the relative frequency of assignment to each class.

The SEM algorithm does not stop by convergence and has to be terminated by simply setting a fixed number of iterations, for example 200. The above-mentioned steps are carried out in each iteration to produce a homogeneous Markov chain of parameter estimates, and the estimate that maximizes the log-likelihood function $LL(\mathbf{\Theta}_{NC})$ is selected. According to [19] this procedure should result in a final solution close to the global maximum likelihood estimate. The algorithm results in a final segmentation and associated model parameters for each class.

**MRF stage**: The pixelwise model is now combined with the MRF model to take into account the spatial correlation between class labels. The classifier based on pixel statistics only is severely affected by overlapping class statistics due to speckle noise. This problem is reduced by incorporating spatial contextual information. The SEM algorithm is modified to include the contextual energy in the calculation of the posterior probabilities, and also the spatial interaction parameter of the Potts MRF model, $\beta$, is estimated in the iteration process. The corresponding log-likelihood function becomes

$$LL(\mathbf{\Theta}_{MRF}) = \sum_{s \in S} \left(\log\left(\sum_{l \in \mathcal{L}} \pi_l(s) p_l(\mathbf{C}_s; \boldsymbol{\theta}_l)\right)\right), \quad (29)$$

where $\mathbf{\Theta}_{MRF} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_J, \beta)$, and the mixture proportions $\pi_l$ have been replaced by the local priors $\pi_l(s)$, defined as the conditional probabilities in (18). We specifically note that $\pi_l(s)$ depends on the site $s$ through the labels of its entire neighborhood $\mathcal{N}(s)$. Starting from an initial parameter vector $\mathbf{\Theta}_{MRF}^{(0)}$, estimated as the maximum of the log-likelihood function of the non-contextual stage, an iteration of the MRF stage SEM algorithm consists of three steps:

1. An *E-step*, where the posterior probabilities associated with the clusters are estimated for each given observation, using the current component PDFs and local priors (see Eq. (25)).

2. An *S-step*, which randomly samples the label of each sample covariance matrix according to the current estimated posterior PMF.

3. An *M-step*, which updates all class parameters, i.e., the $\mathcal{K}$-Wishart parameters and the parameter of the Potts MRF model, $\beta$, discussed in Section III-B.

We use the mode-field approach to approximate the likelihood function, which makes the SEM algorithm computationally affordable [29]. Contextual SEM continues after pixelwise SEM on the segmentation result produced in the non-contextual stage. The same method is used to

terminate the contextual SEM as discussed above for the non-contextual stage. Due to the added influence of spatial contextual information among neighboring pixels, we terminate the algorithm with less number of iterations, for example 15.

### B. Number of Classes

One of the key issues in unsupervised image segmentation is to determine the appropriate number of segments, $J$. In most algorithms, $J$ is assumed to be known. In our case, the number of clusters is automatically determined in a preclustering process in the way proposed by *Doulgeris et al.* [7]. According to this method, the number of mixture classes is dynamically determined in a goodness-of-fit stage, which regularly tests how well each class histogram matches a theoretically predicted PDF model. The algorithm will keep clusters with a good match unaltered; clearly mixed, multi-modal clusters will be split into two clusters, whereas virtually identical competing clusters will be merged. Note that the non-contextual SEM stage could be replaced by the automated clustering algorithm proposed by *Doulgeris et al.* However, we do not do this in order to have consistent comparison of SEM algorithm with and without MRF. We therefore need to fix both the number of clusters and the underlying SEM algorithm in both processing steps.

### C. Initialization

The non-contextual SEM clustering process can be initialized in two ways.

*K-means initialization*: *K*-means clustering is applied to a vector image consisting of the logarithmic intensities of the individual polarimetric channels. Then, the original MLC image is given the same partitioning. The MoMLC technique is subsequently used to estimate the parameters of each partition.

*Random initialization*: In this case the image is randomly partitioned into a given number of segments. In the MRF stage, the Potts MRF model is initialized with $\beta = 1$, and the partition result from the non-contextual stage.

### D. Effective number of looks (ENL) estimation

One of the parameters in the pixelwise PDF is the number of independent looks, $L$. Because there is some correlation between pixels in real images, this number is different from the nominal number of looks used in the multi-looking. Hence, the nominal number of looks needs to be replaced with an estimated *effective number of looks* (ENL). We incorporate the ENL estimation technique from [36], where the ENL is estimated using MoMLC in a pre-analysis of the image.

### V. Experimental Results and Analysis

In this section, we analyze the performance of the proposed algorithm on both simulated and real polarimetric SAR data. It is noted that even though the theory review in section II was carried out for the full polarimetric case, the algorithm can also be applied to dual and single

polarimetric data. Our study includes the following four segmentation experiments:

i) A pixelwise standard Wishart model, i.e., the non-textured case, denoted $\mathcal{W}$.

ii) A pixelwise standard Wishart model and an MRF model for spatial context, i.e., the non-textured contextual case, denoted $\mathcal{W}$-MRF.

iii) A pixelwise $\mathcal{K}$–Wishart model, i.e., the textured case, denoted $\mathcal{KW}$.

iv) A pixelwise $\mathcal{K}$-Wishart model and an MRF model for spatial context, i.e., the textured contextual case, denoted $\mathcal{KW}$-MRF.

The data sets include a simulated dual-pol image, a quad-pol airborne scene over an agricultural area in Foulum, Denmark, and an ENVISAT ASAR scene over the Kongsvegen glacier on Svalbard. The simulated data allows a quantitative analysis of the classification accuracies, whereas the real data only will permit a visual comparison due to insufficient ground truth data.

### A. Simulated SAR Image with Seven Classes

The simulated image is $250 \times 250$ pixels in size, it is dual-pol, and generated with 8-look, $\mathcal{K}$-Wishart distributed matrix data. The covariance matrix and texture parameters are taken from a real dataset to simulate classes with properties of a real image. Fig. 3(a) shows the Quasi-Pauli image (dual-pol) of the simulated test pattern and intensity variations of each simulated class is plotted with its shape parameter in Fig. 3(b). Highly skewed curves indicate high texture regions and less skewed curves represent low texture regions. Class 5 (pink curve) in Fig. 3(b) clearly has extreme skewness compared to the others. This class was simulated with the statistical properties of an urban area. Classes 2, 3 and 4 come from a forest area with moderate texture. Class 6 (cyan curve) is representative of a homogeneous area and has the lowest texture. As seen in Fig. 3(b), the cyan PDF is the narrowest and the standard Wishart distribution can model such a class for the given number of looks. Other classes are simulated with the statistics of agricultural crops and vegetation areas.

To initialize the SEM process, each observation is assigned randomly to one of seven classes. As the generated samples are independent, the ENL is equal to the nominal number of looks. Then, the first stage of the classification is implemented for each of the textured ($\mathcal{KW}$) and non-textured ($\mathcal{W}$) models. Figs. 3(c) and 3(d) show the results for the $\mathcal{W}$ and $\mathcal{KW}$ MAP classifiers on the same $\mathcal{K}$-Wishart test image. Both classifiers already provide good classification performances in the non-contextual step, with accuracies higher than 80% for most classes and overall accuracy of approximately 89% and 97%, respectively. The biggest difference in performance is observed for the highly non-Gaussian urban class (class 5), which has a classification accuracy of 49.45% for the $\mathcal{W}$ and 84.8% for the $\mathcal{KW}$. Since the $\mathcal{W}$ model has limited variance, it has difficulties with grouping the highly non-Gaussian classes into single clusters, and therefore it fits two or more Gaussian-constrained
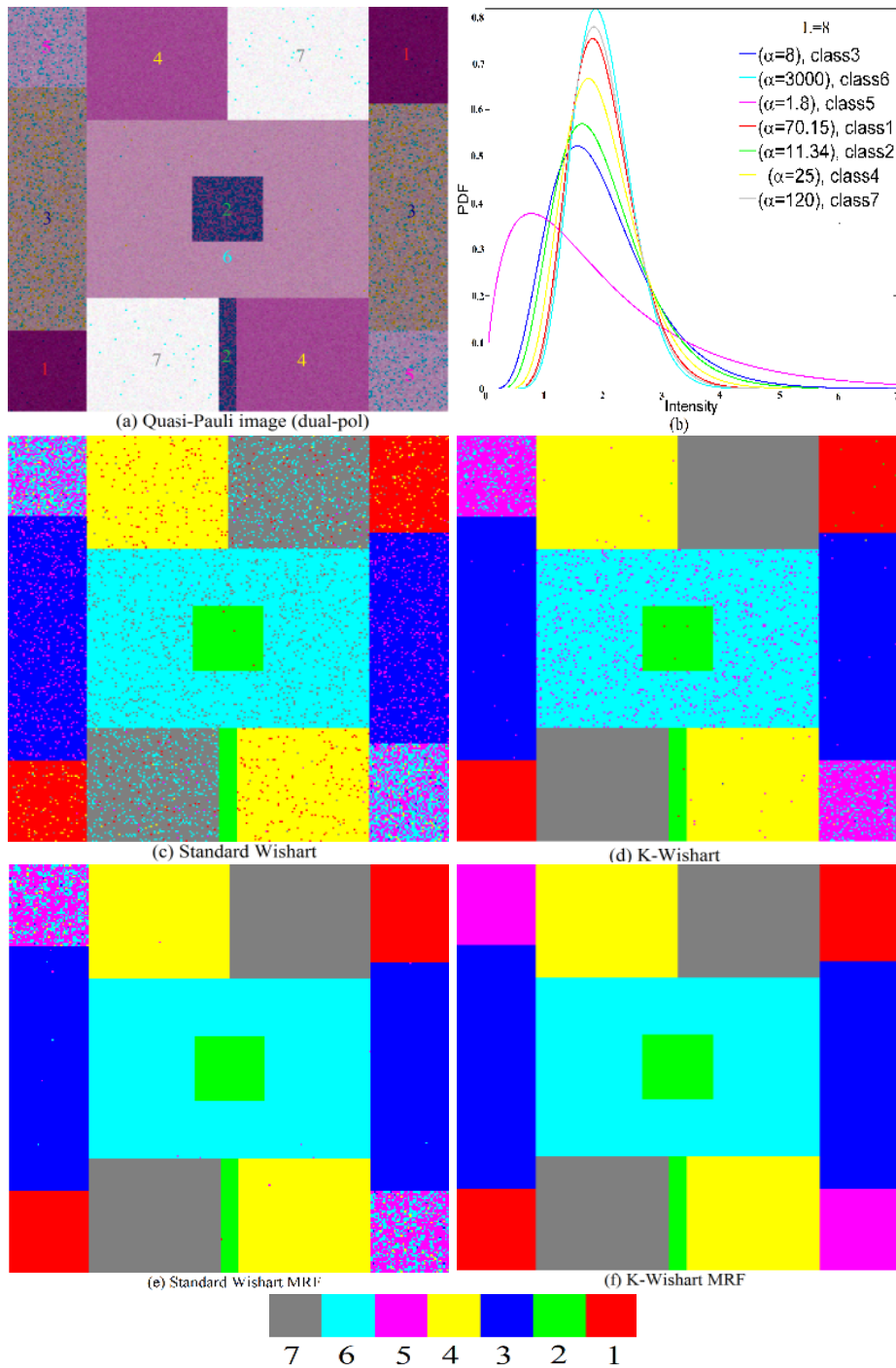
Fig. 3. Experiment with 8-look, dual-pol, $\mathcal{K}$-Wishart distributed simulated pattern: (a) Quasi-Pauli (dual-pol) RGB image with class labels; and (b) textured model intensity curves for each simulated class indicating the texture variation among the classes; (c) Standard Wishart ($\mathcal{W}$); (d) $\mathcal{K}$-Wishart ($\mathcal{KW}$); (e) Standard Wishart MRF ($\mathcal{W}$-MRF), and (f) $\mathcal{K}$-Wishart MRF ($\mathcal{KW}$-MRF) segmentation results.

TABLE I

CLASSIFICATION ACCURACIES FOR EXPERIMENT 1, INCLUDING OVERALL ACCURACY (OA), THE KAPPA COEFFICIENT AND ITS VARIANCE.

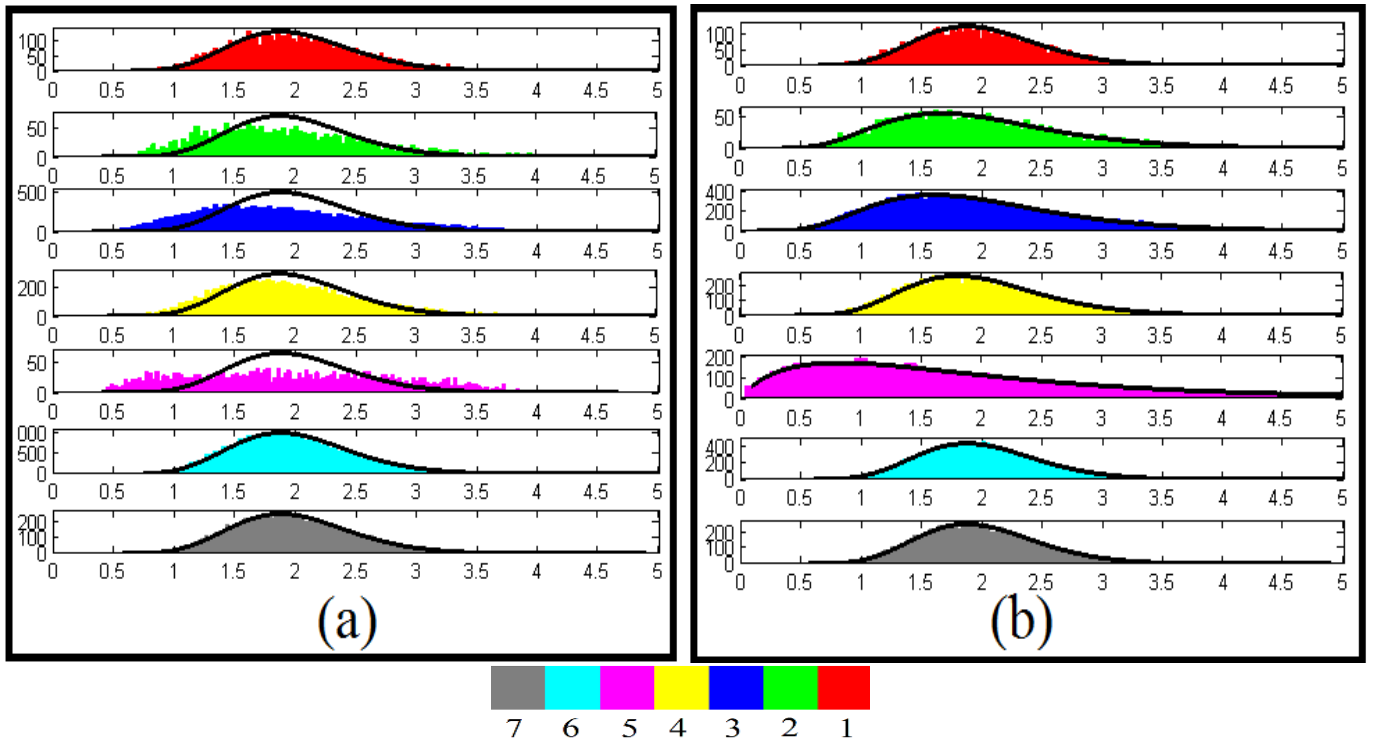| Classifier | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 | Class 6 | Class 7 | OA | $\hat{K}$ | $\hat{\sigma}_K^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mathcal{W}$ | 94.81% | 99.87% | 90.20% | 96.92% | 49.45% | 93.09% | 87.30% | 89.46% | 0.8893 | 4.33e-6 |
| $\mathcal{KW}$ | 99.83% | 99.74% | 99.85% | 99.77% | 84.80% | 93.89% | 100% | 97.25% | 0.9725 | 9.93e-7 |
| $\mathcal{W}$- MRF | 100 % | 99.96% | 99.94% | 99.98% | 66.78% | 99.97% | 100% | 97.58% | 0.9759 | 9.88e-7 |
| $\mathcal{KW}$- MRF | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 1.0000 | 3.89e-7 |

Fig. 4. Comparison between fitted models and class histograms of trace($\mathbf{\Sigma}^{-1}\mathbf{C}$) for all clusters found in the simulated test pattern, (a) The left figure shows that the $\mathcal{W}$-MRF model is not fitted well to the histograms of some clusters. (b) The right figure shows the good visual fit of the $\mathcal{KW}$-MRF model to all class histograms.

distributions to the non-Gaussian data classes [10]. The $\mathcal{KW}$ can better model such intensity variations in one class, yielding significant improvement and a 35.35% increase in the accuracy for the urban class. Table I summarizes a quantitative evaluation of the classification performance, including the classification accuracies for all classes, overall accuracy, the Kappa coefficient, and its asymptotic variance [37] for the proposed models. Even though the considered operational setting is unsupervised, such a quantitative accuracy analysis is feasible for this dataset, since the true label of each pixel is known. As expected, classification results appear noisy because the contextual information is disregarded by the pixelwise SEM technique.

The contextual $\mathcal{W}$-MRF and $\mathcal{KW}$-MRF classifiers improve the results, yielding nearly perfect and perfect segmentation of the seven classes, respectively. Figs. 3(e) and 3(f) demonstrate the results. The contextual stage results in a strong smoothing of homogeneous areas, while class boundaries are preserved. As unsupervised algorithms do not produce a unique class numbering, the labels were adjusted to make all results comparable, before computing the confusion matrices for all classifiers. Figs. 4(a) and 4(b) depict class histograms and fitted model PDFs of the entity trace($\mathbf{\Sigma}^{-1}\mathbf{C}$) for all segments produced by the $\mathcal{W}$-MRF and $\mathcal{KW}$-MRF classifiers, respectively. The transformation trace($\mathbf{\Sigma}^{-1}\mathbf{C}$) compacts the matrix-variate data into a scalar such that the model goodness-of-fit can be visualized. The theoretical models for trace($\mathbf{\Sigma}^{-1}\mathbf{C}$) are derivable for the respective models for $\mathbf{C}$, because the quantity is a linear combination of $d$ whitened terms and each

model possess infinite divisibility [38]. Therefore, Wishart data leads to a Wishart model, $\mathcal{W}(Ld, d)$ for this quantity, which is equivalent to the one dimensional gamma distribution. Similarly, $\mathcal{KW}$ data leads to a $\mathcal{KW}(Ld, \alpha, d)$, which is equivalent to the intensity $\mathcal{K}$-distribution.

For *low texture and homogeneous regions* (class 1, 6 and 7), both the $\mathcal{W}$-MRF and the $\mathcal{KW}$-MRF models produce partitions that excellently match the original simulated image. The model fit of the class histograms is also good. For *moderate texture areas* (classes 2, 3 and 4), even though the $\mathcal{W}$-MRF provides excellent classification accuracies, the model fit is not good in Fig. 4(a). Most notably, the $\mathcal{W}$-MRF has problems with the urban class (class 5). The MRF model cannot fully compensate for the weakness of the pixelwise model, which does not capture the large variance of this *highly textured class*. Still, the contextual smoothing improves the accuracy from 49.45% for the $\mathcal{W}$ classifier to 66.78% for the $\mathcal{W}$-MRF classifier. The model is a bad fit for several class histograms. Conversely, the $\mathcal{KW}$-MRF model provides a good visual fit to all classes.

The estimated Kappa coefficient, $\hat{K}$, and asymptotic variance, $\hat{\sigma}_K^2$, are given in Table I for each classification. We can evaluate significant differences between any pair of classification results by a test statistic $\Delta\hat{K}$ given in [39] as

$$\Delta\hat{K} = \frac{\left|\hat{K}_2 - \hat{K}_1\right|}{\sqrt{\hat{\sigma}_{K_2}^2 + \hat{\sigma}_{K_1}^2}}. \tag{30}$$

At the 95% confidence level two results may be considered

significantly different if $\Delta\widehat{K}$ exceeds 1.96. By computing $\Delta\hat{K}$ for each pair of classifiers, as seen in Table II, we found that *all* classification results are significantly different in a statistical sense. This confirms that adding texture, adding context, or adding both, significantly improves the segmentation. As seen in Table I, the $\mathcal{KW}$ and $\mathcal{W}$-MRF classifiers produce similar accuracies, but this is almost certainly only a coincidence for this particular dataset. The relative significance between texture and context is irrelevant since our proposal exploits both simultaneously.

TABLE II

$\Delta\hat{K}$ values for Test of Significant Differences between pairs of classifiers ($\Delta\hat{K}{>}1.96$ is significant)

| | $\mathcal{W}$ | $\mathcal{KW}$ | $\mathcal{W}$- MRF | $\mathcal{KW}$- MRF |
|---|---|---|---|---|
| $\mathcal{W}$ | 0.00 | 36.07 | 37.53 | 50.98 |
| $\mathcal{KW}$ | | 0.00 | 2.376 | 23.413 |
| $\mathcal{W}$- MRF | | | 0.00 | 20.604 |
| $\mathcal{KW}$- MRF | | | | 0.00 |

### B. Foulum Example

In the second experiment, a small section of the image of an agricultural area from an airborne EMISAR L-band quad-pol acquisition over Foulum, Denmark, from 1998 is tested. An enhanced Pauli composite image (R=HH-VV, G=HV, B=HH+VV) is shown in Fig. 5. The SLC data is multi-looked to 8-looks. The ENL value was estimated to 6.4 in a pre-analysis of the image (as discussed in Section IV-D). The appropriate number of classes for this real dataset is automatically determined in advance by the goodness-of-fit testing method proposed by *Doulgeris et al.* [7]. As expected theoretically, this gives a different number of classes depending on the constraints of the chosen model. The $\mathcal{KW}$ classifier found 14 classes for this data. However, to more easily compare results, the number of classes is fixed to 14 the other classifiers. As seen from Figs. 6(a) and 6(b), the $\mathcal{W}$ classifier splits fields A and B into two classes (red and dark blue), whereas the $\mathcal{KW}$ classifier groups them as a single class (red). This happens also for field C, where the brown, light green, and yellow classes in the $\mathcal{W}$ result correspond to the (mostly) brown and yellow class in the $\mathcal{KW}$ result. For forest areas (field D and E), the $\mathcal{KW}$ classifier seems to distinguish significant differences in forest density correctly, at least by visual inspection.

Again, the segmentations are noisy due to the low number of looks and the pixelwise SEM procedure, whereas the $\mathcal{W}$-MRF and $\mathcal{KW}$-MRF classifiers are able to generate spatially homogeneous segmentation results. Figs. 6(c) and 6(d) show these results. In field C it is observed that the $\mathcal{KW}$ classifier already does a good job in smoothing the field into a quite homogeneous area. The field is further smoothed by the $\mathcal{KW}$-MRF classifier, but without removing small distinct targets. Specifically, the brown segment in Fig. 6(d) is now relatively smooth, and the purple and yellow features retained are clearly visible in

the Pauli image. The main conclusion of this experiment is that the combination of a non-Gaussian density model with an MRF model improves the segmentation results. The $\mathcal{KW}$-MRF classifier, therefore, produces the most reliable clustering, as seen in Fig. 6(d).

### C. Kongsvegen Glacier Example

The final experiment uses an ENVISAT ASAR C-band dual-pol image over the Kongsvegen glacier, Svalbard, from 2005. The polarimetric channels recorded were VV and VH. A Quasi-Pauli composite image (R=VV, G=VH, B=VV) is shown in Fig. 7(a). The SLC data are geocoded and multi-looked to produce MLC images with 30 m resolution and 24 looks. The ENL was estimated to be 18 for this image. A mask is applied to mask out mountains and isolate the glacier pixels for classification. The segmentation algorithm was therefore working with $2 \times 2$ covariance matrix data.

The three major zones of interest to glaciologists: glacier ice, superimposed ice and firn, can be visibly identified in Fig. 7(a) by their dark, medium and bright intensities, respectively. Comparing the $\mathcal{W}$ and $\mathcal{KW}$ results: the $\mathcal{W}$ found two subclasses (red and gray) for the firn area; and the $\mathcal{KW}$ also found those classes. Also note that both classifiers have a bit of trouble in distinguishing the superimposed ice/glacier ice boundary due to overlap between classes. Figs. 7(b) and 7(c) show the results.

The regularization and homogenization obtained by a subsequent MRF stage can be seen in the Figs. 7(d) and 7(e). The $\mathcal{W}$-MRF has trouble with firn area, and it seems that the MRF cannot compensate for the restriction imposed by the pixelwise model (similar to the urban class of the Exp. 1). The $\mathcal{KW}$-MRF smooths the firn region perfectly into one class and the gray speckly parts (class 7) has mostly disappeared after incorporating the MRF.

### VI. Conclusions

In this paper, a novel unsupervised contextual segmentation algorithm for PolSAR imagery has been developed by combining an advanced statistical distribution with spatial contextual information. The algorithm has been constructed based on a Markov random field (MRF) model that integrates a $\mathcal{K}$-Wishart ($\mathcal{KW}$) distribution for the PolSAR data statistics conditioned to each image cluster and a Potts model for the spatial context. The method is totally unsupervised, which is an advantage since in many cases ground truth is not available.

The added value of combining the flexible non-Gaussian $\mathcal{KW}$ distribution and the Potts MRF model were tested on three examples. The segmentation results before and after MRF modeling for both the standard Wishart and the $\mathcal{KW}$ classifier have been obtained. The segmentations have been compared in terms of discriminability of non-Gaussian regions with $\mathcal{KW}$ with respect to standard Wishart model and contextual smoothing with MRF. The effectiveness of MRF models in improving the accuracy (quantified for simulated data) and reliability of PolSAR image clustering has been remarked for all examples. The
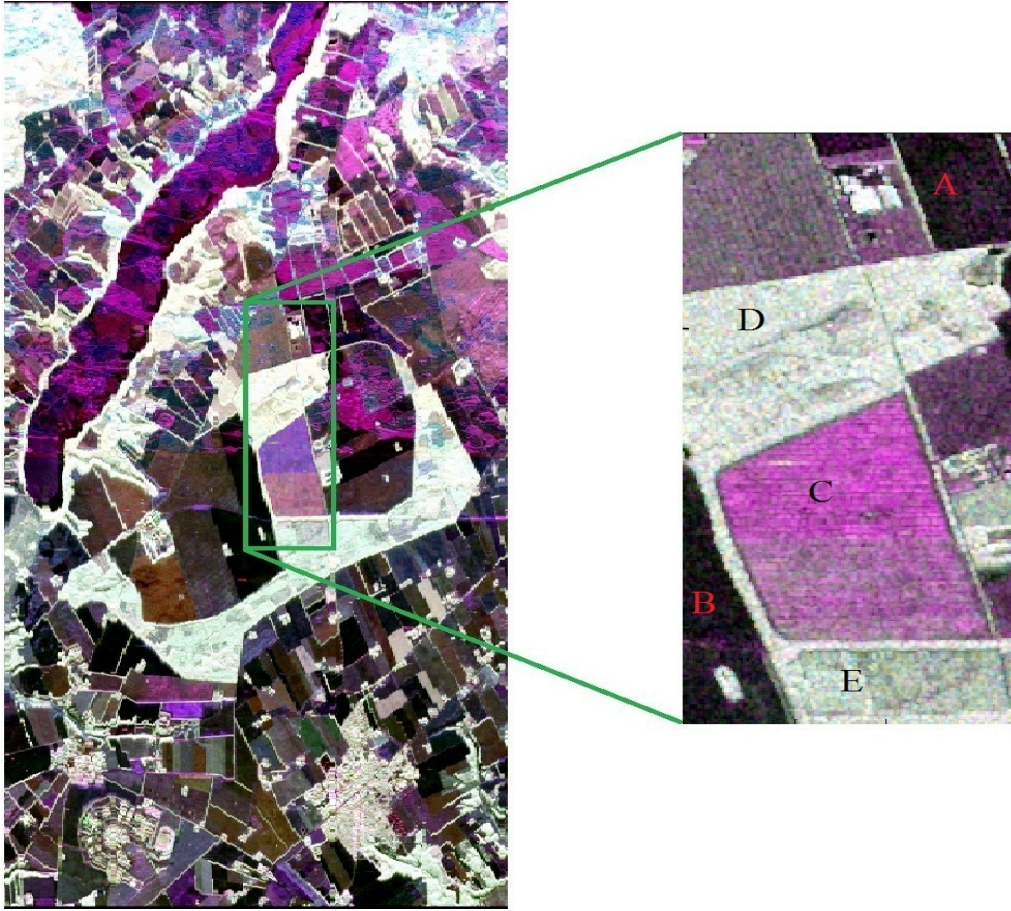
Fig. 5. Left: EMISAR L-band quadpol acquisition over Foulum, Denmark, multi-looked with 8 looks, shown as Pauli decomposition composite image. Right: Enlarged version of test area annotated with field labels referenced in the discussion.
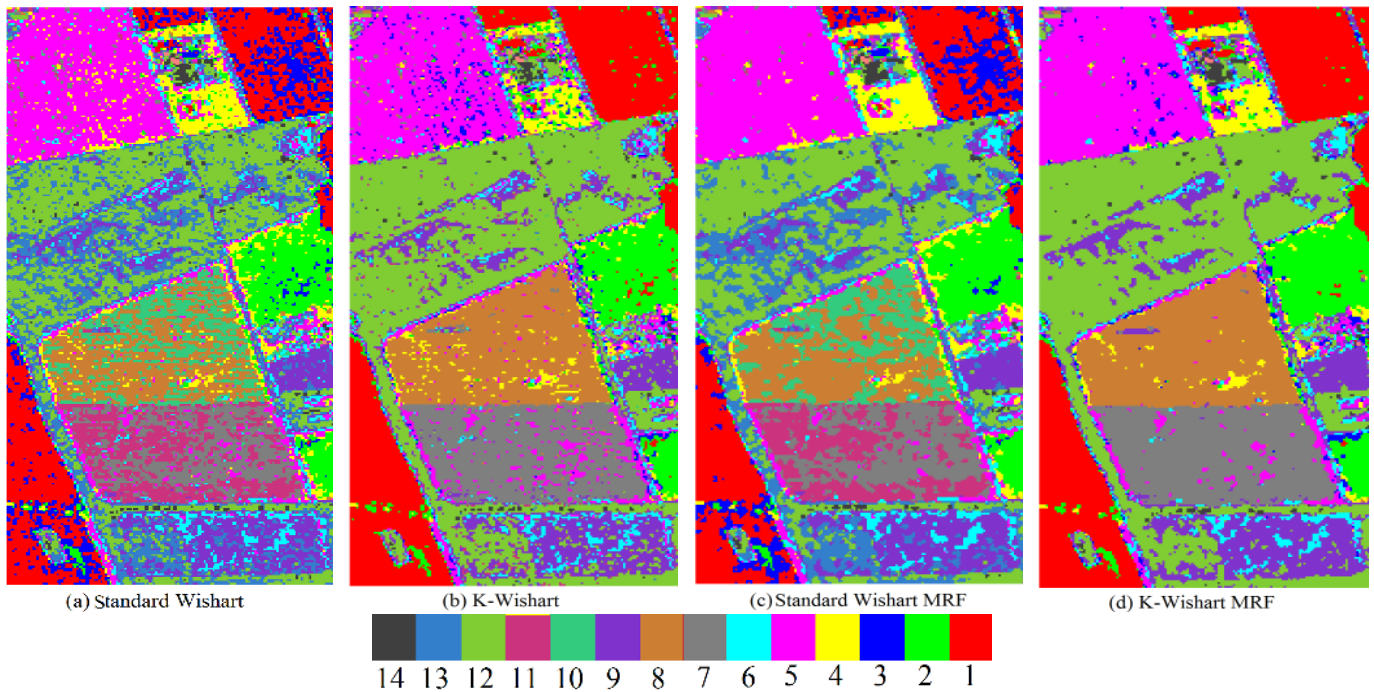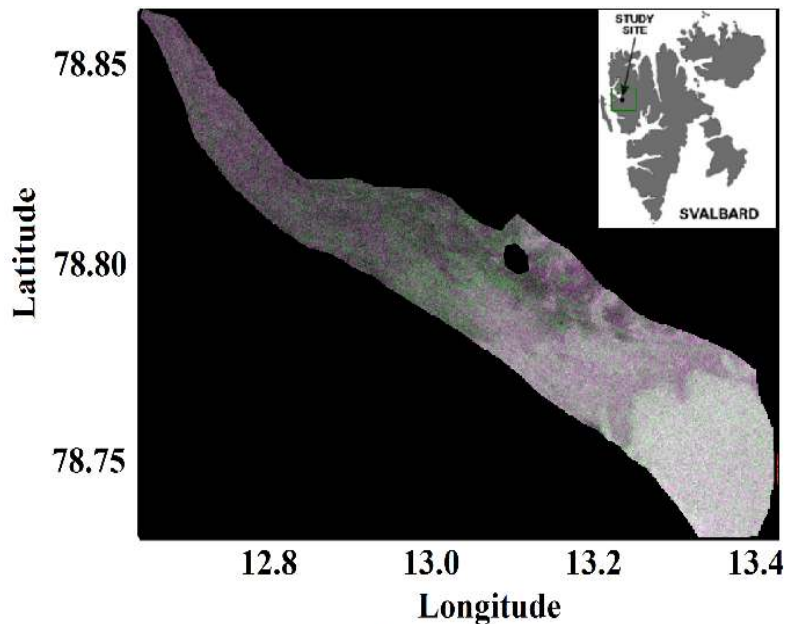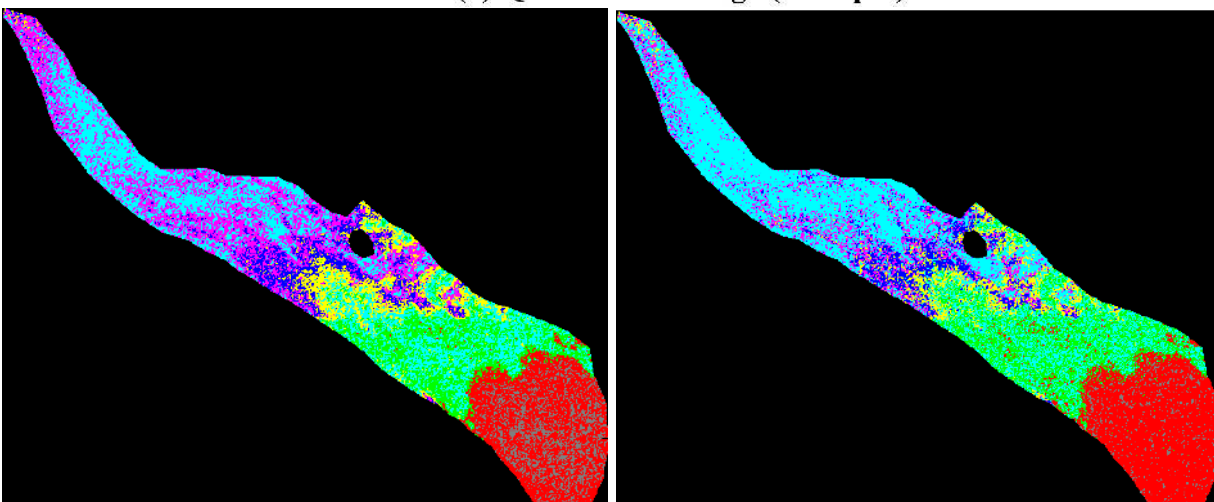


Fig. 6. 14 class segmentation results for (a) Standard Wishart (b) $\mathcal{K}$-Wishart (c) Standard Wishart MRF, and (d) $\mathcal{K}$-Wishart MRF classifier.
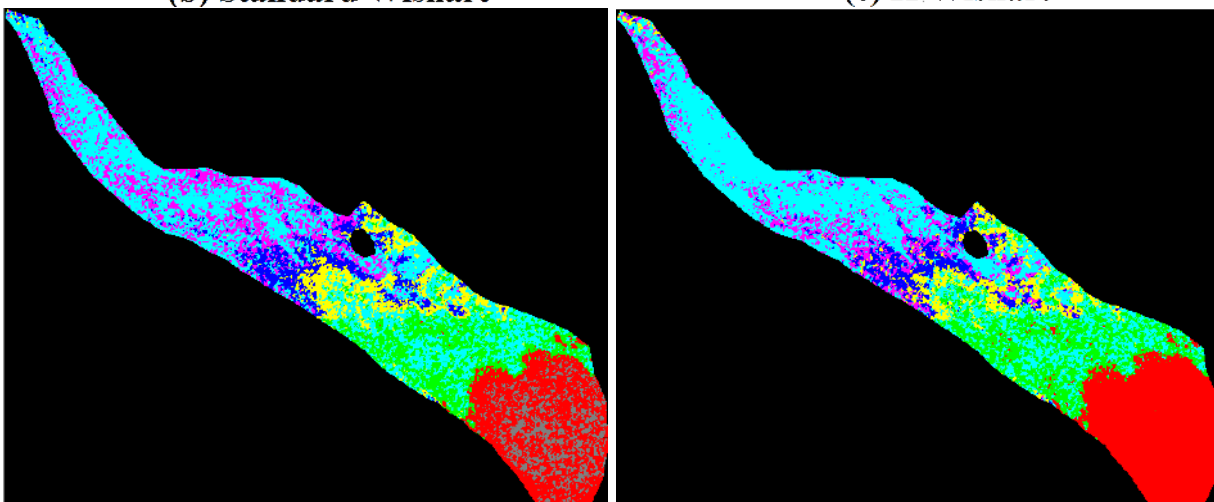
Fig. 7. Dual-pol ENVISAT image multi-looked with 24 looks ( ENL=18): (a) Quasi-Pauli image of Kongsvegen glacier and its location on Svalbard. The three major zones of the image: glacier ice, superimposed ice, and firn, can be visibly identified by their dark, medium, and bright intensities, respectively. (b) Standard Wishart (c) $\mathcal{K}$-Wishart (d) Standard Wishart MRF, and (e) $\mathcal{K}$-Wishart MRF segmentation results, 7 classes found.

results show improvement with respect to segmentation of pixelwise clustering. Future developments of this study will include extending the proposed MRF modeling to change detection applications with PolSAR data. In regard to the computation time, the whole process is slightly slower than the original pixelwise SEM algorithm due to the additional MRF stage in the clustering scheme. Even on the basis of data with a low number of looks (and therefore a high degree of speckle), the proposed approach is able to generate homogeneous and reliable clustering results.

## References

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

[2] J. J. van Zyl, "Unsupervised classification of scattering mechanisms using radar polarimetry data," *IEEE Trans. Geosci. Remote Sensing*, vol. 27, no. 1, pp. 36-45, Jan. 1989.

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

[4] C. Oliver and S. Quegan, "Understanding Synthetic Aperture Radar Images. SciTech Publishing," Raleigh, USA, 2nd edition, 2004.

[5] J.-S. Lee, D. L. Schuler, R. H. Lang, and K. J. Ranson, "$\mathcal{K}$-Distribution for multi-look processed polarimetric SAR imagery," *in Proc. IEEE Int. Geosci. and Remote Sensing Symp.*, pp. 2179-2181, Pasedena, USA, 1994.

[6] C. C. Freitas, A. C. Frery, and A. H. Correia, "The polarimetric $\mathcal{G}$ distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13-31, 2005.

[7] A. P. Doulgeris, S.N. Anfinsen, and T. Eltoft, "Automated non-Gaussian clustering of polarimetric synthetic aperture radar Images," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 10, pp. 3665-3676 , Dec. 2011.

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

[9] V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, "Supervised high resolution dual polarization SAR image classification by finite mixtures and copulas," *IEEE Journal of Selected Topics in Signal Procesing*, vol. 5, no. 3, pp. 554-566, Jun. 2011.

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

[11] S. B. Serpico and G. Moser, "Weight parameter optimization by the Ho-Kashyap algorithm in MRF models for supervised image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no.12, pp. 3695-3705, Dec. 2006.

[12] G. Moser and S. B. Serpico, "Unsupervised change detection from multichannel SAR data by Markovian data fusion," *IEEE Trans. Geosci. Remote Sensing*, vol. 47, no. 7, pp. 2114-2128, Jul. 2009.

[13] A. H. Solberg, T. Taxt, and A. K. Jain, "A Markov random field model for classification of multisource satellite imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, no. 1, pp. 100-113, Jan, 1996.

[14] A. Reigber, M. Jager, M. Neumann, and L. Ferro-Famil, "Classifying polarimetric SAR data by combining expectation methods with spatial context," *Int. J. Remote Sensing*, vol. 31, no. 3, pp. 727-744, Feb. 2010.

[15] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721-741, Nov. 1984.

[16] R. Fjørtoft, Y. Delignon, W. Pieczynski, Sigelle M, and F., Tupin, "Unsupervised classification of radar images using hidden Markov chains and hidden random fields," *IEEE Trans. Geosci. Remote Sensing*, vol. 41, no. 3, pp. 675-686, Mar. 2003.

[17] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE Trans. on Medical Imaging*, vol. 20, no.1, pp. 45-57, Jan. 2001.

[18] J. Schou and H. Skriver, "Restoration of polarimetric SAR images using simulated annealing," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 9, pp. 2005-2016, Sep. 2001.

[19] G. Celeux and J. Diebolt. "The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem," *Computational Statistics Quaterly*, vol. 2, no. 1, pp. 73-82, 1985.

[20] S. Li, "Markov random field modeling in image analysis", 3th ed. London, England: Springer-Verlag, 2009.

[21] M. V. Ibanez and A. Simo, "Parameter estimation in Markov random field image modeling with imperfect observations. A comparative study," *Pattern Recognit. Lett.*, vol. 24, no. 14, pp. 2377-2389, 2003.

[22] G. Moser, J. Zerubia, S. B. Serpico, "Dictionary-based stochastic expectation-maximization for SAR amplitude probability density function estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 1, pp. 188-200, Jan. 2006.

[23] P. R. Kersten, J. S. Lee, and T. L. Ainsworth, "Unsupervised classification of polarimetric synthetic aperture radar images using fuzzy clustering and EM clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 516-527, Mar. 2005.

[24] S. N. Anfinsen and T. Eltoft, "Application of the matrix-variate Mellin transform to analysis of polarimetric radar images," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 6, pp. 2281-2295, Jun. 2011.

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

[26] S. N. Anfinsen, "Statistical analysis of multilook polarimetric radar images with the Mellin transform," Ph.D. dissertation, Univ. Tromsø, Tromsø, Norway, May, 2010.

[27] J.-M. Nicolas, "Introduction aux statistique de deuxième espèce: Application des logs-moments et des logs-cumulants à l'analyse des lois d'images radar," *Traitement du Signal*, vol. 19, no. 3, pp.139-167, 2002. In French, English translation in [26].

[28] S. N. Anfinsen and Torbjørn Eltoft, "Analysis of multilook polarimetric radar data with the matrix-variate Mellin transform," Proc. EUSAR 2012, *Proc. EUSAR 2010 – 8th Eur. Conf. on Synthetic Aperture Radar*, Aachen, Germany, 4 pp., 7-10 Jun. 2010.

[29] G. Celeux, F. Forbes, and N. Peyrand, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognition*, vol. 36, no. 1, pp. 131-144, 2003.

[30] S. Geman and D. Gemanm, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721-741, Nov. 1984.

[31] A. L. M. Levada, D. A. Mascarenhas, and A. Tannus, "Pseudo-likelihood equations for Potts MRF model parameter estimation on higher order neighborhood systems," *IEEE Geosci. and Remote Sensing letters*, vol. 5, no. 3, pp. 522-526, Jul. 2008.

[32] J. E. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. R. Stat. Soc., B*, vol. 36, no. 2, pp. 192-236, 1974.

[33] L. Wang, J. Liu, and S.Z. Li, "MRF parameter estimation by MCMC method," *Pattern Recognition*, vol. 33, pp. 1919-1925, 1999.

[34] Y. Yu and Q. Cheng, "MRF parameter estimation by an accelerated method," *Pattern Recognit. Lett.* vol. 24, pp. 1251-1259, 2003.

[35] Q. Jackson and D. Landgrebe, "Adaptive Bayesian contextual

classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454-2463, Nov. 2002.

[36] S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft, " Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery,"*IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3795-3809, Nov. 2008.

[37] X. Song, Z. Duan, X. Jiang, "Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image," *International Journal of Remote Sens.*, vol. 33, no. 10, pp. 3301-3320, May 2012.

[38] A. P. Doulgeris, S. N. Anfinsen, Y. Larsen, K. langley, and T. Eltoft, "Evaluation of polarimetric configuration for glacier classification," *in Proc. POLinSAR 2009*, Frascati, Italy, 8 pp., Jan., 2009.

[39] R. G. Congalton and R. A. Mead, "A review of three discrete multivariate analysis techniques used in assessing the accuracy of remotely sensed data from error matrices," *IEEE Trans. Geosci. Remote Sens.*, vol. 24, no. 1, pp. 169-174, Jan. 1986.

**Vahid Akbari** (S'10) received the M.Sc. degree (*summa cum laude*) in Remote Sensing from the University of Tehran, Tehran, Iran, in 2009. He is currently working toward the Ph.D. degree in Physics at the University of Tromsø, Norway. He is also affiliated with the Barents Remote Sensing School.

His current research interets lies in the field of statistical analysis of polarimetric synthetic aperture radar images and radar interferometry. In particular, his work concerns statistical modeling for segmentation and multitemporal change detection of multi-polarization radar imagery.

**Anthony P. Doulgeris** (S'06–M'12) received the B.Sc. degree in physics from the Australian National University, Canberra, Australia, in 1988, the M.Sc. degree and the Ph.D. degree in physics from the Department of Physics and Technology, University of Tromsø, Tromsø, Norway, in 2006, and 2011, respectively.

He is continuing his research in applied Earth observation as a postdoctoral research fellow at the Department of Physics and Technology, University of Tromsø, Tromsø, Norway, and is also affiliated with the Barents Remote Sensing School. His research interests focus on investigating remote sensing, pattern recognition and multidimensional statistical modeling, in particular with polarimetric synthetic aperture radar images.

**Gabriele Moser** (S'03–M'05) received the laurea (M.Sc.) degree (*summa cum laude*) in telecommunications engineering and the Ph.D. degree in space sciences and engineering from the University of Genoa, Genoa, Italy, in 2001 and 2005, respectively. Since 2001, he has cooperated with the Signal Processing and Telecommunications Research Group (SPT) of the Department of Telecommunications, Electronic, Electrical, and Naval Engineering, University of Genoa, in the field of remote-sensing image analysis. From January to March 2004, he was a visiting student at the Institut National de Recherche en Informatique et en Automatique (INRIA), Sophia Antipolis, France, working with the Ariana research group on the problem of SAR data modeling. He is currently Assistant Professor of Telecommunications at the University of Genoa.

His research activity is focused on image-processing and pattern-recognition methodologies for remote-sensing data interpretation. In particular, his current research interests include SAR data analysis, multitemporal image classification, hyperspectral image analysis, contextual classification, and geo/biophysical parameter estimation. He has been a reviewer for several international journals. Since 2008, he has been an Associate Editor of the IEEE Geoscience and Remote Sensing Letters.

**Torbjørn Eltoft** (M'92) received the degrees of Cand.Real. (M.Sc.) and Dr.Scient. (Ph.D.) from University of Tromsø, Tromsø, Norway, in 1981 and 1984, respectively.

In 1988 he joined the Faculty of Science, University of Tromsø, and is presently Professor in electrical engineering at the Department of Physics and Technology, where he also is head of the Earth Observation Laboratory and of the BARESS Remote Sensing School. He also holds a position as Adjunct Professor at Norut, Tromsø. His current research interests include multidimentional signal and image analysis with application in remote sensing, statistical models, neural networks, and machine learning.

Dr. Eltoft has been associate editor of the journal *Pattern Recognition* for five years. He was awarded the year 2000 Outstanding Paper Award in Neural Networks by IEEE Neural Networks Council, and Honorable Mention for the 2003 Pattern Recognition Journal Best Paper Award.

**Stian N. Anfinsen** (S'06–M'10) received the Cand.Mag. (B.Sc.), Cand.Scient. (M.Phil.) and Ph.D. degrees in physics from the University of Tromsø, Tromsø, Norway in 1997, 2000 and 2010, respectively. He received the M.Sc. in Communications, Control and Digital Signal Processing with distinction from the University of Strathclyde, Glasgow, U.K. in 1998.

He was with satellite ground station system provider Kongsberg Spacetec in Tromsø from 2001 to 2005. He then moved to the University of Tromsø, Department of Physics and Technology, where he is currently a postdoctoral research fellow. He is also affiliated with the Barents Remote Sensing School. His current research focuses on statistical modelling, parameter estimation, classification, change detection and target detection in polarimetric synthetic aperture radar images.

**Sebastiano B. Serpico** (M'87-SM'00-F'08) received the Laurea degree in electronic engineering and the doctorate in electrical communications from the University of Genoa, Italy. A full professor of telecommunications at the Faculty of Engineering of the University of Genoa, he is the Head of the Signal Processing and Telecommunications laboratory of the Department of Telecommunications, Electronic, Electrical, and Naval Engineering (formerly the Department of Biophysical and Electronic Engineering) of the University of Genoa. His current research interests include pattern recognition for remote sensing images and biomedical images.

He is the chairman of the Institute of Advanced Studies in Information and Communication Technologies (ISICT). He has been the project manager of numerous research contracts and an evaluator of project proposals for various programmes of the European Union. He is the author (or coauthor) of over 200 scientific articles published in journals and conference proceedings. He is an associate editor of the international journal IEEE Transactions on Geoscience and Remote Sensing (TGRS). He was a guest editor of two Special Issues of TGRS on the subject of the analysis of hyperspectral image data (July 2001 issue) and the subject Advances in techniques for the analysis of remote sensing data (March 2005 issue). From 1998 to 2002, he was the chairman of a SPIE/EUROPTO series of conferences on Signal and Image Processing for Remote Sensing. He is a Fellow of the IEEE Society.