High resolution SAR-image classification

Vladimir Krylov, Josiane Zerubia

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

High resolution SAR image classification

Vladimir Krylov — Josiane Zerubia

N° 7108

November 2009

Thème COG

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--7108--FR+ENG

High resolution SAR image classification

Vladimir Krylov* , Josiane Zerubia†

Thème COG — Systèmes cognitifs
Projet Ariana

Rapport de recherche n° 7108 — November 2009 — 36 pages

Abstract: In this report we propose a novel classification algorithm for high and very high resolution synthetic aperture radar (SAR) amplitude images that combines the Markov random field approach to Bayesian image classification and a finite mixture technique for probability density function estimation. The finite mixture modeling is done by dictionary-based stochastic expectation maximization amplitude histogram estimation approach. The developed semiautomatic algorithm is extended to an important case of multi-polarized SAR by modeling the joint distributions of channels via copulas. The accuracy of the proposed algorithm is validated for the application of wet soil classification on several high resolution SAR images acquired by TerraSAR-X and COSMO-SkyMed.

Key-words: SAR image classification, dictionary, amplitude probability density, stochastic expectation maximization, Markov random field, copula

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

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

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

Classification des images RSO haute résolution

Résumé : Dans ce rapport, nous proposons une nouvelle approche pour la classification des images de type Radar à Synthèse d'Ouverture (RSO) haute résolution. Cette approche combine la méthode des champs Markoviens pour la classification bayésienne et un modèle de mélange fini pour l'estimation des densités de probabilité. Ce modèle de mélange fini est realisé grace à une approche fondée sur une espérance-maximisation stochastique, à partir d'un dictionnaire, pour l'estimation des densités de probabilité d'amplitude. Cette approche semi-automatique est étendue au cas important des images RSO avec plusieurs polarisations, en utilisant des copulas pour modéliser les distributions jointes. Des résultats expérimentaux, sur plusieurs images RSO réelles (Dual-Pol TerraSAR-X et Singe-Pol COSMO-SkyMed), pour la classification de zones humides, sont présentés pour montrer l'efficacité de l'algorithme proposé.

Mots-clés : classification d'image RSO, dictionnaire, densité de probabilité d'amplitude, espérance-maximisation stochastique, champ Markovien, copula

Contents

1 Introduction

In the context of modern remote sensing the use of Synthetic Aperture Radar (SAR) represents a very important source of information for Earth observation. Working in the domain of microwaves, SAR is an active imagery system which can be operational regardless of the weather conditions and time of the day. Thanks to this attractive property, SAR images are becoming widely used nowadays in various applications, e.g., in flood/fire monitoring, epidemiological surveillance, agriculture assessment, urban mapping. One of basic problems concerned in all of these applications is the need for accurate and fast classification. The results of classification can either be directly used in one of the above applications or serve as input to further SAR processing problems like segmentation or change detection.

Contemporary SAR systems are capable of providing fully polarized images (HH, HV, VH, VV). Polarimetric SAR imagery, in comparison to single channel SAR data, have the advantage of a more complete description of landcover's scattering behavior. The possible gain in accurate classification from data in several polarizations compared to one channel case explains the special interest to multi-polarized image classification. In this work we investigate the multi-polarization SAR case, as well as single-polarization SAR as a special case.

In this report we develop a semiautomatic algorithm for SAR classification using solely the amplitude data and not the complex data as the approaches above. This is considered as an important classification framework both because several image products provided by novel high resolution satellite SAR systems are geocoded ellipsoid-corrected amplitude (intensity) images and because this modality was the main one for several earlier coarser resolution sensors (e.g., ERS). The proposed algorithm combines the Markov random field (MRF) approach to Bayesian image classification with the dictionary-based stochastic expectation maximization (DSEM) amplitude histogram estimator. MRFs represent a general family of probabilistic image models that provide a convenient and consistent way to characterize context dependent data [1]. DSEM was developed for the purpose of modeling the distribution of amplitudes of a SAR image. Contrary to the single probability density function (pdf) methods proposed earlier [2][3][4][5][6], it uses a mixture of several distinct pdfs to accurately model the amplitude statistics, thus extending the modeling capacity to heterogenous images. This makes the developed classification algorithm robust with respect to possibly complicated shapes of class histograms. The DSEM-MRF technique is extended to polarimetric SAR images by modeling the joint distributions of single channels via copulas, resulting in Copula-DSEM-MRF approach (CoDSEM). Copulas [7] is a rapidly developing instrument in statistics designed for constructing joint distributions from marginals with a wide variety of allowed dependence structures. Compared to classical fitting into multivariate distributions of specific types, copulas provide a significant increase in flexibility [7]. Thus CoDSEM algorithm has good modeling properties based on two flexible statistical modeling concepts.

The research report is organized as follows. First, in section 2 we present the statistical modeling tools involved in our approach: subsection 2.1 covers the DSEM approach, subsection 2.2 gives an introduction to copula theory and, specifically, to bivariate copulas and

subsection 2.3 describes the construction of MRFs and hidden MRFs for image classification. Section 3 describes the used MRF parameter estimation procedure. In section 4 the proposed CoDSEM algorithm is described. Section 5 presents the employed energy optimization procedure. Section 6 reports results of testing the developed algorithm on high resolution Dual-Pol TerraSAR-X and Single-Pol COSMO-SkyMed images. Finally, conclusions are drawn in Section 7.

2 Statistical Modeling

2.1 Dictionary based Stochastic Expectation Maximization

In this section we provide an overview of the dictionary-based stochastic expectation maximization approach (DSEM). A detailed overview of DSEM with experimental validation can be found in [8][9][10].

2.1.1 Finite Mixture Model approach and dictionary

In order to formalize the common scenario when several distinct land-cover typologies are present in the same SAR image, a finite mixture model (FMM) [11][12] for the distribution of grey levels is assumed. Specifically, the SAR image is modeled as a set $\mathcal{I} = \{r_1, r_2, \ldots, r_N\}$ of independent and identically distributed (i.i.d.) samples drawn from the mixture pdf:

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

where $p_i(r|\theta_i)$ are pdfs dependent on vectors θ_i of parameters, taking values in a set $\Theta_i \subset \mathbb{R}^{\ell_i}$ and $\{P_1, \cdots, P_K\}$ is a set of mixing proportions (i.e., $\sum_{i=1}^{K} P_i = 1$ with $0 \leq P_i \leq 1, \quad i = 1, 2, \ldots, K$). Thus the aim is to estimate the parameter vector $\theta = (\theta_1, \cdots, \theta_K; P_1, \cdots, P_K)$. This so-called "i.i.d. approach" is widely accepted in the context of estimation theory [13] and corresponds to discarding the contextual information associated to the correlation between neighboring pixels in the image during the estimation process, thus exploiting only the greylevel information. However in the developed in this report CoDSEM algorithm, this underlying assumption of DSEM is justified by the use of a hidden MRF model (as will be described in Sec. 4), which incorporates in itself the modeling of contextual dependencies.

Each component $p_i(r|\theta_i)$ is modeled by resorting to a finite dictionary $\mathcal{D} = \{f_1, \cdots, f_R\}$ (see Table 1) of $R = 4$ SAR-specific distinct parametric pdfs $f_i(x|\theta_i)$, parameterized by vectors $\theta_i \in A_i, i = 1, \ldots, R$. For descriptions of distributions and their physical properties see [10]. The method integrates the number of mixture component estimation, the selection of the optimal model for each mixture component inside the dictionary and the parameter estimation for each mixture component.

As shown in Table 1, we use the dictionary $\mathcal{D} = \{$Weibull, Lognormal, Nakagami-Gamma, Generalized Gamma$\}$ of 4 components. Whereas in [10] the use of larger dictionary of

Family	Probability distribution function	MoLC equations
Log-normal	$f_1(r) = \frac{1}{\sigma r\sqrt{2\pi}} \exp\left[-\frac{(\ln r - m)^2}{2\sigma^2}\right],$	$\kappa_1 = m$ $\kappa_2 = \sigma^2.$
Weibull	$f_2(r) = \frac{\eta}{\mu^\eta} r^{\eta-1} \exp\left[-\left(\frac{r}{\mu}\right)^\eta\right],$	$\kappa_1 = \ln\mu + \Psi(1)\eta^{-1}$ $\kappa_2 = \Psi(1,1)\eta^{-2}.$
Nakagami	$f_3(r) = \frac{2}{\Gamma(L)} (\lambda L)^L r^{2L-1} \exp\left(-\lambda L r^2\right),$	$2\kappa_1 = \Psi(L) - \ln\lambda L$ $4\kappa_2 = \Psi(1, L).$
Generalized Gamma	$f_4(r) = \frac{\nu}{\sigma\Gamma(\kappa)} \left(\frac{r}{\sigma}\right)^{\kappa\nu-1} \exp\left\{-\left(\frac{r}{\sigma}\right)^\nu\right\},$	$\kappa_1 = \Psi(\kappa)/\nu + \ln\sigma$ $\kappa_2 = \Psi(1,\kappa)/\nu^2$ $\kappa_3 = \Psi(2,\kappa)/\nu^3.$

Table 1: Pdfs and MoLC equations for the parametric families included in the considered dictionary $\mathcal{D}$. Here $\Gamma(\cdot)$ is the Gamma function [14], $\Psi(\cdot)$ the Digamma function [14] and $\Psi(\nu,\cdot)$ the νth order polygamma function [14].

8 pdf has been suggested, adding $\mathcal{D}' = \{$Fisher, K-root, Generalized Gaussian Rayleigh, Symmetric α-Stable$\}$ to the considered here dictionary $\mathcal{D}$. The motivation for using the smaller dictionary will be discussed later.

2.1.2 SEM and parameter estimation

As discussed in [8][10], considering the variety of estimation approaches for FMMs a reasonable choice for this particular estimation problem is the stochastic expectation maximization (SEM) scheme [15]. Thanks to the stochastic sampling involved in this scheme, one gains numerical tractability along with the better exploring capabilities as compared to EM scheme and higher chances of avoiding the local maximuma. Indeed, the sequence of parameter estimates generated by SEM is a discrete time random process that does not converge pointwise nor almost surely; it 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 [15] (in other words, the maximum likelihood estimate of mixture parameters is asymptotically equivalent to the mathematical expectation of θ^t taken with respect to stationary distribution of SEM-process).

SEM is an iterative estimation scheme dealing with the problem of *data incompleteness*. The incomplete data can in general be an unobserved part of the data or either corrupt data. The incompleteness issue is formalized by assuming a "complete" data vector x to be unavailable and observable only through an "incomplete" data vector $y = \Phi(x)$ obtained by a many-to-one mapping $\Phi : X \to Y \subset \mathbf{R}^m$. Thus, a given realization y of the incomplete data may have been generated by any realization $x \in \Phi^{-1}(y)$; this does not allow, for instance, a direct computation of an ML estimate. SEM tries to avoid these difficulty by iteratively and randomly sampling a complete data set and using it to compute a standard ML estimate.

Specifically, the FMM approach is regarded as being affected by a data incompleteness problem, since it is not known from which of the available statistical populations (corresponding directly to mixture components) involved in (1) a given image sample has been drawn. In this manner, the complete data for FMM is represented by the set $\{(r_i, s_i), i = 1, \ldots, N\}$, where r_i is the observed part (SAR amplitude) and s_i the lacking data (label). Given an FMM with K components, s_i takes value in $\{1, \ldots, K\}$ and denotes to which out of the K components the i-th pixel belongs.

The classic SEM procedure consists of 3 steps:

- Expectation step (**E-step**), where the probabilities for missing information, i.e. pixel labels, are calculated based on the information collected on the previous iteration,
- Stochastic step (**S-step**), where the labels are sampled with respect to the distributions estimated on E-step, and
- Maximization step (**M-step**), where the parameter estimate is updated by computing a standard ML estimate according to the complete data realization sampled on S-step.

For detailed description see [15].

Instead of calculating ML estimates on the M-step, the Method of Log-Cumulants (MoLC) is adopted [3][8], which have been proved to be a feasible and effective estimation tool for common SAR parametric models [3] and also for all pdfs in our dictionary [6][8]. MoLC has recently been proposed as a parametric pdf estimation technique suitable for distributions defined on $[0, +\infty)$; it has been widely applied in the context of SAR-specific parametric families for amplitude and intensity data modeling, e.g., to Nakagami distribution [3].

In [16] method of moments is adopted for random variables on $[0, +\infty)$, using the Mellin transform [14] by analogy to the Laplace transform in moment generating function:

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

The νth order log-cumulants are defined as $\kappa_\nu = [\ln \phi_u]^{(\nu)}(1)$, $\nu = 1, 2, \ldots$, where $^{(\nu)}$ stands for the νth derivative. In case of the transform convergence, the following MoLC equations take place [16]:

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

This allows stating a set of equations relating the unknown parameters of a given parametric model with one or more log-cumulants. The solution of such equations allows computing the desired parameter estimates [3][6]. These equations have one solution for any observed log-cumulants for all of the pdfs in $\mathcal{D}$ (see in Table 1).

Thus, instead of M-step in the SEM two steps appear:
- MoLC step, where for every component the parameters of all R models from the dictionary are estimated, and

- Model Selection step (MS-step) where the pdf that provides the highest value of likelihood is picked.

The other crucial point in the FMM estimation is the choice of the number of components. We adopt the strategy developed in [10]: SEM is initialized with un upper bound $K_0 > K^*$ of the number of components; then, if after the t-th iteration the value P_i^{t+1} goes below some predefined threshold, the component is eliminated and at the $(t+1)$-th iteration the mixture consists of $(K-1)$ components.

The resulting algorithm is organized as follows:

- **E-step:** compute, for each greylevel z and i-th component, the posterior probability estimates corresponding to the current pdf estimates, i.e. $z = 0, \cdots, Z-1, \quad i = 1, \cdots, K$:

$$\tau_i^t(z) = \frac{P_i^t p_i^t(z)}{\sum_{j=1}^K P_j^t p_j^t(z)};$$

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

- **K-step:** for all i, $i = 1, \cdots, K$: if P_i^t is below a given threshold, eliminate the i-th component, update K;

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

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

  where $b = 2, 3$, $h(z)$ is the image histogram, $Q_{it} = \{z : s^t(z) = \sigma_i\}$ is the set of grey levels assigned to the i-th component; then, solve the corresponding MoLC equations (see Table 1) for each parametric family $f_j(\cdot|\theta_j)$ $(\theta_j \in A_j)$ in the dictionary, thus computing the resulting MoLC estimate $\theta_{ij}^t, \quad i = 1, \cdots, K, \quad j = 1, \cdots, R$;

- **MS-step:** for the i-th mixture component, compute the log-likelihood of each estimated pdf $f_j(\cdot|\theta_{ij}^t)$ according to the data assigned to the i-th component:

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

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

2.2 Copulas

For the purpose of modeling the joint distribution of a multivariate random variable given marginal distributions we employ copulas. It is worth nothing that many current SAR sensors include a Dual-Pol acquisition modality; this further enforces the interest in the classification in two-dimensional SAR amplitude/intensity spaces. Indeed, as mentioned in introduction, many parametric models have been proposed for the marginal statistics of SAR amplitudes or intensities, but only very few models are available for the joint distribution of several SAR amplitudes (e.g., Nakagami-Gamma model developed in [17]). In order to overcome this limitation, we combine the marginal pdfs provided by DSEM with copulas.

In this section we present the basic definitions and facts necessary for using the concept of copulas. For detailed overview we refer to the classical textbook on copulas [7]. For the sake of simplicity we present only the 2D case of copulas, though all the concepts presented can be extended to multivariate cases.

2D copula is a bivariate joint distribution defined on $[0,1]^2$ such that marginal distributions are uniform on the interval [0, 1]. Specifically,

Definition 1. A *bivariate copula* is a function $C : [0,1]^2 \to [0,1]$, which satisfies the following properties:

1. both marginals are uniformly distributed on $[0,1]$;
2. for every u,v in $[0,1]$:

$$C(u,0) = C(0,v) = 0 \quad \text{and} \quad C(u,1) = u, \quad C(1,v) = v;$$

3. for every $u_1 \leq u_2$, $v_1 \leq v_2$ in $[0,1]$:

$$C(u_2,v_2) - C(u_1,v_2) - C(u_2,v_1) + C(u_1,v_1) \geq 0. \tag{2}$$

The second property in the definition provide the necessary boundary conditions. The third is the so-called *2-increasing property*: for the cumulative distribution function (CDF) $C(x,y)$ the left hand part of Eq. (2) represents the probability that the r.v. with joint CDF $C(\cdot,\cdot)$ is inside the rectangle $[u_1,u_2] \times [v_1,v_2]$, thus guarantying this probability to be above zero.

The importance of copulas in statistics is explained by Sklar's Theorem:

Theorem. *Let X and Y be random variables with joint distribution function H and marginal distribution functions F and G, respectively. Then there exists a copula C such that*

$$H(x,y) = C(F(x),G(y)) \tag{3}$$

for all x, y in $\mathbb{R}$. If F and G are continuous, then C is unique. Otherwise, the copula C is unique on the range of values of the marginal distributions F and G. Conversely, if C is a copula and F and G are distribution functions, then the function H defined by (3) is a joint distribution function with marginals F and G.

Thus copulas link joint distribution functions to their one-dimensional margins. A proof of this theorem can be found in [7].

The simplest example of a copula is the *product copula*

$$\Pi(u, v) = uv, \tag{4}$$

which characterizes independent random variables when the distribution functions are continuous.

Another useful representation of copulas can be given in terms of pdfs. Given absolutely continuous random variables with pdfs $f(x)$ and $g(y)$ with corresponding CDFs $F(x)$ and $G(y)$ the pdf of the joint distribution $h(x, y)$ is given by

$$h(x, y) = f(x)g(y)\frac{\partial^2 C}{\partial x \partial y}(F(x), G(y)), \tag{5}$$

where $\frac{\partial^2 C}{\partial x \partial y}(x, y)$ is pdf corresponding to the copula $C(x, y)$.

An important family of copulas are archimedean copulas, which have a simple analytical form and yet provide a large variety of modeled dependence structures. Unlike other copula families, most archimedean copulas have closed-form solutions and don't need to be derived from the multivariate distribution functions by means of Sklar's Theorem.

Definition 2. An *archimedean copula* is a bivariate copula C defined as

$$C(u_1, u_2) = \phi^{-1}(\phi(u_1) + \phi(u_2)), \tag{6}$$

where the *generator function* $\phi(u)$ is any function satisfying the following properties:

1. $\phi(u)$ is continuous on $[0, 1]$;
2. $\phi(u)$ is decreasing, $\phi(1) = 0$;
3. $\phi(u)$ is convex.

The product copula (4) is an archimedean copula with generator function $\phi(u) = -ln(x)$.

Whereas the product copula represents a fairly simple spatial structure, i.e. the diagonal on sample scatter plot, other copulas present more sophisticated dependence between marginal random variables which can be appreciated from example scatter plots on Fig. 1.

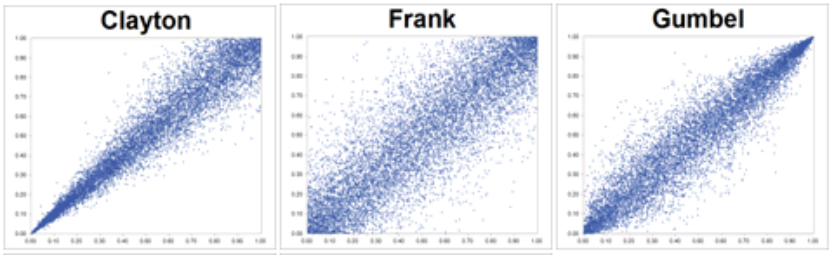


Figure 1: Scatter plots for Clayton, Frank and Gumbel copulas.

Copula	$C(u,v)$	θ interval
Product	uv	-
Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-1/\theta}$	$\theta \in (0,+\infty)$
Ali-Mikhail-Haq	$\frac{uv}{1-\theta(1-u)(1-v)}$	$\theta \in [-1,1]$
Gumbel	$\exp\left(-\left[(-\log(u))^\theta+(-\log(v))^\theta\right]^{1/\theta}\right)$	$\theta \in [0,1]$
Frank	$-\frac{1}{\theta}\log\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$	$\theta \neq 0$
Farlie-Gumbel-Morgenstern (FGM)	$uv(1+\theta(1-u)(1-v))$	$\theta \in [-1,1]$
Marchal-Olkin	$\min\left(u^{1-\theta}v, uv^{1-\theta}\right)$	$\theta \in [0,1]$
A12	$\left(1+\left[(u^{-1}-1)^\theta+(v^{-1}-1)^\theta\right]^{1/\theta}\right)^{-1}$	$\theta \in [1,+\infty)$
A14	$\left(1+\left[(u^{-1/\theta}-1)^\theta+(v^{-1/\theta}-1)^\theta\right]^{1/\theta}\right)^{-\theta}$	$\theta \in [1,+\infty)$
Raftery	$u-\frac{1-\theta}{1+\theta}u^{\frac{1}{1-\theta}}\left(v^{-\frac{\theta}{1-\theta}}-v^{\frac{1}{1-\theta}}\right), u \leq v$ $v-\frac{1-\theta}{1+\theta}v^{\frac{1}{1-\theta}}\left(u^{-\frac{\theta}{1-\theta}}-u^{\frac{1}{1-\theta}}\right), u > v$	$\theta \in [0,1]$

Table 2: List of copulas under consideration.

In this work we consider 10 explicitly defined copulas: 7 archimedean (product, Clayton, Ali-Mikhail-Haq, Gumbel, Frank, A12, A14), 1 copula with a quadratic section (Farlie-Gumbel-Morgenstern) and 2 non-archimedean copulas with simultaneous presence of an absolutely continuous and a singular component (Marchal-Olkin, Raftery). The names for A12 and A14 copulas originate from their positions in the list of archimedean copulas in [7]. This choice of copulas is capable of modeling a wide variety of dependence structure and cover a significant part of copula applications [18]. We summarize the information about copulas we use in this work in Table 2; for further information see [7].

In the context of archimedean copulas the problem of copula estimation turns into the estimation of generator function. In practice, most generator functions are one parameter functions $\phi(u) = \phi_\theta(u)$, though most of them have extensions to multivariate generators thus defining multivariate archimedean copulas [7].

A common way to estimate the copula is by using its connection with ranking dependence measures such as Kendall's tau and Spearman's rho [7] [19]. In this work we implement the estimation based on the former. For more details on the choice between Kendall's tau and Spearman's rho see [7].

Definition 3. *Kendall's tau* is a concordance and discordance measure between two independent realizations (Z_1, Z_2) and $(\hat{Z}_1, \hat{Z}_2)$ from the same law $H(x,y)$ and is expressed by

$$\tau = P\{(Z_1-\hat{Z}_1)(Z_2-\hat{Z}_2)>0\} - P\{(Z_1-\hat{Z}_1)(Z_2-\hat{Z}_2)<0\}. \quad (7)$$

Given the two realizations $z_{1,l}$ and $z_{2,l}$, $l = 1, \ldots, N$, the empirical estimator of Kendall's tau is

$$\hat{\tau} = \frac{\sum_{l=1}^{N-1} \sum_{k=l+1}^{N} z_{1,lk}\, z_{2,lk}}{\binom{N}{2}}, \qquad \text{where } z_{n,lk} = \begin{cases} 1, & \text{if } z_{n,l} \leq z_{n,k} \\ -1, & \text{otherwise} \end{cases} \quad \text{for } n = 1, 2. \tag{8}$$

The probabilities in (7) can be evaluated by integrating over the distribution of $(\hat{Z}_1, \hat{Z}_2)$. So that, in terms of copulas, we get

$$\tau_C = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1, \tag{9}$$

where C is the copula associated with $H(x, y)$.

In case of an archimedean copula $C(u, v)$ defined by generator $\phi(u)$ (9) turns into

$$\tau = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt, \tag{10}$$

where the integral has an analytical closed-form solution in most cases. By plugging the empirical estimate $\hat{\tau}$ in place of τ we get an estimate $\hat{\theta}$ for parameter θ. E.g., in case of the Clayton copula we have $\hat{\theta} = \frac{2\hat{\tau}}{1-\hat{\tau}}$.

One more important subject concerning parameter estimation for copulas via Kendall's tau is the question of τ-relevance intervals for copulas. Given an estimator of tau $\hat{\tau}$ one should decide whether the specific copula under consideration is appropriate for modeling the dependence of this level $\hat{\tau}$. Some copulas are specific to marginals with low level of correlation, others deal with strongly correlated marginals and the third are capable of modeling all levels of rank correlation. So in every case the list of copulas is limited to those which are capable of accurately modeling the specific value $\hat{\tau}$.

The dependencies between θ and τ, along with τ-relevance intervals of the considered copulas are presented in Table 3.

2.3 Markov random fields

Let S be a finite set of sites with a neighborhood system defined on it. Here, it is the two-dimensional lattice with a second-order neighborhood system. For each site, the neighbours are the eight surrounding pixels. A set of sites C is called a clique, if it contains sites that are mutually neighbours. A discrete MRF is defined as a collection of discrete random variables $\mathbf{Z} = \{Z_i, i \in S\}$ defined on S, each Z_i taking value in $\{0, 1\}^M$ with M being the number of classes, whose joint probability distribution satisfies

$$\forall z, P(z_i | z_{S \setminus \{i\}}) = P(z_i | z_j, j \in N(i)) \tag{11}$$

$$\forall z, P(z) > 0, \tag{12}$$

Copula	dependence between θ and τ	τ interval
Product	-	$\tau \in \{0\}$
Clayton	$\theta = \frac{2\tau}{1-\tau}$	$\tau \in (0, 1]$
Ali-Mikhail-Haq	$\tau = \frac{3\theta-2}{3\theta} - \frac{2}{3}\left(1-\frac{1}{\theta}\right)^2 ln(1-\theta)$	$\tau \in [-0.181726, \frac{1}{3}]$
Gumbel	$\theta = \frac{1}{1-\tau}$	$\tau \in [0, 1]$
Frank	$\tau = 1 - \frac{4}{\theta^2}\int_0^\theta \frac{t}{e^{-t}-1}dt$	$\tau \in [-1, 0) \cup (0, 1]$
FGM	$\theta = \frac{9}{2}\tau$	$\tau \in [-\frac{2}{9}, \frac{2}{9}]$
Marchal-Olkin	$\theta = \frac{2\tau}{\tau+1}$	$\tau \in [0, 1]$
A12	$\theta = \frac{2}{3-3\tau}$	$\tau \in [\frac{1}{3}, 1]$
A14	$\theta = \frac{1+\tau}{2-2\tau}$	$\tau \in [\frac{1}{3}, 1]$
Raftery	$\theta = \frac{3\tau}{2+\tau}$	$\tau \in [0, 1]$

Table 3: Dependence $\tau(\theta)$ and τ-relevance intervals.

where $z_{S\setminus\{i\}}$ denotes the realization of the field restricted to $S\setminus\{i\}$ and $N(i)$ denotes the set of neighbours of i. Property (11) means that the interactions between site i and the other sites actually reduce to interactions with neighbours $N(i)$. The property (12) is necessary for the Hammersley-Clifford theorem, which states that the joint probability distribution of a Markov field is a Gibbs distribution

$$P_G(z) = W^{-1}\exp(-H(z)), \tag{13}$$

where H is the energy function

$$H(z) = \sum_c V_c(z_c).$$

The V_c functions are the clique potentials and may depend on parameters, the normalizing factor $W = \sum_z \exp(-H(z))$ is also referred to as partition function. In this case only the cliques of size two are considered, i.e. only the 8 pairwise interactions. An approximation of the likelihood (13) is the pseudo-likelihood introduced in [20]

$$PL(z) = \prod_{i\in S} P_G(z_i|z_{N(i)}). \tag{14}$$

Each term in this product is computed as

$$P_G(z_i|z_{N(i)}) = \frac{\exp(-\sum_{c\ni i} V_c(z_c))}{\sum_{z_i}\exp(-\sum_{c\ni i} V_c(z_c))}.$$

Expression (14) introduces a genuine probability distribution only when the variables are independent.

Image classification involves observed as well as unobserved data to be recovered. In case of hidden MRFs, the unobserved data is modeled by an MRF Z as in (13), with the energy function depending on a parameter β. The observations Y are conditionally independent given Z with the density

$$f(y|z,\theta) = \prod_{i\in S} f_i(y_i|z_i,\theta).$$

The complete likelihood is given by

$$P_G(y,z|\theta,\beta) = f(y|z,\theta)P_G(z|\beta) = W(\beta)^{-1}\exp\{-H(z|\beta) + \sum_{i\in S}\log f_i(y_i|z_i,\theta)\}.$$

Thus the conditional field Z given $Y = y$ is an MRF with energy function

$$H(z|y,\theta,\beta) = H(z|\beta) - \sum_{i\in S}\log f_i(y_i|z_i,\theta), \tag{15}$$

where the contextually dependent part is

$$H(z|\beta) = \sum_c V_c(z_c|\beta). \tag{16}$$

The aim of classification is to recover the unknown image z, interpreted as a classification into a finite number of M labels.

3 MRF parameter estimation

As discussed in Sec. 2.3, to make the parameter estimation problem computationally feasible, the likelihood function $p(x|\theta)$ is replaced by the pseudo-likelihood function (14), so that

$$\log PL(z|\beta) = \log\left[\prod_{s\in S} p(z_s|z_{S\setminus\{s\}},\beta)\right], \tag{17}$$

where conditional probabilities are defined as

$$p(z_s|z_{S\setminus\{s\}},\beta) = \frac{\exp(-H(z_s|z_{S\setminus\{s\}},\beta))}{\sum\limits_{x_s\in X_S}\exp(-H(x_s|z_{S\setminus\{s\}},\beta))}.$$

Following the algorithm developed in [21], we implement the simulated annealing procedure for β estimation:

Step 0. Initialization of temperature T_0, β_0, $t = 1$, γ and maximum number of iterations N.

Step 1. Sampling $\beta' \sim N(\cdot|\beta_t, 1)$.

Step 2. β' is accepted with probability

$$\alpha(\beta', \beta_t) = \min\left\{1, \exp\left(\frac{\log PL(x|\beta') - \log PL(x|\beta_t)}{T}\right)\right\},$$

if β' accepted then $\beta_{t+1} = \beta'$, otherwise $\beta_{t+1} = \beta_t$.

Step 3. Incrementing t; if $t \leq N$ set $T_t = \gamma T_{t-1}$ and goto step 1, stop otherwise.

By choosing γ and N we thus define the final temperature T_N: $T_N = T_0\gamma^N$.

As discussed in [21], this procedure can be followed by Metropolis-Hastings algorithm with gradient oriented probing, but we did not notice any significant improvement by adding this sampling procedure.

The final estimate β^* is chosen by averaging the estimates on the last t iterations

$$\beta^* = \frac{1}{t}\sum_{i=N-t}^{N} \beta_i.$$

4 Copula-DSEM approach

The approach developed in this report belongs to the class of Bayesian image classification. In considers the contextual dependence in the SAR image via MRF.

We start by presenting the multi-polarized case of the CoDSEM algorithm.

The classification has to be done into M classes and D is the number of polarization channels considered.

DSEM-step. The first step is the estimation of class distributions from the observed data. To do it a learning image is taken with at least a partial ground truth. For every class the distributions $p_{dm}(y|z \in \omega_m)$, where ω_m is the set of pixels in class m, are estimated by DSEM

$$p_{dm}(y_d) = p_{dm}(y_d|\omega_m, \theta) = \sum_{i=1}^{K} P_{dmi} p_{dmi}(y_d|\theta_{dmi}), \quad m = 1, \ldots, M,\ d = 1, \ldots, D. \tag{18}$$

This is done independently for all D polarizations, $y = (y_1, \cdots, y_D)$. The corresponding DSEM-estimates for CDFs are

$$F_{dm}(y_d) = \sum_{i=1}^{K} P_{dmi} F_{dmi}(y_d|\theta_{dmi}). \tag{19}$$

As discussed in Sec. 2.1 the use of DSEM instead of other parametric pdf models proves advantageous when modeling inhomogeneous classes (i.e., classes which contain several different landcover subclasses). However, even in case of homogenous classes, DSEM can be

viewed as an efficient means to choose the best single pdf model from the panel of distributions in the dictionary $\mathcal{D}$.

Copula-step. The joint distributions $p_m(\mathbf{y}|\omega_m)$ for classes $m = 1, \cdots, M$ are modeled then from marginal distributions (18). For this purpose we employ copulas. As described in Sec. 2.2 such a choice is justified by a wide variety of dependence structures provided by copulas. As compared to classical alternative of fitting the marginals into the multivariate pdf of a specific distribution (which can also be simulated by copulas [7]), the resulting joint pdf is significantly more flexible [7]. Thus joint pdfs are constructed by generalization of (5) as

$$p_m(\mathbf{y}|\omega_m) = p_{1m}(y_1) \dots p_{Dm}(y_D) \frac{\partial^D C_m^*}{\partial y_1 \dots \partial y_D}(F_{1m}(y_1), \dots, F_{Dm}(y_D)). \tag{20}$$

To choose the most appropriate copula C_m^* from the list of 10 copulas in Sec. 2.2, first we discard those copulas, for which the current estimate $\hat{\tau}_m$ is outside the τ-relevance interval (see Table 3). Then, for each m, we choose the best fitting copula, i.e. the copula with the highest p-value in Pearson Chi-square test-of-fitness (PCS) [19]. In general, PCS tests a null hypothesis that the frequency distribution of certain events observed in sample is consistent with a particular theoretical distribution.

$$X^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i},$$

where O_i and E_i are the observed and the hypothetical frequencies respectively and n the number of outcomes. PCS is one of the statistical tests whose results are chi-square distribution, i.e. $X^2 \sim \chi^2_{n-r-1}$, where r is the number of reductions of degrees of freedom (typically, the number of parameters for parametric CDFs). The reference to χ^2 distribution allows to calculate p-values for several different null hypothesis.

In our case the null hypothesis in PCS is that the sample frequencies

$$C_c(F_1(y_1), \cdots, F_D(y_D)),$$

where $(y_1, \cdots, y_D)$ are the observed data, are consistent with the theoretical frequencies for $C_c(\cdot)$. This is correct due to the well known fact, that if $\mathbf{x}$ is distributed with CDF $F(u)$, and $x_1, \cdots, x_N$ are the independent observations of $\mathbf{x}$, then $F(x_i)$ $(i = 1, \cdots, N)$ are independent $[0, 1]$-uniformly distributed r.v.s [19]

We can now explain the restrictive choice of dictionary $\mathcal{D}$ size to $R = 4$. To estimate the conditional pdfs by (20) the calculation of CDFs $F_{dm}(y_d)$, $d = 1, \cdots, D$ is needed. For the pdfs in $\mathcal{D}$ there are analytical expressions for CDFs; however for K-root, Generalized Gaussian Rayleigh, Symmetric α-Stable from $\mathcal{D}'$ there is no analytical expression and the numerical computation is costly. In addition, tests with a full dictionary $\mathcal{D} \cup \mathcal{D}'$ reported a minor improvement in modeling accuracy (within 1% of correlation coefficient increase for the same number of components) compared to restricted $\mathcal{D}$. So comparing the significant gain in calculation speed and expected minor or negligible improvement in accuracy, the

choice has been done for the former. The Fisher distribution has been taken out of the dictionary because it has been picked very rarely by DSEM for the mixture components.

MRF-step. The next step is the construction of the energy function (15) for the proposed MRF model.

Having no prior information about the geometrical properties or orientation of analyzed SAR image, anisotropic second-order neighborhood system has been chosen as the neighborhood model. The considered cliques are restricted to size 2, i.e. only pairwise interactions are exploited. Thus (16) becomes

$$H(x|\beta) = \sum_{c} V(x_c|\beta) = \sum_{c=\{s,s'\}\in C} \left[-\beta\, \delta_{x_s=x_{s'}}\right],$$

with $\delta_{x_s=x_{s'}} = 1$ if $x_s = x_{s'}$, and 0 otherwise.

Given the conditional pdfs (20) and no prior information about the proportions of classes on the testing image (which results in giving the same prior weights to all the classes) the resulting local energy function is

$$U(\omega_m|\mathbf{y},\beta) = \sum_{i\in S} \left[-\log p_m(\mathbf{y}|\omega_m) - \beta \sum_{s:\{i,s\}\in C} \delta_{x_i=x_s} \right]. \tag{21}$$

Estimation-step. As can be seen from the energy function specified in (18)-(21) there is only 1 parameter β to estimate. This is thanks to the fact that all the parameter estimation for (20) is incorporated into DSEM.

The choice for β-estimation depends on the learning data at hand. Given a ground truth map including sufficient transition areas from each class into all others one can apply an explicit simulated annealing estimation procedure described in Sec. 3. If, on the other hand, the transition information is sparse or insufficient, then the only choice for β-specification is trial-and-error method, as any β-estimation procedure will be misguided (resulting in a significant overestimate of β) given too few transition areas.

Optimization-step. By this time the MRF model is completely specified.

This step involves the optimization of the test image labels with respect to (18)-(21). For the optimization the Modified Metropolis Dynamics algorithm described in Sec. 5 is used.

In case of single polarization channel we have $D = 1$, so there is no need for the Copula-step, because the estimates after DSEM-step can be directly plugged into (21). So, in single-pol case the algorithm goes

$$\text{DSEM-step} \rightarrow \text{MRF-step} \rightarrow \text{Estimation-step} \rightarrow \text{Optimization-step}.$$

5 Optimization

Given the classical optimization choice between Simulated Annealing (SA) [22] and Iterated Conditional Modes (ICM) [20] one has to specify the goal of optimization. If the goal is the best quality, i.e., global minimum of the energy, the choice is a stochastic SA with appropriate cooling procedure. However the drawback is very slow convergence, in fact one doesn't know in advance the timeframe of convergence [22]. If on the other hand the computational time is of significance, one chooses the deterministic ICM procedure which like all greedy algorithms converges fast. The drawback in this case would be the bane of all greedy algorithms - strong dependence on initialization point and convergence to a local minimum. In most cases though both goals are important - speed and quality, thus a compromise solution has to be found. One such solution is Modified Metropolis Dynamics (MMD) algorithm [23]. It is a deterministic algorithm, which needs a predefined cooling schedule and level α, see the MMD algorithm below.

MMD optimization algorithm

1. sample a random initial configuration ω^0, with $k = 0$ and $T = T_0$;
2. using uniform distribution pick up a global state η which differs exactly in one element from ω^k;
3. **(MMD)** Compute $\Delta U = U(\eta) - U(\omega)$ and accept η according to the rule:

$$\omega^{k+1} = \begin{cases} \eta, & \text{if } \Delta U \leq 0, \\ \eta, & \text{if } \Delta U \geq 0 \text{ and } ln(\alpha) \leq -\frac{\Delta U}{T}, \\ \omega^k, & \text{otherwise.} \end{cases} \quad (22)$$

   where α is a constant threshold ($\alpha \in (0, 1)$), chosen at the start of the algorithm;
4. Decrease the temperature $T = T_{k+1}$ and goto Step 2 until convergence ($\Delta U/U$ below a given threshold).

The difference between SA, MMD and ICM can be appreciated looking at (22), specifically at the second case: if removed, one gets deterministic ICM procedure; if α is chosen once and for all at the start of the algorithm, one gets deterministic MMD; and finally, if $\alpha \in (0, 1)$ is randomly chosen on every iteration, the result is the stochastic SA.

MMD can be viewed as consisting of two steps:

- **Pseudo-stochastic phase:** At high temperature, the energy increase is permitted, thus the behavior of the algorithm can be seen similar to stochastic,
- **Deterministic phase:** Once the temperature goes below a certain threshold, only the decrease of energy is allowed, thus becoming deterministic and converging to a local minimum.

Specifying the level α can be interpreted as adjusting the proportion between these two phases.

As discussed in [23], the MMD algorithm can be a favorable choice, especially in case of poor initialization.

6 Experiments

In this section we present experiments with CoDSEM for the application of wet soil mapping. This application originates from epidemiology monitoring and specifically the search for flooded (water) and wet regions where mosquitoes propagate. The purpose is to classify the scenes into three classes - the flooded (water), wet and dry soil areas.

The experiments were done on 3 SAR images:

SAR1. Dual HH/VV Polarized TerraSAR-X, Stripmap (6 m ground resolution), geocoded, 2-look image acquired over Sanchagang, China (11000x22000 pixels); this image is distributed on a free basis at *http://www.infoterra.de/* (©Infoterra);

SAR2. Dual HH/VV Polarized TerraSAR-X, HR Spotlight (1 m ground resolution), not geocoded, single-look image acquired over Barkedji, Senegal (9100x4200 pixels); provided for experiments by the French Space Agency (©CNES-DLR);

SAR3. Single HH Polarized COSMO-SkyMed, Stripmap (2.5 m ground resolution), geocoded, single-look image acquired over Piemonte, Italy (23000x27400 pixels); provided for experiments by the Italian Space Agency (©ASI).

In all experiments a learning image (from the same big image) with manually created ground truth (GT) map is used (Fig. 1-3 in Appendix). Test images are used for testing (Optimization-step) and the resulting classification maps and confusion matrices are reported.

MRF parameter estimation has been applied as in Sec. 3 for the images with exhaustive GT map, in case of nonexhaustive learning images MRF parameter has been set to $\beta = 1.5$. MMD optimization has been used with the following parameters: $\alpha = 0.3$, $T_{k+1} = 0.97\, T_k$ and the temperature is changed once every three iteration, where one iteration is a full raster scan, initial temperature $T_0 = 10.0$, and the threshold for $\Delta U/U$ is 0.0001.

All the experiment (Exp. 1-10) with images and results are shown in Appendix. Here are some comments to the reported experiments:

SAR1

Exp. 1 Image 700x600 pixels. High classification accuracies reported for all classes.

Exp. 2 Image 500x500 pixels. Compared to results in Exp. 1 and Exp. 3 the accuracy demonstrated here is lower, that is due to the attempt to create a more exhaustive GT for the test image resulting in high misclassification for wet/dry soil classes, which is hard to distinguish manually, especially in transition areas

between classes. Thus, to some extent, the decrease of accuracy is explained by a partially inaccurate GT.

Exp. 3 Image 1200x1400 pixels (same for Exp. 4-6). High CoDSEM classification accuracies reported for all classes. For the sake of comparison, we provide classification maps and accuracies acquired by 2D Nakagami-Gamma model [17] (Exp. 3(d)) and K-nearest neighbors [24] (Exp. 3(e)), with $K^* = 40$ estimated by cross-validation. The respective models were combined with MRF approach, same estimation and optimization procedures were used as described in Sec. 3. We stress here the higher classification accuracy achieved by CoDSEM as compared to both parametric density estimation 2D Nakagami-Gamma approach and non-parametric general classification approach K-NN.

Exp. 4 This experiment is interesting to compare with Exp. 3. It demonstrates the benefit of picking copulas for joint distribution modeling, instead of independence assumption (product copula). In general, the average copula improvement is around 1%.

Exp. 5 This experiment has been done only on one polarization channel HH. Here the benefit of using two polarizations is demonstrated. The classification on the same image with two polarizations HH/VV (Exp. 3) reported improvement in accurate of 6%. More exhaustive testing of DSEM and CoDSEM revealed an average gain of $5 - 10\%$ in total classification accuracy from adding the second polarization channel.

Exp. 6 This experiment demonstrates the result of β estimation from an exhaustive GT. The overall accuracy is lower, however like in Exp. 2, the inaccuracy of inaccuracy introduced by a manually created exhaustive GT is rather high.

SAR2 This image presents a more complicated scene. This is not only due to higher resolution, but also because of the more sophisticated landscape on the scene. The image presents the area around an almost dry river bed and the water cannot be seen directly as it is completely covered by tree canopies. Thus the classification cannot be done into three classes, so the resulting two classes are water-forest and wet soil (not covered by tree canopies).

Exp. 7 Image 1000x1000 pixels. Reasonably good classification accuracies reported, even though inferior to that reported for SAR1. This can be explained by the lower separability of histograms for the considered two classes.

Exp. 8 Image 1300x1000 pixels. Image downsampled (by averaging) from initial image 4 to 1, thus increasing the effective number of looks from 1 to 4. This provided an improvement of classification accuracies as compared to results reported in Exp. 7.

SAR3 Image in Exp. 9 is 1000x400 pixels, 800x500 pixels in Exp. 10. As mentioned above the tests here were done on single-pol COSMO-SkyMed. Both Exp. 9 and Exp. 10

report an average classification accuracy comparable with that of Exp. 5 which was also a single-pol experiment. These results suggest the same level of applicability of the developed method to both TerraSAR-X and COSMO-SkyMed, however to prove that more extensive testing is required.

7 Conclusion

We proposed a novel classification algorithm for high and very high resolution SAR combining the Markov random field (MRF) approach to Bayesian image classification and a finite mixture technique for probability density function estimation. The former is a convenient way of taking into account the contextual information. The latter is done by dictionary-based stochastic expectation maximization (DSEM) amplitude histogram estimation approach, which provides a good level of flexibility for the algorithm, regardless of the type of histogram the considered classes may exhibit. The developed algorithm is extended to an important case of multi-polarized SAR by modeling the joint distributions of channels via copulas. Copulas are an attractive choice for multi-dimensional propagation as they provide a wide variety of dependence structures. The resulting classification algorithm is supervised (model estimation is done on a training image) and semiautomatic (several DSEM and MMD parameters have to be specified).

The accuracy of the proposed algorithm is validated in the application of flooded/wet/dry soil classification on high resolution SAR images acquired by TerraSAR-X and COSMO-SkyMed. The experiments demonstrated a high level of accuracy on both types of images.

Acknowledgment

The first author would like to thank INRIA and the French Space Agency (CNES) for funding his work at INRIA as an invited PhD student. Both authors would like to thank Dr. Gabriele Moser and Prof. Sebastiano B. Serpico from the Dept. of Biophysical and Electronic Engineering (DIBE) of the University of Genoa, Italy, for their helpful advice in this research work.

The authors would like to thank the Italian Space Agency (ASI) for providing the COSMO-SkyMed image of Piemonte (©ASI, 2008) and CNES for providing the TerraSAR-X Barkedji image (©CNES-DLR, 2008). The TerraSAR-X image of Sanchagang was downloaded from *http://www.infoterra.de/*, where it is available for free testing (©Infoterra, 2008).

Appendix

The experiments 1-5 provide the results of tests on SAR1.

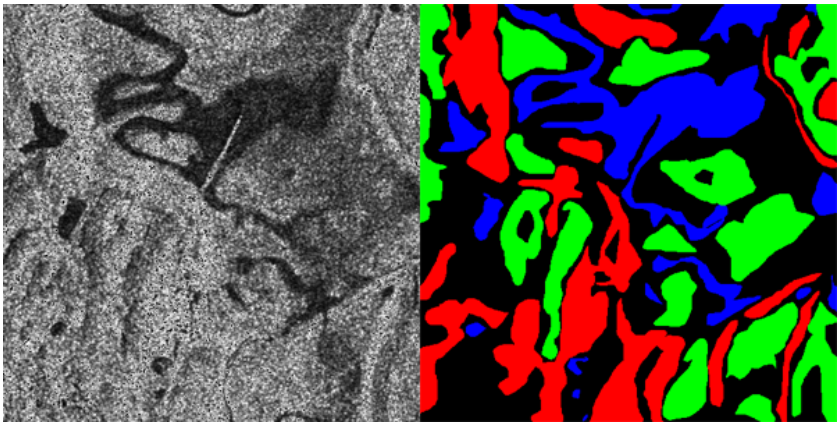
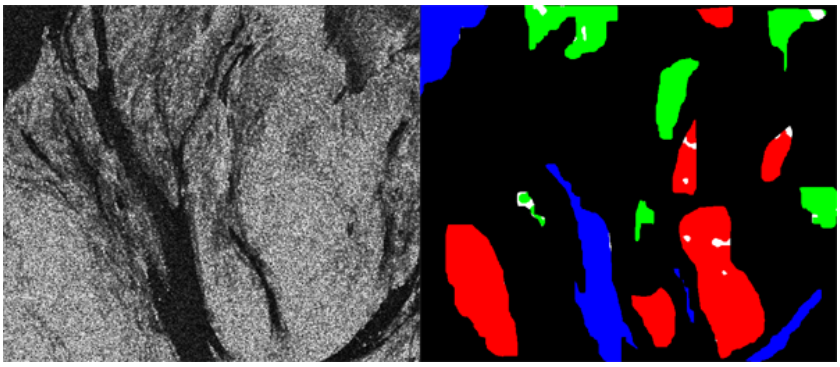
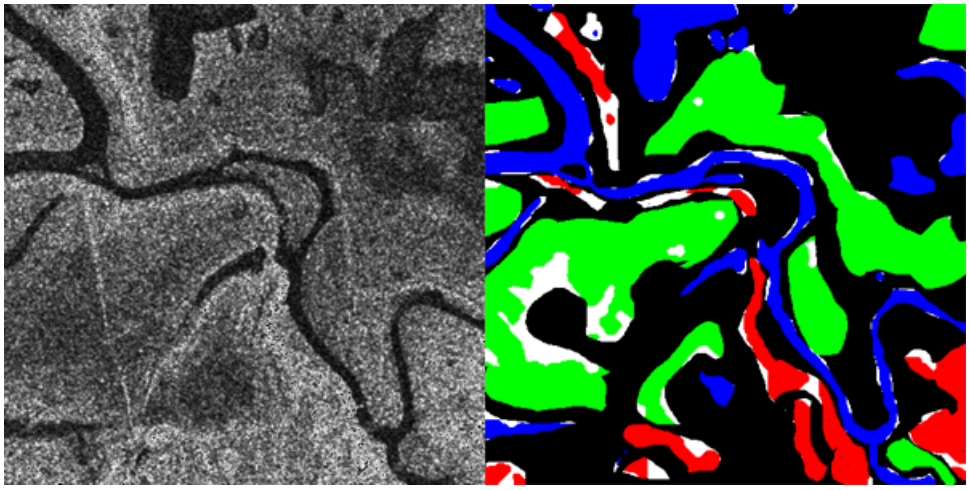


Figure 1: SAR1 Learning image 500x500 pixels. HH polarization (left), partial GT (right): blue - water, green - wet soil, red - dry soil. Copulas: Gumbel, Frank, Frank. It is used for learning in experiments 1-5 (TerraSAR-X, ©Infoterra).



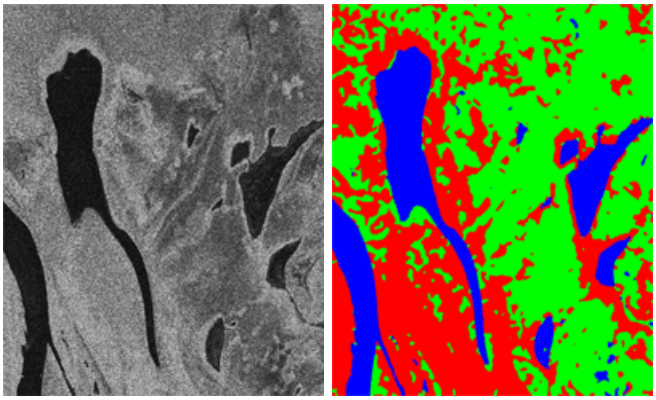
Overall: **97.86%**	**water**	**wet soil**	**dry soil**
water	**99.79%**	0.20%	0.01%
wet soil	3.40%	**94.89%**	1.71%
dry soil	0.01%	1.78%	**98.21%**

Experiment 1: SAR1 Test 1. HH polarization (left), classification (right), confusion matrix (table), where the GT classes are in rows. Black pixels are outside GT, white - misclassification of all types, blue - correct water classification, green - wet soil, red - dry soil (TerraSAR-X, ©Infoterra).

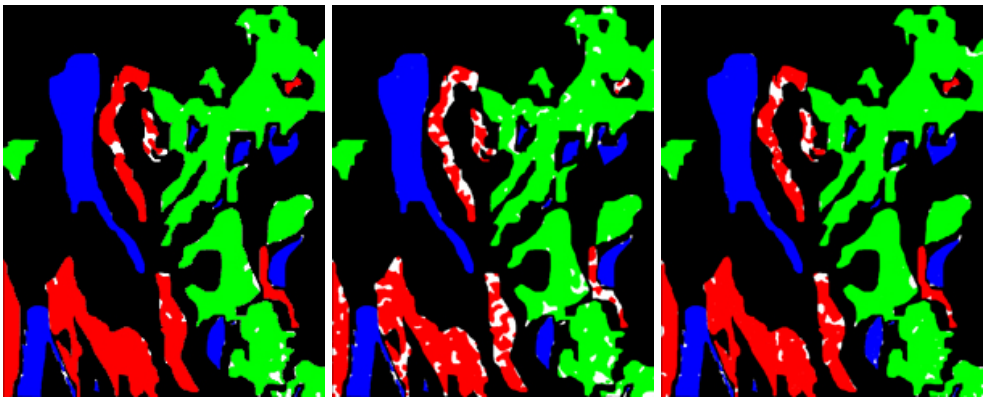


Overall: **88.73%**	**water**	**wet soil**	**dry soil**
water	**93.07%**	6.39%	0.52%
wet soil	9.69%	**90.14%**	0.16%
dry soil	0.41%	22.09%	**77.49%**

Experiment 2: SAR1 Test 2. HH polarization (left), classification (right), confusion matrix (table). Black pixels are outside GT, white - misclassification of all types, blue - correct water classification, green - wet soil, red - dry soil (TerraSAR-X, ©Infoterra).



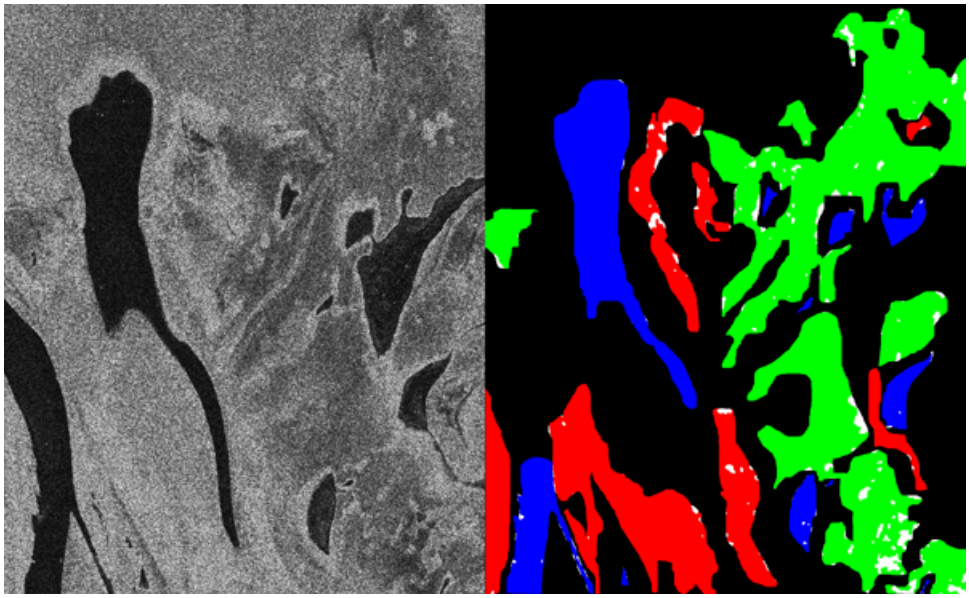
(a) (b)



(c) (d) (e)

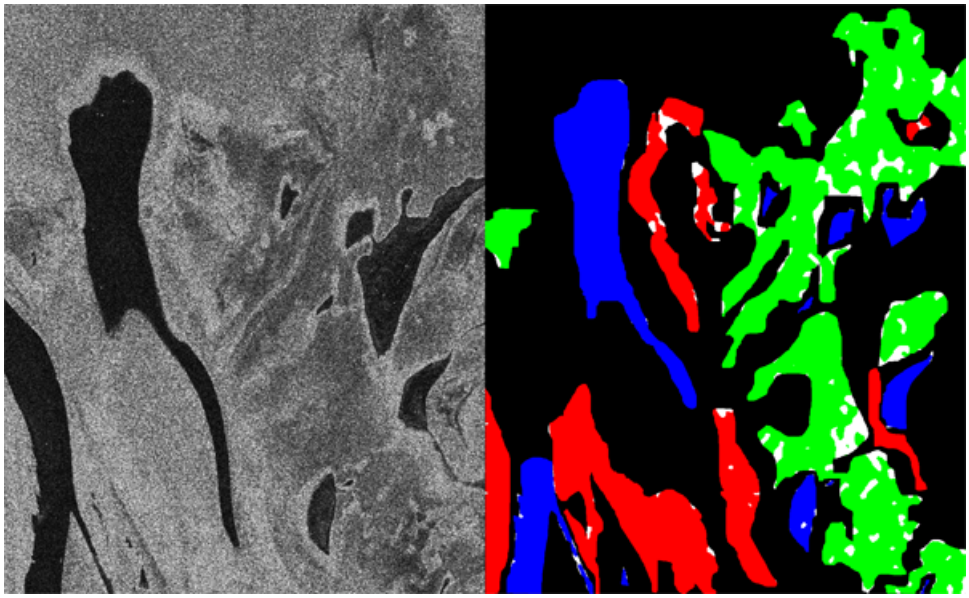
Method	CoDSEM			2D Nakagami-Gamma			K-NN		
Overall	**97.07%**			**90.22%**			**95.92%**		
Class	**water**	wet	dry	**water**	wet	dry	**water**	wet	dry
water	**99.37%**	0.21%	0.40%	**98.75%**	0.92%	0.28%	**98.79%**	0.83%	0.36%
wet	1.00%	**96.83%**	2.15%	1.82%	**94.63%**	3.51%	0.25%	**98.01%**	1.74%
dry	0.09%	4.17%	**95.73%**	0.09%	24.33%	**75.50%**	0.12%	9.07%	**90.80%**

Experiment 3: SAR1 Test 3. HH polarization (a), full CoDSEM-classification map (b) (water - blue, wet soil - green, dry soil - red), CoDSEM-classification map (c) referenced to nonexhaustive GT (black - outside GT, white - misclassification of all types, blue - correct water classification, green - wet soil, red - dry soil), 2D Nakagami-Gamma classification (d), and K-NN classification (e). Confusion matrices for the three classification approaches are presented in the table. (TerraSAR-X, ©Infoterra).



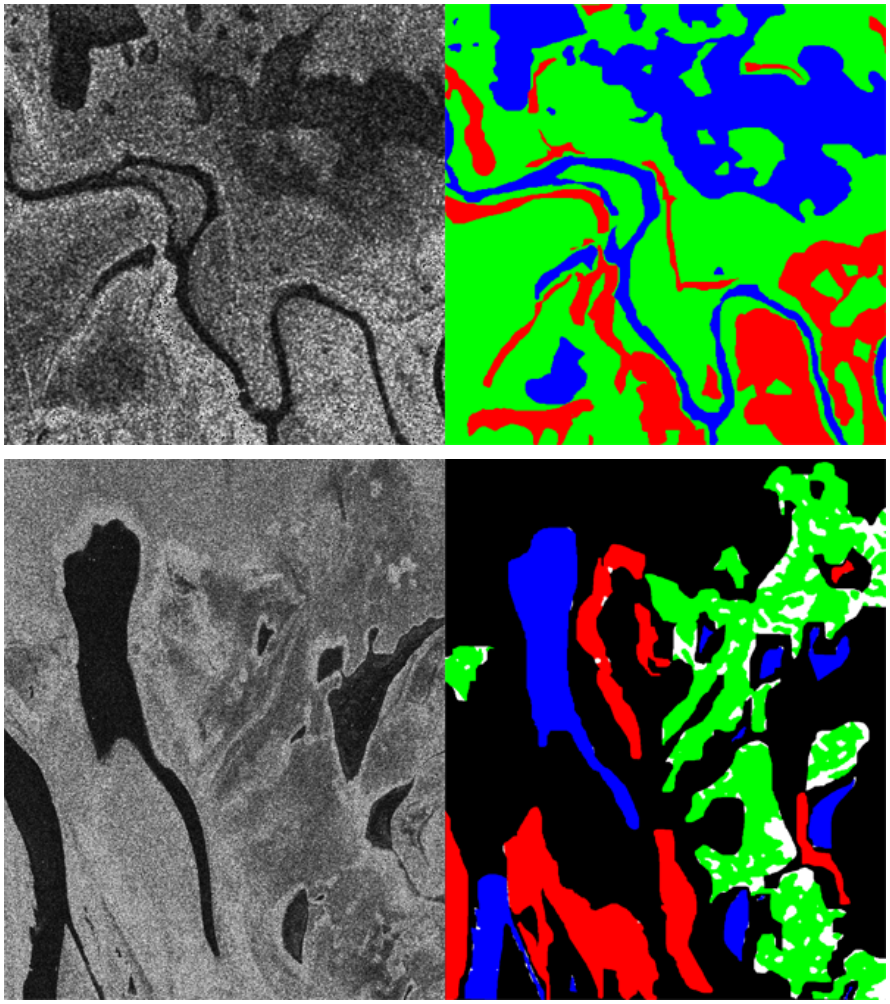
Overall: **96.52%**	**water**	**wet soil**	**dry soil**
water	**99.20%**	0.35%	0.44%
wet soil	2.03%	**94.93%**	3.03%
dry soil	0.10%	2.57%	**97.31%**

Experiment 4: SAR1 Test 4. Test with product copulas. HH polarization (left), classification (right), confusion matrix (table). Black pixels are outside GT, white - misclassification of all types, blue - correct water classification, green - wet soil, red - dry soil (TerraSAR-X, ©Infoterra).



Overall: **93.27%**	**water**	**wet soil**	**dry soil**
water	**99.06%**	0.44%	0.48%
wet soil	0.32%	**88.96%**	10.71%
dry soil	0.08%	3.31%	**96.60%**

Experiment 5: SAR1 Test 5. Test on single HH polarization. HH polarization (left), classification (right), confusion matrix (table). Black pixels are outside GT, white - misclassification of all types, blue - correct water classification, green - wet soil, red - dry soil (TerraSAR-X, ©Infoterra).



Overall: **89.62%**	**water**	**wet soil**	**dry soil**
water	**99.36%**	0.11%	0.52%
wet soil	0.62%	**80.01%**	19.35%
dry soil	0.08%	0.33%	**99.57%**

Experiment 6: SAR1 Test 6. Test with β estimation from the learning image (500x500 pixels) with exhaustive GT ($\hat{\beta} = 1.69$). Learning image (left top), GT (right top), test image HH pol (left bottom), classification (right bottom), confusion matrix (table). Copulas: Gumbel, Gumbel, Gumbel. Black pixels are outside GT, white - misclassification of all types, blue - correct water classification, green - wet soil, red - dry soil (TerraSAR-X, ©Infoterra).

The experiments 7-8 provide the results of tests on SAR2.

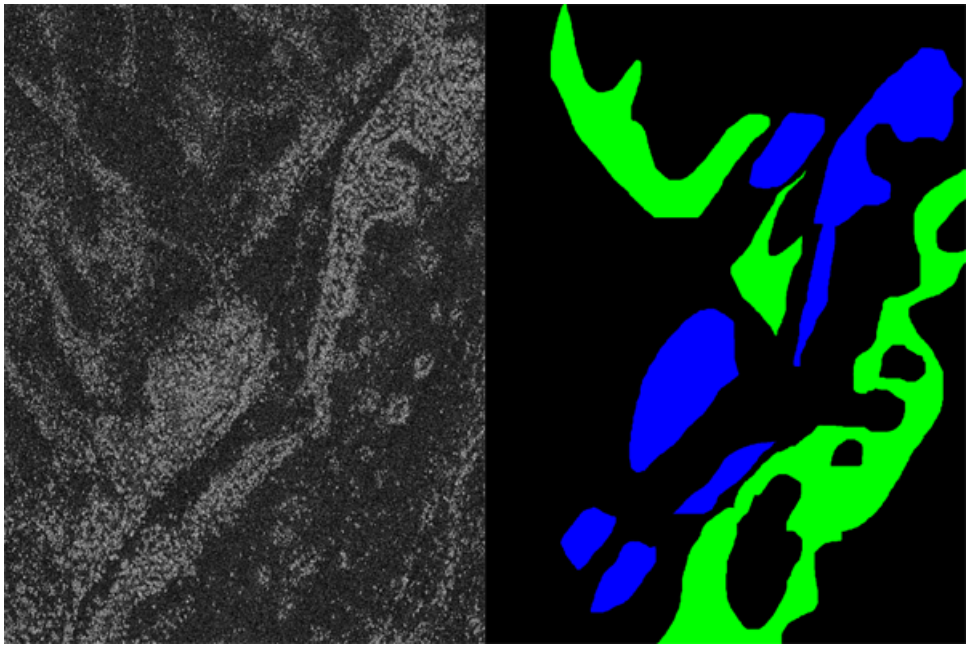
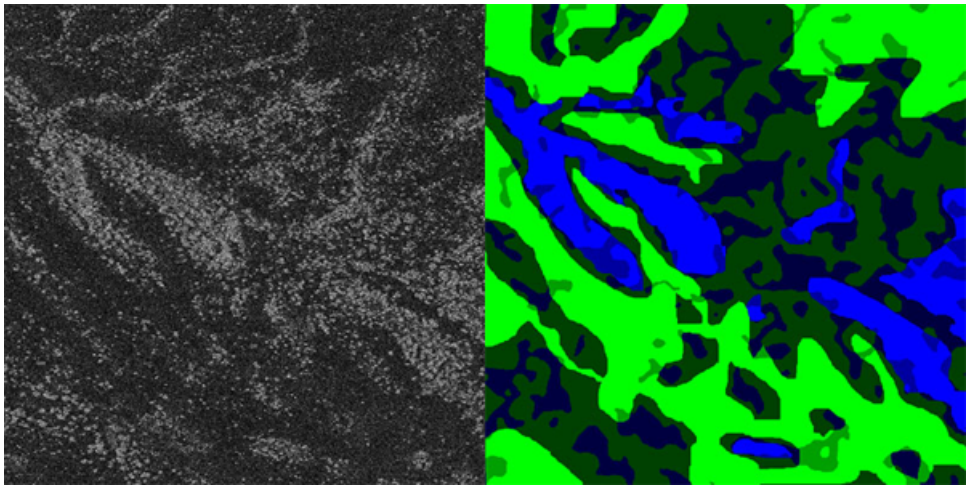
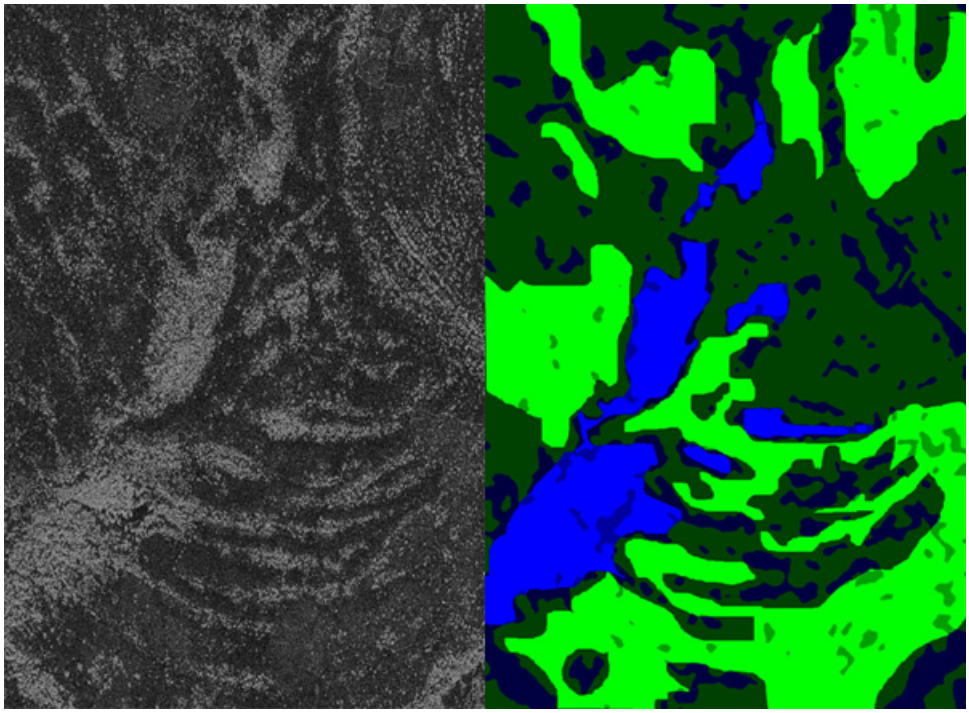


Figure 2: SAR2 Learning image 1000x700 pixels. HH polarization (left), partial GT (right): blue - water-forest, green - wet soil. Copulas: Raftery, Gumbel. It is used for learning in experiments 7-8 (TerraSAR-X, ©CNES).



Overall: **89.17%**	**water-forest**	**wet soil**
water-forest	**81.53%**	18.46%
wet soil	7.61%	**92.39%**

Experiment 7: SAR2 Test 1. HH polarization (left), classification (right), confusion matrix (table). Bright blue pixels show the correct water-forest classification, bright green - wet soil; dark blue - water-forest misclassification and classification outside of defined GT, dark green - same for wet soil (TerraSAR-X, ©CNES).



Overall: **94.07%**	**water-forest**	**wet soil**
water-forest	**89.53%**	10.46%
wet soil	4.59%	**95.41%**

Experiment 8: SAR2 Test 2. Test with downsampling the initial image 4 to 1 thus bringing the number of looks to 4. HH polarization (left), classification (right), confusion matrix (table). Bright blue pixels show the correct water-forest classification, bright green - wet soil; dark blue - water-forest misclassification and classification outside of defined GT, dark green - same for wet soil (TerraSAR-X, ©CNES).

The experiments 9-10 provide the results of tests on SAR3.

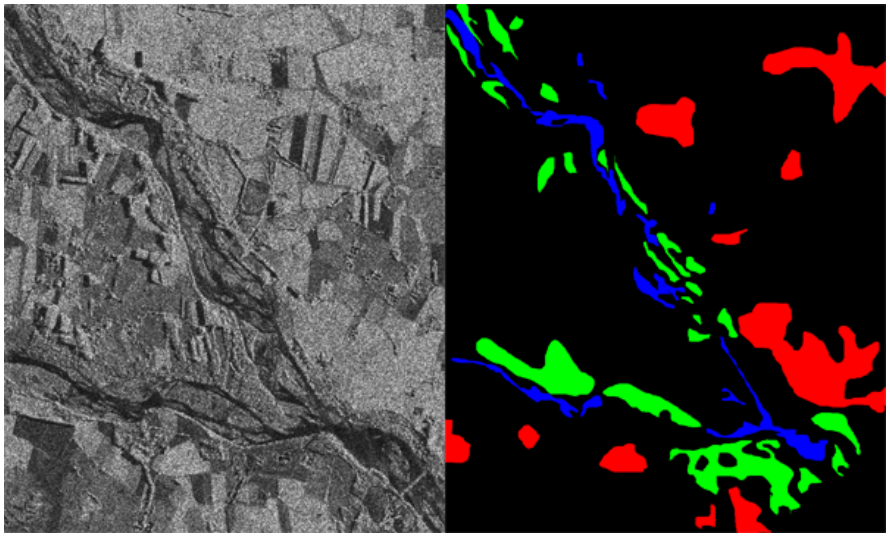
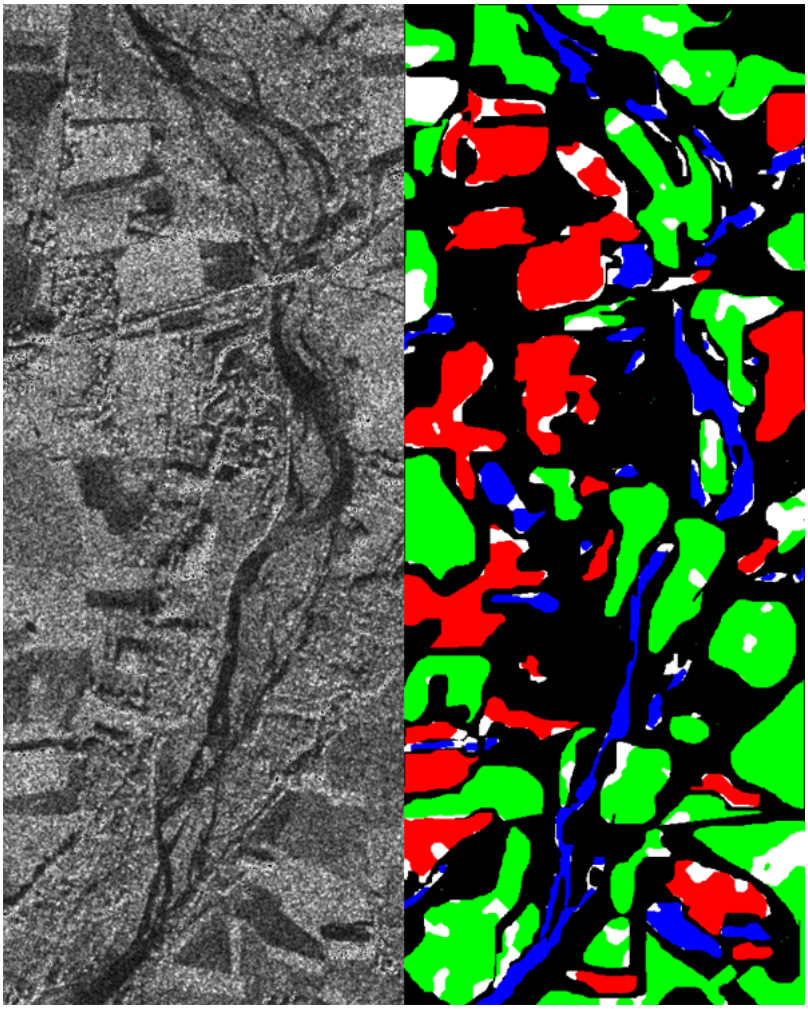
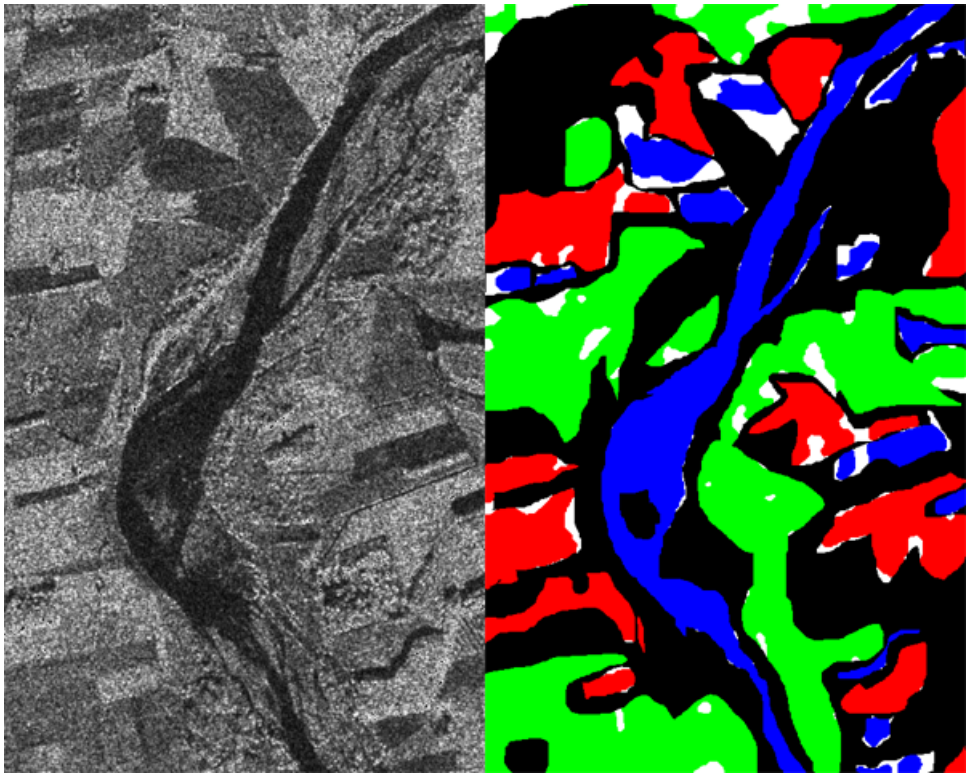


Figure 3: SAR3 Learning image 1200x1000 pixels. HH polarization (left), partial GT (right): blue - water, green - wet soil, red - dry soil. It is used for learning in experiments 9-10 (COSMO-SkyMed, ©ASI).



Overall: **87.42%**	**water**	wet soil	dry soil
water	**86.10%**	13.27%	0.61%
wet soil	5.73%	**86.07%**	8.18%
dry soil	0.04%	9.53%	**90.42%**

Experiment 9: SAR3 Test 1. HH polarization (left), classification (right), confusion matrix (table). Black pixels are outside GT, white - misclassification of all types, blue - correct water classification, green - wet soil, red - dry soil (COSMO-SkyMed, ©ASI).



Overall: **91.36%**	**water**	**wet soil**	**dry soil**
water	**86.66%**	13.20%	0.12%
wet soil	0.60%	**92.21%**	7.17%
dry soil	0.01%	5.71%	**94.27%**

Experiment 10: SAR3 Test 2. HH polarization (left), classification (right), confusion matrix (table). Black pixels are outside GT, white - misclassification of all types, blue - correct water classification, green - wet soil, red - dry soil (COSMO-SkyMed, ©ASI).

References

[1] Z. Kato, J. Zerubia, and M. Berthod. Unsupervised parallel image classification using Markovian models. *Pattern Recognition*, 32(4):591–604, 1999.

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

[3] 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.*, 42(10):2046–2057, 2004.

[4] E. E. Kuruoglu and J. Zerubia. Modelling SAR images with a generalization of the Rayleigh distribution. *IEEE Trans. Image Process.*, 13(4):527–533, 2004.

[5] G. Moser, J. Zerubia, and S. B. Serpico. SAR amplitude probability density function estimation based on a generalized Gaussian model. *IEEE Trans. Image Process.*, 15(6):1429–1442, 2006.

[6] H.-C. Li, W. Hong, and Y.-R. Wu. Generalized Gamma distribution with MoLC estimation for statistical modeling of SAR images. In *Proceedings of APSAR*, pages 525–528, Huangshan, China, 2007.

[7] R. B. Nelsen. *An Introduction to Copulas*. Springer, New-York, 2nd edition, 2007.

[8] G. Moser, S. Serpico, and J. Zerubia. Dictionary-based Stochastic Expectation Maximization for SAR amplitude probability density function estimation. *IEEE Trans. Geosci. Remote Sens.*, 44(1):188–199, 2006.

[9] V. Krylov, G. Moser, S. Serpico, and J. Zerubia. Dictionary-based probability density function estimation for high-resolution SAR data. In *Proceedings of SPIE*, volume 7246, page 72460S, San Jose, USA, 2009.

[10] V. Krylov, G. Moser, S. Serpico, and J. Zerubia. Modeling the statistics of high resolution SAR images. Research Report 6722, INRIA, 2008.

[11] R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood, and the EM algorithm. *SIAM Review*, 26(2):195–239, 1984.

[12] M. A. F. Figueiredo and A. K. Jain. Unsupervised learning of finite mixture models. *IEEE Trans. Patt. Anal. Mach. Intell.*, 24(3):381–396, 2002.

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

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

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

[16] J.-M. Nicolas and A. Maruani. Lower-order statistics: a new approach for probability density functions defined on $\mathbb{R}^+$. In *Proceedings of EUSIPCO*, Tampere, Finland, 2000.

[17] J.-S. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller. Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery. *IEEE Trans. Geosci. Remote Sens.*, 32(5):1017–1028, 1994.

[18] D. Huard, G. Évin, and A.-C. Favre. Bayesian copula selection. *Computational Statistics & Data Analysis*, 51(2):809–822, 2006.

[19] E. L. Lehmann and J. P. Romano. *Testing statistical hypotheses*. Springer, New York, 3rd edition, 2005.

[20] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B*, 48:259–302, 1986.

[21] Y. Yu and Q. Cheng. MRF parameter estimation by an accelerated method. *Pattern Recognit. Lett.*, 24(9-10):1251–1259, 2003.

[22] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intell.*, 6:721–741, 1984.

[23] Z. Kato, J. Zerubia, and M. Berthod. Satellite image classification using a modified Metropolis dynamics. In *Proceedings of ICASSP*, pages 573–576, 1992.

[24] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New-York, 2006.

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