International Journal of Remote Sensing
Vol. 31, No. 3, 10 February 2010, 727–744

Taylor & Francis
Taylor & Francis Group

Classifying polarimetric SAR data by combining expectation methods with spatial context

A. REIGBER*†, M. JÄGER‡, M. NEUMANN§‡ and L. FERRO-FAMIL§
†Microwaves and Radar Institute, German Aerospace Center (DLR),
Oberpfaffenhofen, Germany
‡Computer Vision and Remote Sensing Group, Berlin
University of Technology, Berlin, Germany
§Institute of Electronics and Telecommunications of Rennes,
University of Rennes 1, Rennes, France

(*Received 26 January 2007; in final form 19 May 2008*)

Unsupervised classification is an essential step in the automatic analysis of SAR remote sensing data. Classification results make SAR data easier to interpret and can serve as a starting point for automated analysis techniques that apply to homogeneous regions of the observed scene. Polarimetric SAR data are particularly interesting for unsupervised classification purposes, since they contain a great amount of information, allowing robust statistical clustering of the image content on the one hand and a direct physical interpretation of the result on the other.

This paper proposes a new unsupervised classification approach for polarimetric SAR data. Assuming Wishart-distributed polarimetric covariance matrices, it combines spectral clustering based on the covariance matrices themselves with spatial clustering by statistical analysis of local neighbourhoods. Instead of working with binary assignments of samples to class centres, a soft decision rule is used in which each pixel is assigned to all class centres in the spectral and spatial domains. The local neighbourhood is taken into account by altering the probabilities of class membership by a neighbourhood function, obtained from normalized compatibility coefficients, describing cluster sizes and mutual tolerance. In this way, robust and homogenous classification results can be obtained even in the presence of strong speckle noise.

1. Introduction

Unsupervised classification is an important technique for the automatic analysis of synthetic aperture radar (SAR) data. Polarimetric SAR (PolSAR) data in particular are appealing for this purpose, since they allow sophisticated classification based on the analysis of multiple polarimetric channels. In the literature, many unsupervised classification approaches for polarimetric SAR data have been proposed. One type of algorithms is based on the analysis of physical scattering properties, which has the advantage that some information about class types is available. van Zyl (1989) derived a classification scheme which splits the scene into the three main classes of odd bounce, even bounce and volume scattering. A more advanced approach can be found in Cloude and Pottier (1997), where target decomposition theory is used to calculate the physical scattering parameters: entropy H and α-angle. These two parameters span a two-dimensional feature space,

*Corresponding author. Email: andreas.reigber@dlr.de

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160902897809

which is then segmented into eight classes using physically motivated thresholds. A further refinement can be found in Ferro-Famil *et al.* (2001), where the polarimetric anisotropy A is introduced to fully characterize the eigenvalue spectrum. In this way, 16 classes could be derived by a threshold-based partition in the H-A-α feature space, used in this case for initialization of another iterative classification scheme.

A different type of unsupervised classification algorithm is based purely on statistical clustering. This has the advantage that classes which do not perfectly align with pure physical scattering mechanisms can be identified. Instead, objects with an arbitrary but similar backscattering are grouped. Multi-look polarimetric SAR data are represented by polarimetric covariance matrices and follow the complex Wishart-distribution (Goodman 1963). Based on this distribution, a statistical distance measure can be derived (Lee *et al.* 1994a), which is then used to segment the data in the space of covariance matrices using standard clustering techniques like ISODATA (Ball and David 1965). Several variants of this approach, using expectation maximization techniques (Dempster *et al.* 1977, Davidson *et al.* 2002) or, closely related, fuzzy decision rules (Du and Lee 1996, Chen *et al.* 2003) are found in the literature. Since pure statistical clustering does not consider physical scattering properties, an interesting alternative is the combination of statistical clustering with physically motivated techniques. Lee *et al.* (1999) use a H-α pre-segmentation to initialize an iterative Wishart-clustering in the feature space of covariance matrices. Ferro-Famil *et al.* (2001) extend this idea and also include the anisotropy and some improvements to the iteration scheme. Finally, Lee *et al.* (2004a) propose a statistical clustering scheme, which avoids class assignments that are inconsistent with the dominant polarimetric scattering properties.

However, one disadvantage of all these algorithms is that each pixel is treated independently of its neighbours; spatial context is only indirectly considered during speckle filtering. The local neighbourhood does indeed have a significant influence on a pixel's class membership: when a certain region has already been classified, with high confidence, as belonging to a single class, it becomes comparatively unlikely that a pixel in this region belongs to another class. The much more likely scenario is a misestimation of its covariance matrix due to speckle noise. In fact, as a matter of principle, the noise level of SAR data is always relatively high, even when using sophisticated speckle filtering techniques (Vasile *et al.* 2006, Lee *et al.* 2004b) and classification results often appear inconsistent and noisy. Therefore, the inclusion of local neighbourhoods in the statistical decision concerning class membership is reasonable in view of encouraging homogenous classification results and resolving uncertainties.

A common way to include contextual information in the classification process is to model the labelling process as a Markov random field (MRF). An MRF takes into account a statistical model linking observations to class labels as well as statistical dependencies among neighbouring labels (Chellappa and Jain 1993). In general, Markov random fields establish a spatial adjacency graph between the image pixel and the hidden layer of class labels, whose *a posteriori* probability is maximized by an optimization strategy such as simulated annealing (Hegarat-Mascle *et al.* 1996). MRFs have been widely used for classification of remote sensing data; some newer applications in the context of SAR data can be found, for example, in Fjortoft *et al.* (2003) and Tison *et al.* (2004). Markov random fields are effective for segmentation, but are computationally very expensive to optimize due to the complicated structure of their likelihood function: the function to be maximized does not factorize in a way that individual samples can be treated independently, therefore it needs to cope with an extremely highly dimensional configuration space due to cyclical label dependencies.

An alternative approach for incorporating spatial context is so-called probabilistic label relaxation technique (PLR). As with MRF, it aims to achieve a spatially consistent classification result by incorporating neighbourhood relations, but uses a simpler strategy based on the minimization of costs for transition between different class labels, not taking into account the image data themselves (Townsend 1986). Originally, probabilistic label relaxation was applied as a post-processing step for improving the results of supervised classification algorithms (Harris 1985). However, as described in this paper, PLR can also be successfully applied in the framework of an unsupervised classification scheme for polarimetric SAR data based on the statistical partitioning of the covariance matrix feature space.

This paper is organized as follows: section 2.2 gives a brief overview of spectral clustering with the expectation maximization technique, a method which uses only the statistical properties of the polarimetric covariance matrix. Section 2.3 introduces the concept of probabilistic label relaxation, i.e. clustering using only local neighbourhoods and without considering polarimetric information. In section 2.4, these two concepts are combined into a novel unsupervised classification procedure, which is able to derive homogenous classification results, even from minimally speckle-filtered SAR data. Section 3 demonstrates the potential of the proposed method on real SAR data acquired by DLR's experimental E-SAR sensor and evaluates the results obtained with reference to ground-truth data of an agricultural scene. Finally, section 4 discusses and summarizes the presented methods and results.

2. Neighbourhood-supported polarimetric classification

2.1 *Relative class probabilities*

The backscattering of a monostatic polarimetric SAR system is characterized by the complex scattering vector

$$\boldsymbol{k} = [S_{\mathrm{HH}}, \sqrt{2}S_{\mathrm{HV}}, S_{\mathrm{VV}}] \tag{1}$$

whose elements S_{HH}, S_{HV} and S_{VV} 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. Usually, polarimetric data are transformed into the form of covariance matrices $\mathbf{C}$ when analysing second order statistics or reducing speckle noise:

$$\mathbf{C} = \langle \boldsymbol{k}^{\dagger}\boldsymbol{k} \rangle_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{k}^{\dagger}\boldsymbol{k} \tag{2}$$

with † denoting the adjoint operator. $\mathbf{C}$ stands for the n-look sample covariance matrix, i.e. n denotes the number of independent samples used for averaging. For a homogenous region ω_i with Gaussian backscattering, characterized by a covariance matrix $\Sigma_i = E(\mathbf{C}|\mathbf{C} \in \omega_i)$, $\mathbf{C}$ follows the complex Wishart-distribution (Lee *et al.* 1994a,b)

$$p(\mathbf{C}|\Sigma_i) = \frac{n^{qn}|\mathbf{C}|^{n-q}\exp[-n\mathrm{Tr}(\Sigma_i^{-1}\mathbf{C})]}{|\Sigma_i|^n \pi^{q(q-1)/2}\prod_{j=1}^{q}\Gamma(n-j+1)} \tag{3}$$

with $E(\ldots)$ denoting the expectation value, q the dimensionality of $\mathbf{C}$ (here 3), $\mathrm{Tr}(\ldots)$ the trace of a matrix, $|\ldots|$ the determinant operator and $\Gamma(\ldots)$ the Gamma-function.

In classification problems, a dataset consists of several classes, each associated with a different Σ_i. Following Bayes' theorem (Freund and Walpole 1987), an unknown pixel is assigned to the class ω_i, which fulfils

$$\mathbf{C} \in \omega_i \quad \text{if} \quad p(\mathbf{C}|\Sigma_i)p(\Sigma_i) > p(\mathbf{C}|\Sigma_j)p(\Sigma_j) \qquad \forall j \neq i. \tag{4}$$

$p(\Sigma_i)$ are the *a priori* probabilities of occurrence of class ω_i, which are unknown and assumed to be uniform. Furthermore, given a fixed set of M classes, each covariance matrix $\mathbf{C}$ can be assigned to one of these classes by considering the normalized probability of class membership:

$$p(\mathbf{C} \in \omega_i) = \frac{p(\mathbf{C}|\Sigma_i)}{\sum_{j=1}^{M} p(\mathbf{C}|\Sigma_j)}. \tag{5}$$

Class membership can be determined as in equation (4). If a binary assignment is desired, usually the most probable class is selected in the classification process. One possibility to incorporate neighbourhoods in the decision process is to modify these class memberships to take into account classification results in a local neighbourhood. Such an algorithm has to decide whether, instead of the class with the highest relative membership, another choice, which is more compatible with the local environment, is appropriate. Ideally, such an approach leads to much more homogenous classification results than algorithms based on $p(\mathbf{C}|\Sigma_i)$ alone (Cloude and Pottier 1997, Lee *et al.* 1999).

2.2 *Classification via expectation maximization*

In expectation maximization (EM), each pixel is assigned, with different degrees of class membership, to all possible classes in a way that maximizes the posterior probability of the assignment with respect to a mixture model describing the constellation of M classes. This is significantly different from unsupervised classification in the classical k-means sense, where each pixel is assigned to only one, namely the most likely, class (Lee *et al.* 1999, 2004a). In the literature, a weighted averaging on the basis of probabilities, as opposed to using fixed assignments, is often also referred to as a *fuzzy* decision (Zadeh 1975) and, in fact, expectation maximization is very closely related to fuzzy classification approaches (Du and Lee 1996, Chen *et al.* 2003). In general, it has the advantage that no binary decisions about class memberships have to be made and all possible assignments can be considered in parallel (Bezdek 1981). In the following, a simple iterative classification scheme will be described, which is based on the principles of expectation maximization and serves as the basis for including spatial context in the classification process.

In an expectation maximization classification algorithm, one attempts to assign N pixels to M different classes, while the optimal set of class centres $\Sigma = (\Sigma_1, \ldots, \Sigma_M)$ remains to be found by the algorithm. The number of classes M is usually assumed to be known, or has to be determined by suitable algorithms. For this purpose, one defines the log-likelihood function, the joint sample likehood conditioned upon a set of class centres:

$$\mathcal{L}(\Sigma) = \ln \prod_{i=1}^{N} p(\mathbf{C}_i|\Sigma) = \sum_{i=1}^{N} \ln p(\mathbf{C}_i|\Sigma). \tag{6}$$

The EM algorithm then optimizes $\mathcal{L}$. In case of the assumed mixture of M different probability density functions, the sample likelihood can be written as

$$p(\mathbf{C}_i|\Sigma) = \sum_{j=1}^{M} p(\mathbf{C}_i|\Sigma_j). \tag{7}$$

For $p(\mathbf{C}_i|\Sigma_j)$, the Wishart-distribution of equation (3) is appropriate. Grouping all constant and class independent terms in the factor $K(\mathbf{C}_i)$, one can express this probability as

$$p(\mathbf{C}_i|\Sigma_j) = K(\mathbf{C}_i)\exp\left(-n\mathrm{Tr}(\Sigma_j^{-1}\mathbf{C}_i)\right)/|\Sigma_j|^n. \tag{8}$$

Substituting equation (8) and equation (7) in equation (6), and ignoring the constant term dependent on $K(\mathbf{C})$, one obtains a log-likelihood function of

$$\mathcal{L}(\Sigma) = \sum_{i=1}^{N} \ln \sum_{j=1}^{M} \exp\left(-n\mathrm{Tr}(\Sigma_j^{-1}\mathbf{C}_i)\right)/|\Sigma_j|^n. \tag{9}$$

In the framework of expectation maximization it is now assumed that hidden variables y_i exist, which map observed covariance matrices $\mathbf{C}_i$ to class centres Σ_j. The y_i correspond, in principle, to fixed assignments: if, for instance, $y_i = 2$, then $\mathbf{C}_i$ belongs to class 2. However, since these assignments are unknown, expectation maximization considers y_i as random variables and introduces the corresponding probability density functions $p(y_i\,|\,\mathbf{C}_i, \Sigma) = p(\mathbf{C}_i \in \omega_i)$. Initially, Σ is also unknown. However, with an initial estimation of one of these two parameters, y or Σ, an optimization of $\mathcal{L}(\Sigma)$ by gradient ascent becomes possible.

The EM algorithm starts with an initial guess of a sub-optimal set of class centres $\Sigma^{(0)}$. To do so, as maximum-liklihood estimates of the unknown real Σ_j, the sample means

$$\Sigma_j^{(0)} = \frac{1}{N_j}\sum_{i\in\omega_j} \mathbf{C}_i \qquad \text{with} \qquad N_j = \text{number of pixels in } \omega_j \tag{10}$$

are estimated over M regions ω_j, $1 \le j \le M$, in the scene. Initial seed regions can be determined in one of several ways: they can be defined manually, in the form of training sets, automatically by an H-α pre-classification (Lee *et al.* 1999), or even by a random assignment of pixels to one of the M classes.

Subsequently, one estimates, in the so-called *expectation* step, the *a posteriori* probabilities $p(y_i = j|\mathbf{C}_i, \Sigma^{(0)})$ for each pixel and class, i.e. the probability that a pixel belongs to class j, given its covariance $\mathbf{C}_i$ and a set of class centres $\Sigma^{(0)}$. In the case of polarimetric data, this is achieved by substituting equation (8) into equation (5). Denoting $p_{ij} = p(y_i = j|\mathbf{C}_i, \Sigma_j) = p(\mathbf{C}_i \in \omega_j)$ one then obtains

$$p_{ij}^{(k)} = \frac{p(\mathbf{C}_i|\Sigma_j^{(k)})}{\sum_{l=1}^{M} p(\mathbf{C}_i|\Sigma_l^{(k)})} = \frac{\exp\left(-n\mathrm{Tr}\left((\Sigma_j^{(k)})^{-1}\mathbf{C}_i\right)\right)/|\Sigma_j^{(k)}|^n}{\sum_{l=1}^{M}\exp\left(-n\mathrm{Tr}\left((\Sigma_l^{(k)})^{-1}\mathbf{C}_i\right)\right)/|\Sigma_l^{(k)}|^n}. \tag{11}$$

The superscript k indicates the current iteration step; after initialization $k = 0$ and $\Sigma = \{\Sigma_1^{(0)}, \Sigma_2^{(0)}, \ldots\}$. The definition of p_{ij} according to equation (5) implies that

$$p_{ij} \in [0,1] \quad \forall i,j \qquad \text{and} \qquad \sum_{j=1}^{M} p_{ij} = 1 \quad \forall i. \tag{12}$$

Using the values obtained for p_{ij}, an updated set of class centres can be computed in the subsequent *maximization* step:

$$\Sigma_j^{(k+1)} = \frac{\sum_{i=1}^{N} p_{ij}^{(k)} \mathbf{C}_i}{\sum_{i=1}^{N} p_{ij}^{(k)}}. \tag{13}$$

In the case of the simplified Wishart-distribution of equation (8), equation (11) and equation (13) together lead to an improvement of the log-likelihood with respect to $\Sigma_j^{(k)}$ and assignments y (Bezdek 1981, Du and Lee 1996, Chen *et al.* 2003). To reach a local maximum, the expectation and maximization steps are carried out iteratively until a certain termination criterion is met. This could be the convergence of p_{ij} between two subsequent iterations (Chen *et al.* 2003), i.e.

$$\frac{1}{N}\sqrt{\sum_{ij}\left(p_{ij}^{(k)} - p_{ij}^{(k-1)}\right)^2} < \text{threshold}. \tag{14}$$

Alternatives are the convergence of the class centres themselves (Du and Lee 1996), the percentage of pixels changing their most likely class between iterations falling below a certain threshold, or simply a fixed number of iterations.

Figure 1 shows a quantitative comparison of the expectation classification technique with threshold classification of the H-α feature space (Cloude and Pottier 1997) and Wishart k-means classification (Lee *et al.* 1999). Covariance matrices were obtained using a mean filter with only nine independent samples, which is usually not sufficient for speckle reduction. For Wishart and EM classification, a random initialization and 10 iterations were used. Since the EM algorithm does not assign pixels to a single class, the colour in figure 1(*d*) was chosen according to the most likely class of each pixel (see equation 4). As expected, all classification results appear very noisy due to insufficient speckle filtering. Over forested areas, EM seems to deliver slightly more homogenous results, while H-α results appear slightly more homogenous over agricultural areas. In general, both results are quite similar. However, EM additionally delivers information concerning the probability with which a pixel falls into a class other than the most likely one. This is important when considering the introduction of probabilistic relaxation techniques.

2.3 *Probabilistic label relaxation*

In a neighbourhood-supported classification, the classification result obtained in a pixel's surrounding is used to resolve uncertainties in class membership where the local covariance is ambiguous. This idea is based on the assumption that two neighbouring pixels are not entirely statistically independent: in reality, spatially random classification results are not very likely; instead, continuous areas of certain sizes are to be expected. If the direct surroundings of a pixel are already classified, with high confidence, into a certain class ω_j, it becomes more likely that the observed pixel also falls into class ω_j.

This circumstance is the basis for the probabilistic label relaxation technique. The starting point is the introduction of so-called *a priori* compatibility coefficients $p(y_i|y_j)$: the conditional probability that a pixel falls into class ω_i if a neighbouring pixel belongs to class ω_j. The joint probability of pixel n having class label y_i and pixel m having class label y_j can then be written as

$$p(n, y_i; m, y_j) = p(y_i|y_j)p(m, y_j) \tag{15}$$

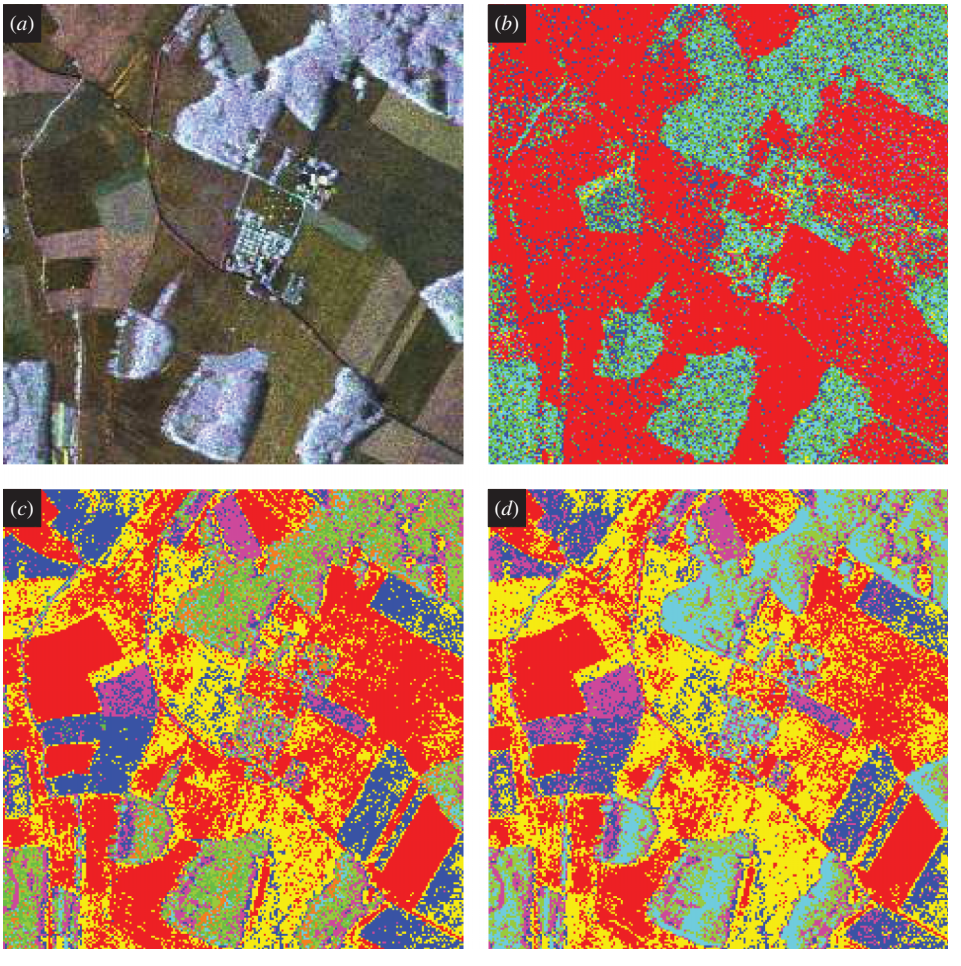

Figure 1. (*a*) Image amplitude in Pauli-decomposition (HH = red, VV = green, HV = blue, nine looks). (*b*) Entropy-alpha classification. (*c*) Classification into eight classes with Wishart *k*-means. (*d*) Classification into eight classes with Wishart expectation maximization.

i.e. the probability of the neighbourhood pixel's class assignment (derived by other means), multiplied with its mutual compatibility. The joint probability $p(n, y_i; m, y_j)$ in equation (15) gives information about class membership of pixel n solely by examination of its neighbourhood and without considering the content of the pixel itself. In general, M possible class assignments are possible; furthermore, it is possible to consider an arbitrary neighbourhood of pixel n consisting of L pixels. Based on this, a neighbourhood function

$$q(n, y_i) = \sum_{m=1}^{L} \sum_{j=1}^{M} p(y_i|y_j) p(m, y_j) \tag{16}$$

can be defined, which represents the cumulative influence of the adjacent pixel labels onto the class assignment of pixel n. It has to be noted that $q(n, y_i)$ is not a probability function: the additive structure of equation (16) relates to *costs* for class

transitions between pixels (Harris 1985), even though $p(y_i|y_j)$ can be interpreted as a probability. At this point, PLR differs significantly from the concept of MRF where joint probabilities are used. However, PLR allows us in a simple way directly to consider long-range effects by computing equation (16) over a large neighbourhoods with lower weighting of farther neighbours. Additionally, equation (16) can be formulated as a convolution, which leads to a computationally very efficient implementation.

After the expectation maximization procedure, the probabilities of class membership, derived from the polarimetric covariance matrices and the current set of model parameters (according to equation 8) are known: $p(m, y_j) = p(\mathbf{C}_m|\Sigma_j)$, and equation (16) can be evaluated. This results in two kinds of initial evidence for class membership: one, $q^{(0)}(n, y_i)$, based only on the spatial neighbourhood and calculated as in equation (16), and the other, $p_q^{(0)}(n, y_i) = p(\mathbf{C}_n|\Sigma_i)$, based on the local observation only. Combining both pieces of evidence, a new probability measure for determining class membership can be defined:

$$p_q^{(h+1)}(n, y_i) = \frac{p_q^{(h)}(n, y_i) q^{(h)}(n, y_i)}{\sum_{j=1}^{M} p_q^{(h)}(n, y_j) q^{(h)}(n, y_j)}. \tag{17}$$

PLR uses an iterative process to adjust the conditional probabilities $p_q^{(h)}(n, y_i)$ and the neighbourhood function: in a subsequent step, $p_q^{(h)}(n, y_i)$ is used to compute an update of $q^{(h)}(n, y_i)$:

$$q^{(h+1)}(n, y_i) = \sum_{m=1}^{L} \sum_{j=1}^{M} p(y_i|y_j) p_q^{(h)}(m, y_j) \tag{18}$$

with h indicating the current iteration step. The iterations over equations (17) and (18) are repeated several times until an acceptable convergence of p_q is reached. Each iteration enlarges the size of the effectively considered neighbourhood. This results in long-range effects and prevents complete convergence of the process. Therefore, and due to efficiency reasons, a termination after $H = 5$–10 iterations is reasonable.

The question of how the compatibility coefficients are to be determined remains. Ideally, a spatial model of the area under investigation is known, e.g. derived from a geoinformation system. In agricultural areas, certain assumptions about the probability of classification inhomogeneities may be permissable when the sensor resolution and characteristic field sizes are known. In general, however, it has to be assumed that the compatibility coefficients are unknown. In this case, it is reasonable to determine all compatibility coefficients from a single value α by

$$p(y_i|y_j) = \begin{cases} \alpha & \text{if} \quad y_i = y_j \\ 1 - \alpha & \text{if} \quad y_i \neq y_j \end{cases} \qquad \text{with} \quad 0 \leq \alpha \leq 1. \tag{19}$$

The value of α quantifies how much more probable a meeting between equal classes is than a meeting between different classes. A high value of α, for example, strongly encourages homogenous areas in the classification result. In the following, the more descriptive ratio $\alpha/(1 - \alpha)$ is used for denoting the strength of the compatibility coefficients.

2.4 *Combined approach*

In the context of classifying polarimetric SAR data, the usage of probabilistic label relaxation as a post-processing step on a result like that in figure 1 is not very effective. The poor classification quality of classifiers like expectation maximization, caused by the speckle effect inherent to SAR data, leads to an inaccurately estimated set of class covariances Σ. A subsequent relaxation process would indeed homogenize the classification result by incorporating spatial context, but is not able to correct the suboptimal set of class centres. Therefore, it is advisable to integrate the described relaxation process directly into the expectation maximization iterations in order to ensure a 'soft' harmonization of the two, possibly competing, terms of polarimetric and spatial information. Such an approach improves the estimation of class covariances $\Sigma_j^{(k)}$, while preserving the effect of spatial homogenization of class memberships.

A reasonable neighbourhood-supported classification using expectation maximization and probabilistic label relaxation (EM-PLR) starts with an initial estimation of mean covariance matrices according to equation (10). Subsequently, for each pixel and for each class, the *a posteriori* probabilities of class membership are estimated as in equation (11). After this step, a probabilistic label relaxation process is used to iteratively update the conditional probabilities p_q for each pixel and class using equations (17) and (18), and a fixed number of K iterations (e.g. five). Equation (13) can then be applied to obtain an updated set of class centres from the values of $p_q^{(k)}$

$$\Sigma_i^{(k+1)} = \frac{\sum_{n=1}^{N} p_q^{(k)}(n, y_i)\mathbf{C}_n}{\sum_{n=1}^{N} p_q^{(k)}(n, y_i)} \tag{20}$$

after which the expectation maximization iteration resumes. This procedure is continued, as in normal expectation maximization, until a certain termination criterion is met. For example, after each set of EM and PLR iterations, a decision about the most likely class can be made:

$$\mathbf{C}_n \in \omega_i \quad \text{if} \quad p_q(n, y_i) > p_q(n, y_j) \quad \forall j \neq i. \tag{21}$$

If only a small percentage of pixels (e.g. less than 1%) changes its class membership in consecutive iterations, the process is stopped and the current classification is considered final. An alternative termination criterion is the convergence of the conditional probabilities themselves (see equation 14), i.e. if $||p_q^{(k)} - p_q^{(k-1)}||$ falls below a certain threshold.

The entire proposed iteration scheme is outlined in figure 2 as a block diagram. In practice, for the case of random initialization, it was also experimentally established to be advantageous to perform some pure expectation maximization iterations in the beginning, i.e. without taking into account the neighbourhood relations. This encourages a certain level of stability in Σ and leads to a faster convergence of the entire process, since initial highly uncertain assignments of p and q are avoided.

3. Experimental results

In order to evaluate the characteristics and potential of the proposed approach, different classifications of an agricultural scene featuring relatively large homogenous fields have been produced. The data were acquired over the test site of Alling,

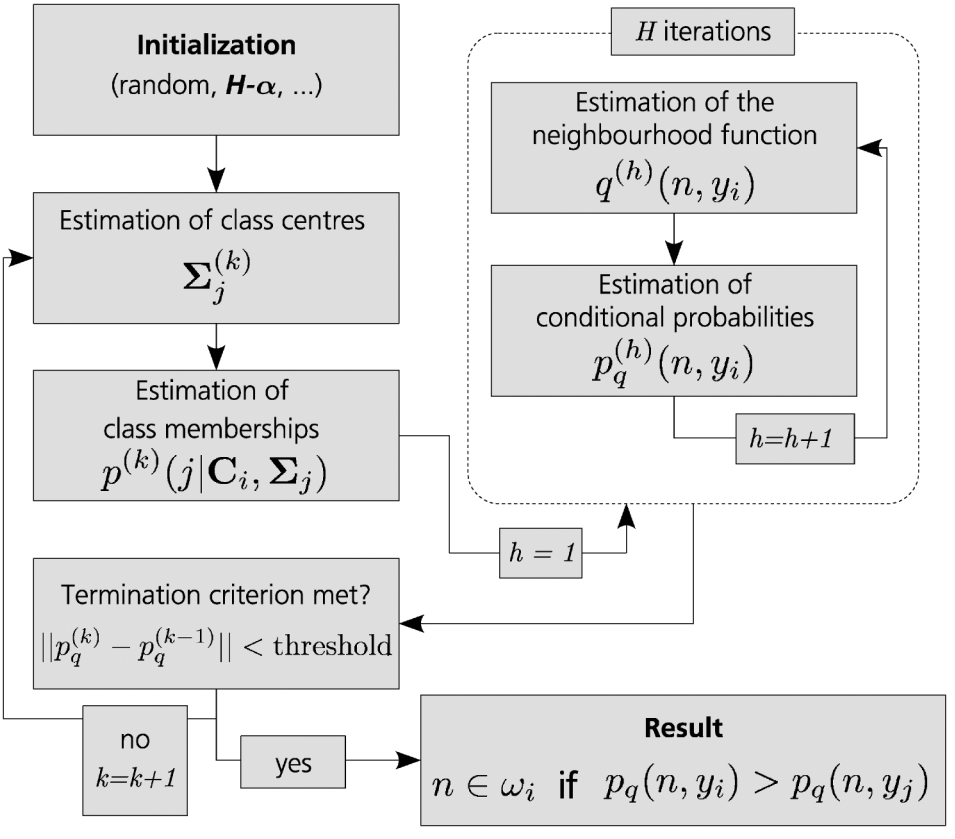

Figure 2. Block diagram of the proposed unsupervised polarimetric classification using expectation maximization and probabilistic label relaxation.

Germany, by DLR's experimental SAR sensor E-SAR at L-band. The image amplitude in the Pauli representation is shown in figure 3 (*a*). A 3×3 boxcar filter was used to average the covariance matrices, resulting in input data with only nine looks. Such slight averaging allows for high spatial resolution but leads to badly estimated covariance matrices due to the strong influence of speckle. This is desirable in this case, since it should be demonstrated how noise degradation can be countered by introducing spatial homogenization through probabilistic label relaxation.

To initialize the expectation maximization process, each observation is assigned randomly to one of eight classes. This type of initialization initially results in very badly defined and fragmented classes, which would be erroneously homogenized by incorporating spatial context. Therefore, as mentioned above, five pure expectation maximization iterations (equations 11 and 13) are performed before starting iterations on the neighbourhood function. Such a procedure is not necessary with other initialization strategies and corresponds, in principle, to an initialization with the classification result obtained after five expectation maximization iterations.

The neighbourhood function according to equation (16) is computed using a 5×5 neighbourhood with Gaussian weighting in all shown examples. This means that a total of 24 adjacent pixels are considered in the summation, where farther pixels are weighted less. Very similar results are obtained when using smaller

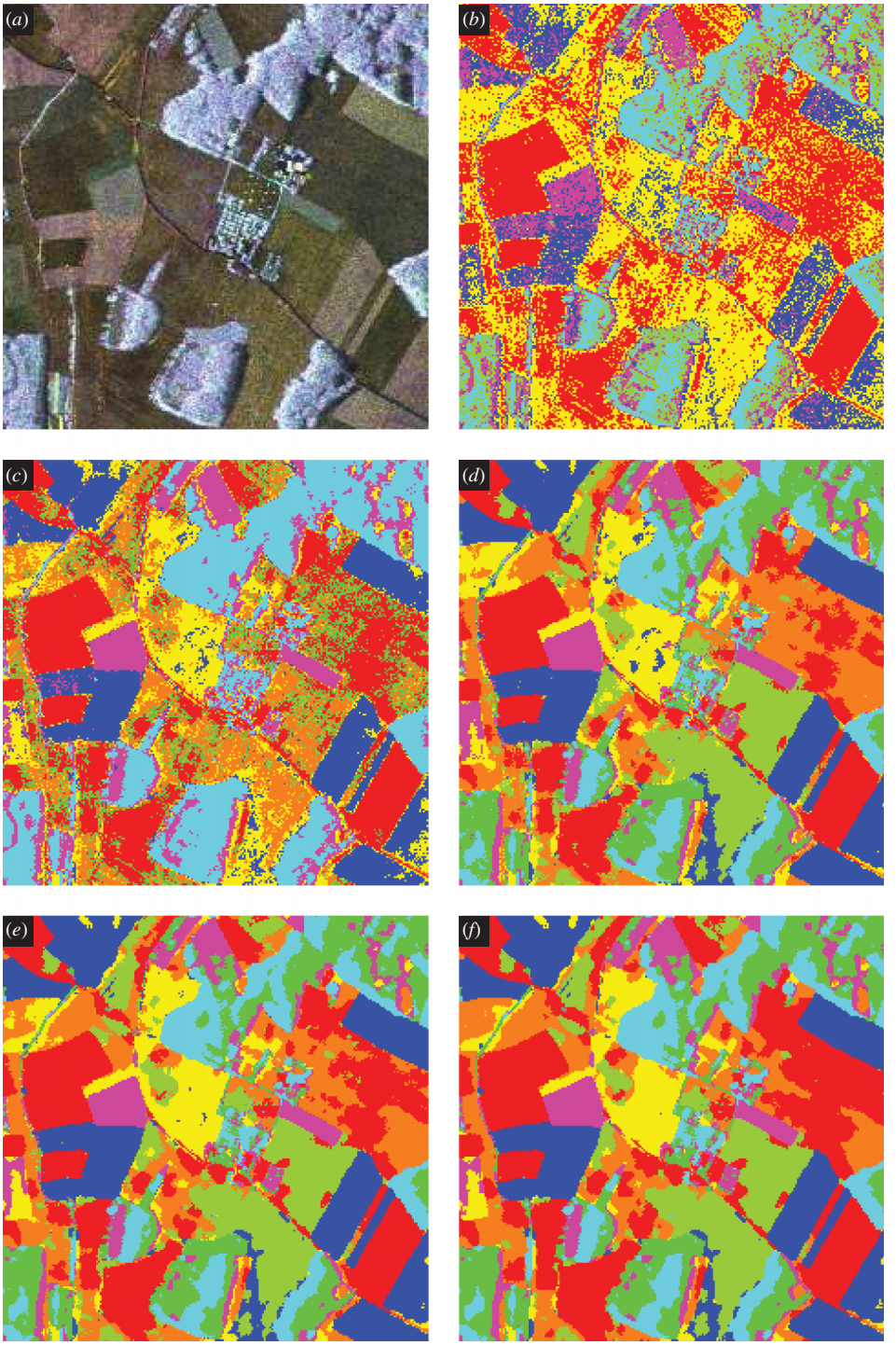

Figure 3. Classification results of neighbourhood supported expectation maximization with different ratios of compatibility coefficients: (*a*) image amplitude, (*b*) identical compatibility coefficients, (*c*) equal class two times more probable than unequal class, (*d*) five times more probable, (*e*) 10 times more probable, (*f*) 100 times more probable.

neighbourhoods, like a 3×3 Gaussian or a simply 4-connectedness: since $p_q(n, y_i)$ is estimated iteratively, pixels beyond the immediate neighbourhood are, eventually, taken into account in any case. The convergence of the conditional probability $||p_q^{(k)} - p_q^{(k-1)}||$ is used as termination criterion. If its value falls below 1%, the iteration process is stopped and each pixel is assigned to the most likely class based on the current p_q.

Figure 3 shows classification results using different values for the compatibility coefficients. In all cases, the estimation of the neighbourhood function is iterated five times. In figure 3(*b*), the probabilities of a transition from one class to the same class and a transition from one class to another class are equal. In this case, the neighbourhood function loses its significance and the result corresponds to that of expectation maximization alone. As expected, the result appears relatively noisy due to the low number of looks of the input data. In figure 3(*c*), a compatibility coefficient is set which is two times higher for equal classes than for different classes. The classification results are noticeably more homogenous. As the ratio between the compatibility coefficients is increased further (figure 3(*d*)–(*f*)), more and more homogenization is achieved. Despite the low number of looks of the data, qualitatively appealing and high-resolution classification results can be achieved. For example, several small point targets persist as small points in the classification result, even when using very strong homogenization (figure 3(*f*), middle).

Figure 5(*a*) illustrates the convergence of the conditional probability. In all cases, the proposed iteration scheme converges relatively quickly after 10–15 iterations. Certain differences can be observed between the individual curves, which result from the fact that a convergence over two competing terms, the covariances and the neighbourhood, is being sought. This can cause discrepancies, when the two terms contradict each other to a certain extent in early iterations, a situation that is resolved at different rates. In general, the overall rate of convergence is seen to be largely independent of the choice of compatibility coefficients.

Figure 4 shows classification results from the same test-site, now varying the number of iterations on the neighbourhood function. This time, the compatibility coefficients are kept constant at a value of 10 times higher probability for a transition between like classes than between different classes. When using a single iteration on the neighbourhood function (shown in figure 4(*a*)), a relatively noisy classification result is obtained, even after more than 15 iterations. This is due to the fact that only the direct neighbourhood is considered and that the neighbourhood function has no time to converge before a new set of class centres Σ is derived. As the number of iterations on the neighbourhood function increases (figure 4(*b*)–(*f*)), a more and more homogenized classification result is achieved, since the size of the neighbourhood considered increases and is no longer confined to the size of the 5×5 Gaussian neighbourhood. Depending on the test site, it is up to the user to choose a reasonable number of iterations; in the presented dataset sufficiently homogenized classification results are achieved after 5–10 iterations.

Figure 5(*b*) shows the convergence of the conditional probabilities $p_q(n, y_i)$ as the number of iterations on the neighbourhood function varies. As before, the method converges relatively quickly after 10–15 iterations in all cases. Again, the rate of convergence seems to depend only marginally on the number of internal iterations on the neighbourhood function.

In general, it appears that a too large number of iterations on the neighbourhood function is not desirable: the result is an overemphasis of the spatial context, in which

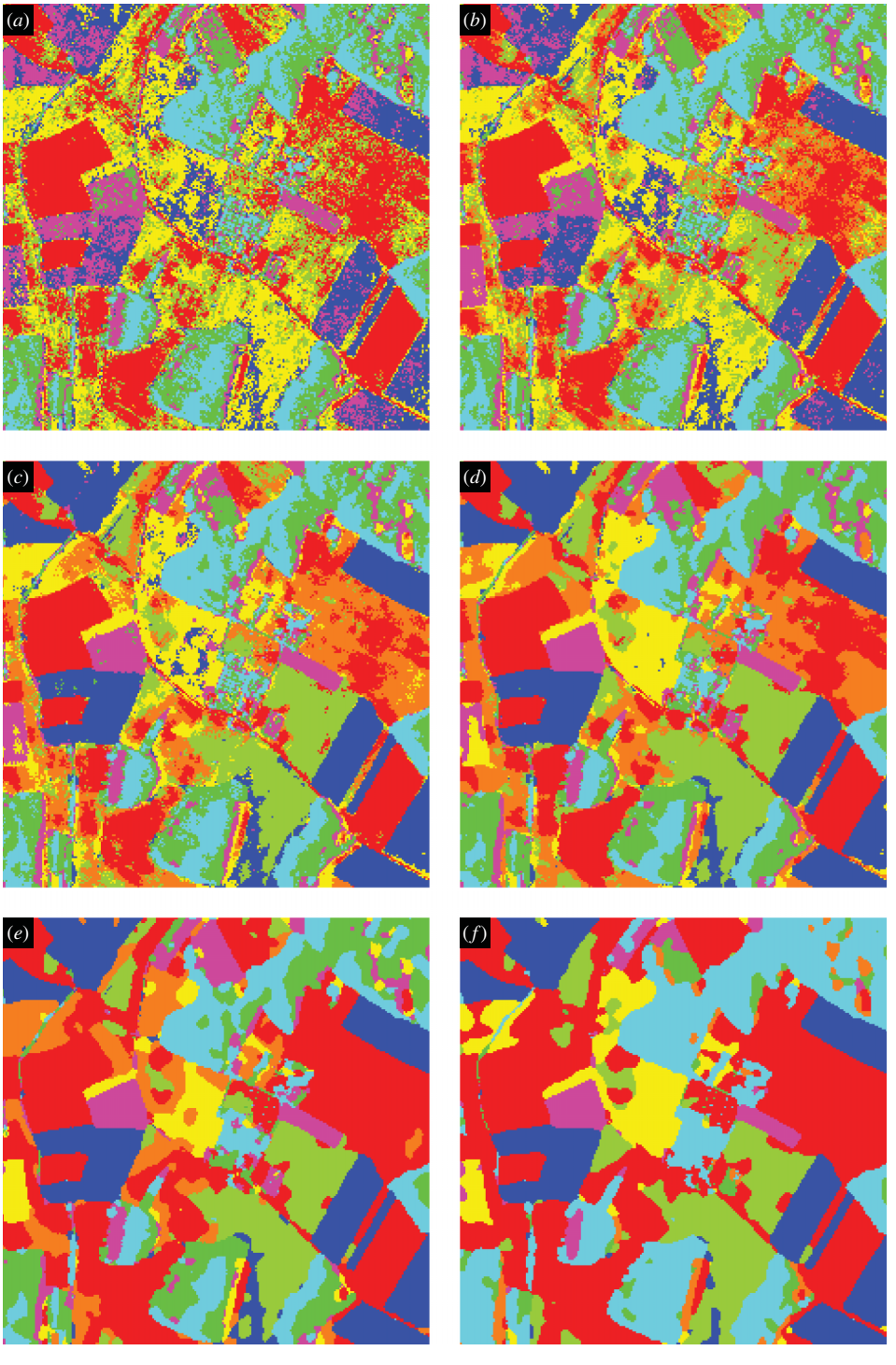

Figure 4. Classification results of neighbourhood supported expectation maximization with different number of iterations on the neighbourhood function: (*a*) one iteration, (*b*) two iterations, (*c*) three iterations, (*d*) five iterations, (*e*) 10 iterations, (*f*) 20 iterations.

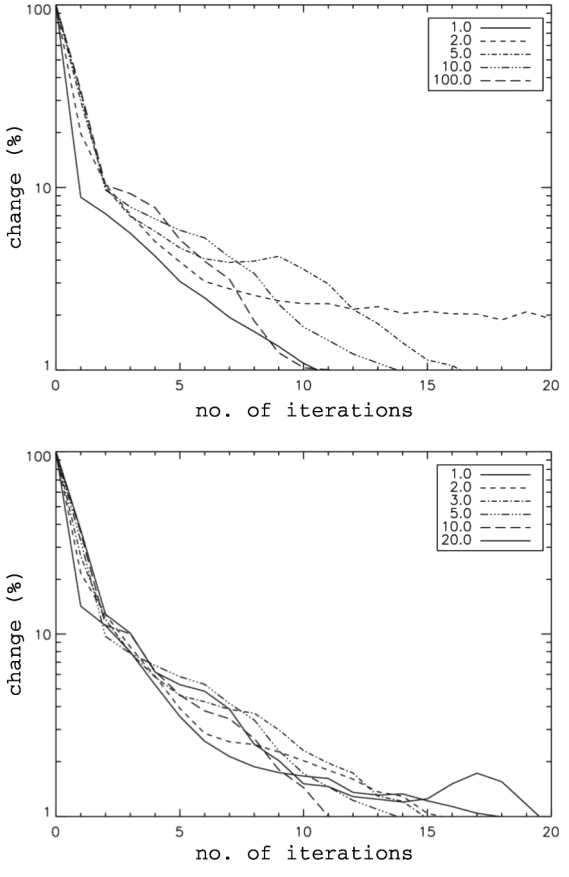

Figure 5. Top: convergence of the conditional probabilities $||p_q^{(k)} - p_q^{(k-1)}||$ using different ratios of compatibility coefficients (compare figure 3). Bottom: convergence of the conditional probabilities $||p_q^{(k)} - p_q^{(k-1)}||$ using different number of iterations on the neighbourhood function (compare figure 4).

class assignments essentially neglect the Wishart-distribution over observed covariance matrices. As a result, it is likely that areas, almost identical in terms of polarimetric information, become separated in early iterations and are subsequently homogenized by the neighbourhood function. Later iterations are not able to correct this error, since the costs of spatially inhomogeneous class labels exceed the benefit of a better polarimetric assignment. Too high compatibility coefficients are also disadvantageous, since these unnecessarily hamper a change into another class. Particularly with regards to relatively fine structures in the analysed scene, the compatibility coefficient should be kept rather small in order to avoid excessive homogenization. 5–10 iterations on the neighbourhood function and a ratio between compatibility coefficients between 10 and 100 have proved to be reasonable. However, these values have to be considered to be strongly scene dependent.

A remaining question is the overall classification accuracy of the proposed algorithm, compared to other known approaches for unsupervised classification of polarimetric data. For this purpose, a manual classification into eight classes, based on the

image amplitude shown in figure 3(*a*), was prepared and serves as a reference. Figure 6 shows this manual classification, as well as results obtained with different algorithms: figures 6(*b*) and (*c*) depict the result of Wishart *k*-means and expectation maximization, respectively, both derived from data filtered by a 7 × 7 refined-Lee speckle filter. Figure 6(*d*) depicts the result of the proposed approach, derived from data filtered with a 3 × 3 boxcar filter with a compatibility ratio of 10 and 10 iterations on the neighbourhood function. As unsupervised algorithms do not necessarily derive an identical class numbering, the colouring was adjusted manually.

Even on the basis of data with a much lower number of looks, the proposed approach is able to generate more homogenous classification results. The adaptive speckle filtering was clearly not sufficient to ensure a classification homogeneity over the distributed targets that is comparable to the proposed algorithm. Table 1 contains a quantitative evaluation of the classification performance. For each of the classes in

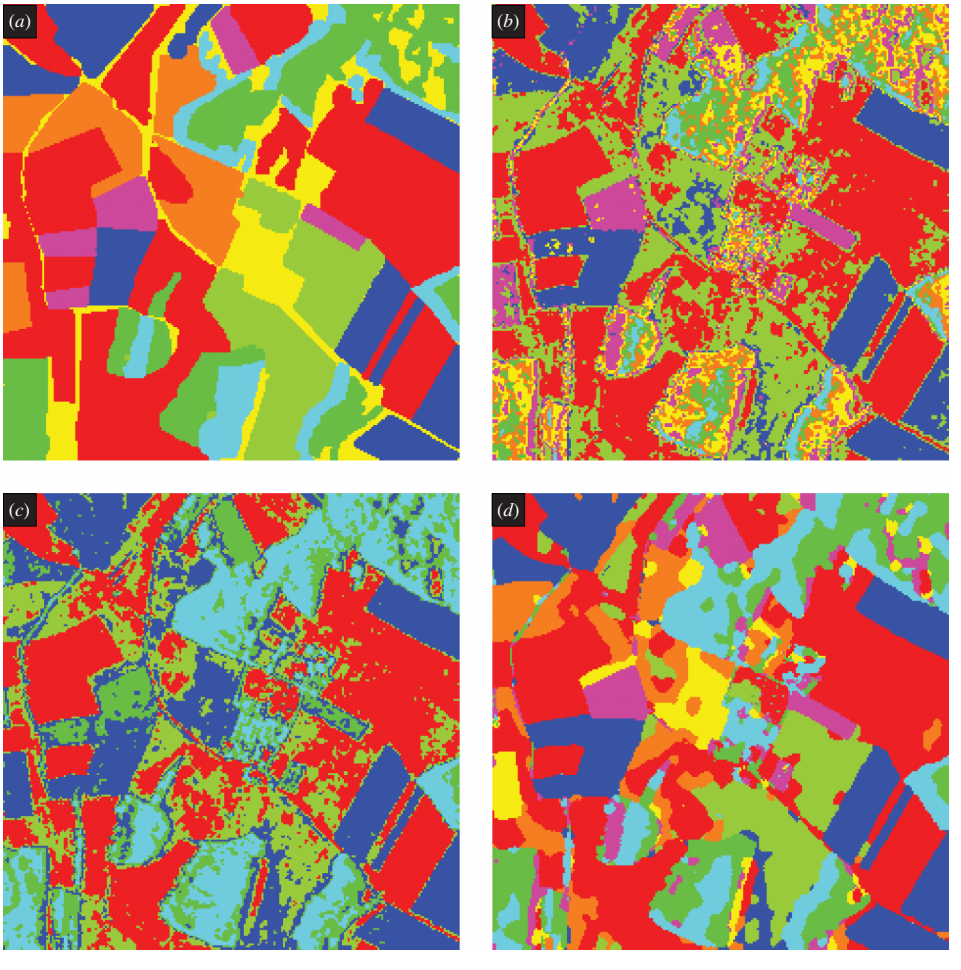

Figure 6. Comparison of classification results of different algorithms with ground truth generated by manual segmentation into eight classes: (*a*) visual classification. (*b*) Wishart *k*-means after 7 × 7 refined-Lee filtering. (*c*) EM after 7 × 7 refined-Lee filtering. (*d*) EM-PLR after 3 × 3 boxcar filtering.

Table 1. Classification accuracy (in %) of different classifiers.

	1	2	3	4	5	6	7	8	Ø
H-α thresholds (nine looks)	30	78	78	42	30	42	90	59	56
H-α thresholds (refined-Lee filtered)	42	78	79	43	46	48	94	41	59
Wishart *k*-means (nine looks)	24	73	76	32	48	25	51	60	49
Wishart *k*-means (refined-Lee filtered)	24	85	75	32	61	31	53	63	53
EM (nine looks)	18	53	61	52	65	63	37	43	49
EM (refined-Lee filtered)	21	53	67	54	73	68	37	46	53
EM-PLR (nine looks)	30	92	38	86	71	92	67	63	67
EM-PLR (refined-Lee filtered)	31	88	37	86	71	92	69	55	66

the manual classification, the percentage of pixels recognized by the various algorithms was calculated. It has to be noted that class numbering is variable in unsupervised algorithms. Therefore, for each class and method, a unique correspondence was determined to ensure the largest possible resemblance. Even so, some unresolvable ambiguities in class correspondence might remain, particularly in case of severely misestimated classes. All in all, the values of table 1 clearly suggest that the proposed algorithm performs significantly better than the other algorithms tested. Its mean recognition rate of 67% is still not ideal, which might be an indicator of the fact that the chosen class number of eight does not adequately describe this test-site.

4. Discussion and conclusions

This paper proposes an efficient unsupervised approach for the classification of polarimetric SAR data, which is designed to yield homogenous classification results. In the proposed approach, the spatial context is explicitly taken into account during the statistical relaxation process of the likelihood function. This is achieved by combining an iterative expectation maximization approach with the principle of probabilistic label relaxation. Both techniques together provide spatially homogenous and polarimetrically motivated classification results even in the presence of strong speckle noise.

In current standard approaches homogenization is usually ensured by smart speckle filtering of the covariance matrices; the actual classification process, however, consists of standard clustering techniques on a pixel-by-pixel basis using the statistical characteristics of SAR data. Advanced techniques which take into account spatial homogeneity constraints, like, e.g. Markov random fields, are effective, but computationally very expensive to optimize due to their highly dimensional configuration space. The proposed technique provides a simple alternative, which is easy to implement and computationally very efficient.

Probabilistic label relaxation results in a strong smoothing of homogenous areas, while class boundaries and point targets are preserved at full resolution. In other words, strong adaptive speckle filtering prior to classification is rendered unnecessary by a spatially adaptive classification approach. This is appealing, since only marginal averaging of the covariance matrices causes pronounced speckle noise but preserves high spatial resolution. As has been shown, the proposed neighbourhood-supported classification is superior to known standard techniques for unsupervised polarimetric classification, particularly in case of a low number of looks.

The disadvantages of the proposed approach are typical for all unsupervised classification methods: the chosen number of classes can have a strong influence on the classification result, and has to be determined carefully. The same holds for the chosen iteration strategy. In general, the proposed technique is potentially of interest when homogenous classification results, rather than pointwise assignments, are desired, as is usually the case in agricultural or forested areas. In these contexts, the proposed approach yields results of higher quality than conventional approaches, based solely on the analysis of covariance matrices.

Acknowledgements

The authors wish to thank QinetiQ Corporation and the German Aerospace Center (DLR) for providing the SAR data in the framework of the SPARC (Surface Parameter Retrieval Collaboration) project. This work was partially supported by the German Research Foundation (DFG) under project number RE 1698/2.

References

Ball, G.H. and David, J., 1965, *ISODATA, A Novel Method of Data Analysis and Pattern Classification* (Menlo Park, CA: Stanford Research Institute).

Bezdek, J.C., 1981, *Pattern Recognition with Fuzzy Objective Function Algorithms* (Dordrecht: Kluwer Academic Publishers).

Chellappa, R. and Jain, A., 1993, *Markov Random Fields: Theory and Application* (New York: Academic Press).

Chen, C.T., Chen, K.S. and Lee, J.S., 2003, The use of fully polarimetric information for the fuzzy neural classification of SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, pp. 2089–2100.

Cloude, S.R. and Pottier, E., 1997, An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing*, **35**, pp. 68–78.

Davidson, G., Ouchi, K., Saito, G., Ishitsuka, N., Mohri, N. and Uratsuka, S., 2002, Polarimetric classication using expectation methods. In *Polarimetric and Interferometric SAR Workshop,* Communications Research Laboratory, Tokyo.

Dempster, A.P., Laird, N.M. and Rubin, D.B., 1977, Maximum-Likehood from incomplete data via the expectation maximisation algorithm. *Journal of Royal Statistical Society B*, **39**, pp. 1–38.

Du, L.J. and Lee, J.S., 1996, Fuzzy classification of earth terrain covers using multi-look polarimetric SAR image data. *International Journal of Remote Sensing*, **17**, pp. 809–826.

Ferro-Famil, L., Pottier, E. and Lee, J.S., 2001, Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha-Wishart classifier. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, pp. 2332–2342.

Fjortoft, R., Delignon, Y., Pieczynski, W., Sigelle, M. and Tupin, F., 2003, Unsupervised classification of radar images using hidden Markov chains and hidden random fields. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, pp. 675–686.

Freund, J.E. and Walpole, R.E., 1987, *Mathematical Statistics*, 4th ed. (Englewood Cliffs, NJ: Prentice Hall).

Goodman, N.R., 1963, Statistical analysis based on a certain multivariant complex Gaussian distribution. *The Annals of Mathematical Statistics*, **34**, pp. 152–177.

Harris, R., 1985, Contextual classification post-processing of Landsat data using a probabilistic relaxation model. *International Journal of Remote Sensing*, **6**, pp. 847–866.

Hegarat-Mascle, S.L., Madjar, D.V. and Olivier, P., 1996, Applications of simulated annealing to SAR image clustering and classification problems. *International Journal of Remote Sensing*, **17**, pp. 1761–1776.

LEE, J.S., GRUNES, M.R., AINSWORTH, T.L., DU, L.J., SCHULER, D.L. and CLOUDE, S.R., 1999, Unsupervised classification using polarimetric decomposition and the complex Wishart classifier. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, pp. 2249–2258.

LEE, J.S., GRUNES, M.R. and KWOK, R., 1994a, Classification of multi-look polarimetric SAR imagery based on the complex Wishart distribution. *International Journal of Remote Sensing*, **15**, pp. 2299–2311.

LEE, J.S., GRUNES, M.R., POTTIER, E. and FERRO-FAMIL, L., 2004a, Unsupervised terrain classification preserving polarimetric scattering characteristics. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, pp. 722–731.

LEE, J.S., HOPPEL, K.W., MANGO, S.A. and MILLER, A.R., 1994b, Intensity and phase statistics of multilook polarimetric and interferometric imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **32**, pp. 1017–1028.

LEE, J.S., SCHULER, D.L. and AINSWORTH, T.L., 2004b, Scattering model based speckle filtering of polarimetric SAR data. In *Proceedings of EUSAR'04,* Ulm, Germany, pp. 203–207.

TISON, C., NICOLAS, J., TUPIN, F. and MAÎRE, H., 2004, A new statistical model of urban areas in high resolution SAR images for Markovian segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, pp. 2046–2057.

TOWNSEND, F.E., 1986, The enhancement of computer classifications by logical smoothing. *Photogrammetric Engineering and Remote Sensing*, **52**, pp. 213–221.

VAN ZYL, J.J., 1989, Unsupervised classification of scattering mechanisms using radar polarimetry data. *IEEE Transactions on Geoscience and Remote Sensing*, **27**, pp. 36–45.

VASILE, G., TROUVÉ, E., LEE, J.S. and BUZULOIU, V., 2006, Intensity-driven adaptive-neighborhood technique for polarimetric and interferometric SAR parameters estimation. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, pp. 1609–1621.

ZADEH, L.A., 1975, *Fuzzy Sets and Their Applications to Cognitive and Decision Processes* (New York: Academic Press).