

5. POLARIMETRIC SAR DATA CLASSIFICATION

5.1 Classification of polarimetric scattering mechanisms

- Direct interpretation of decomposition results
- Cameron classification
- Lee classification based on Freeman decomposition

5.1.1 H-A- α classification of scattering mechanisms

Cloude and Pottier proposed an algorithm to identify in an unsupervised way polarimetric scattering mechanisms in the H- α plane. The key idea is that entropy arises as a natural measure of the inherent reversibility of the scattering data and that α can be used to identify the underlying average scattering mechanism.

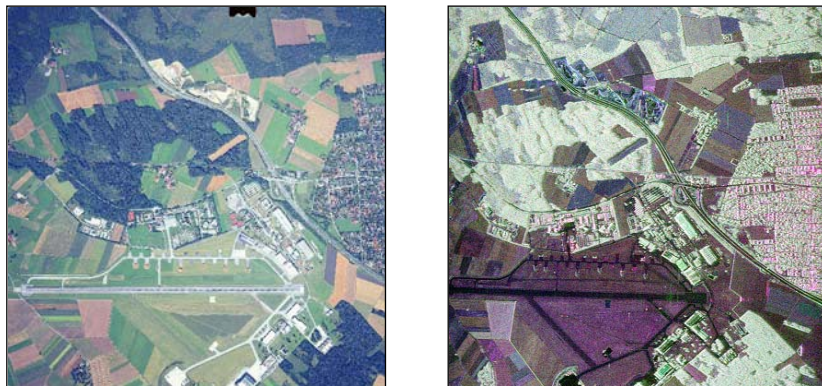


Figure 1 Optical image (left), polarimetric color coded image (right) of the Oberpfaffenhofen test site.

The H- α classification plane is sub-divided into 8 basic zones characteristic of different scattering behaviors. The basic scattering mechanism of each pixel of a polarimetric SAR image can then be identified by comparing its entropy and α parameters to fixed thresholds. The different class boundaries, in the H- α plane, have been determined so as to discriminate surface reflection (SR), volume diffusion (VD) and double bounce reflection (DB) along the α axis and low, medium and high degree of randomness along the entropy axis. Detailed explanations, examples and comments concerning the different classes can be found in the publication from Cloude and Pottier.

Figure 2 shows the H- α plane and the occurrence of the studied polarimetric data into this plane.

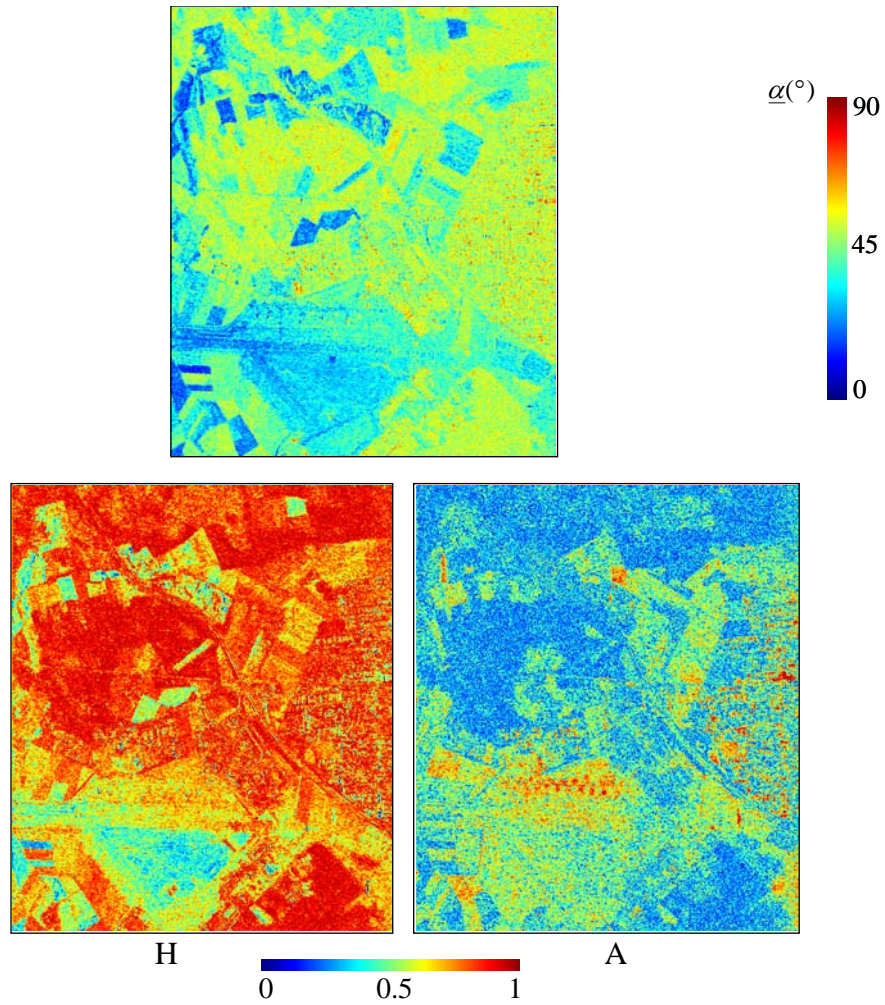


Figure 2 Polarimetric decomposition main parameters: H, A and $\underline{\alpha}$.

Grey zones in the H- $\underline{\alpha}$ plane correspond to unfeasible areas.

It can be seen, in **Figure 3**, that the largest densities in the occurrence plane correspond to volume diffusion and double bounce scattering with moderate and high randomness. Medium and low entropy surface scattering mechanisms are also frequently encountered in the scene under examination.

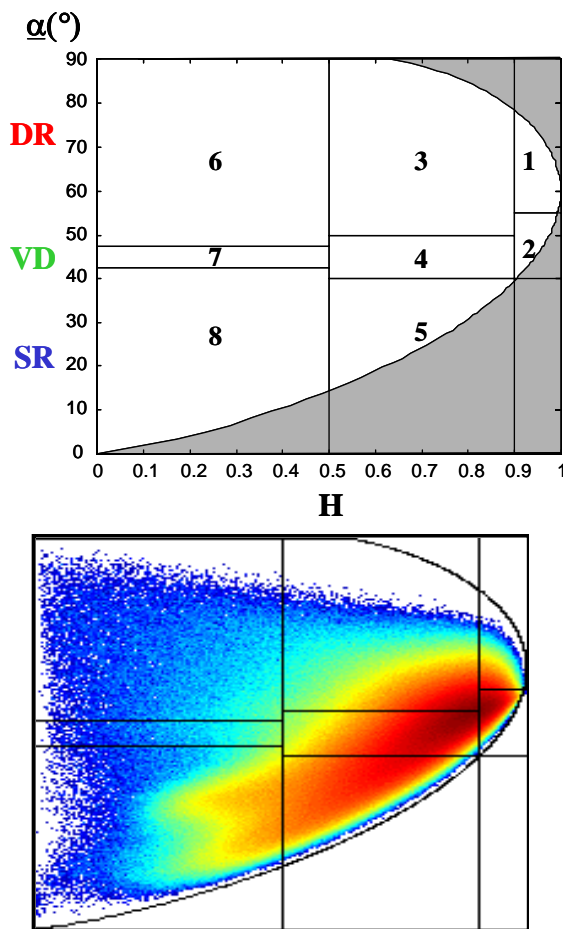


Figure 3 H- α scattering mechanism identification plane (top). Polarimetric data occurrence in the H- α plane (bottom).

Data distribution in the H- α plane does not show, for the considered scene, distinct natural clusters belonging to a single scattering mechanism class. Therefore, identification results may highly depend on segmentation thresholds.

Results of the unsupervised identification procedure are presented in **Figure 4**.

It can be observed in **Figure 4** that the proposed segmentation in the H- α plane permits to identify in a macroscopic way the type of scattering mechanism. Agricultural fields and bare soils are characterized by surface scattering. Scattering over forested areas is dominated by volume diffusion while urban areas are mainly characterized by double bounce scattering. It may be noted that the identification process slightly overestimates volume diffusion and double bounce scattering over surfaces.

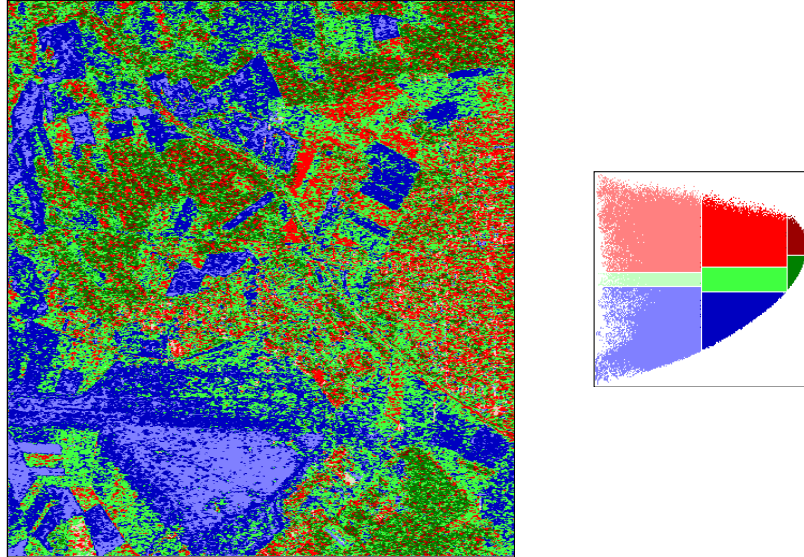


Figure 4 Unsupervised scattering mechanism identification in the H- α plane.

5.2 Polarimetric data statistical segmentation

5.2.1 Maximum Likelihood Supervised segmentation

5.2.1.1 Bayes optimal decision rule

The problem of optimal segmentation over a SAR image over a fixed number of clusters may be formulated as follows.

How to assign, in an optimal way, a SAR image pixel, p , to one of the M possible clusters $\{\Theta_1, \dots, \Theta_M\}$, according to a SAR observable, x ?

The answer is given by Bayes optimal decision rule using the conditional probability of the different clusters

$$\text{Decide } p \in \Theta_i \text{ if } P(\Theta_i|x) > P(\Theta_j|x) \quad \forall j \neq i \quad (1)$$

A pixel is assigned to the most probable cluster conditionally to the observation of x over the pixel under consideration.

The probability of error related to this decision is obtain from the a posteriori probabilities of the unselected clusters by

$$P(\text{error}|x) = \sum_{j \neq i} P(\Theta_j|x)P(x) \quad (2)$$

It is obvious, that assigning a pixel to the cluster with the highest a posteriori probability minimizes the conditional error probability.

The total probability is then computed using

$$P(error|x) = \int P(error|x)P(x)dx \quad (3)$$

The set of *a posteriori* probabilities $P(\Theta_i|x)$ is generally uneasy to derive. One prefers then to express such quantities using Bayes rule

$$P(\Theta_i|x)P(x) = P(x|\Theta_i)P(\Theta_i) \quad (4)$$

Inserting this expression in (1), the expression of the optimal decision rule becomes

$$\text{Decide } p \in \Theta_i \text{ if } P(x|\Theta_i)P(\Theta_i) > P(x|\Theta_j)P(\Theta_j) \quad \forall j \neq i \quad (5)$$

Where $P(x|\Theta_i)$ is the likelihood of the observation x , given Θ_i and $P(\Theta_i)$ is the prior probability of Θ_i .

Additionally, if the prior probabilities are supposed to be equal, the optimal decision rule may written using only cluster likelihood as

$$\text{Decide } p \in \Theta_i \text{ if } \Theta_i = \text{Arg max } P(x|\Theta_i) \quad (6)$$

5.2.1.2 Supervised segmentation scheme

A supervised segmentation scheme may be represented as a three-step process described in the following.

First training data, separated into M distinct groups or classes, are supplied to the segmentation algorithm. Training inputs may be :

- user supplied under the correct format
- sampled from the image under study
- sampled from SAR images acquired over other scenes

During the second phase, the segmentation algorithm learns from the training data sets how to discriminate the M classes and establishes decision rules. In the case of supervised segmentation based on Bayes optimal decision rule, the segmentation algorithm learns the different statistical quantities useful to compute the expression mentioned in (6).

Once the learning phase has converged, the algorithm assigns each pixel of the SAR image under study to one of the M clusters provided in the first phase, according to decision procedure derived during the second step.

5.2.1.3 Supervised ML segmentation of [S] matrix images

In the case of [S] matrix images, the SAR observation x is replaced by a 3 element complex target vector $\underline{k} = [S_{11}, \sqrt{2}S_{12}, S_{22}]^T$.

It has been verified that when the radar illuminates an area of random surface of many elementary scatterers, \underline{k} can be modeled as having a multivariate complex circular gaussian probability density function $N_c(0, [\Sigma])$ of the form :

$$P(\underline{k}) = \frac{\exp(-\underline{k}^\dagger [\Sigma]^{-1} \underline{k})}{\pi^q |\Sigma|} \quad (7)$$

where q stands for the number of elements of \underline{k} , equal to three in the monostatic case, $|\Sigma|$ represents the determinant, and $[\Sigma] = E(\underline{k}\underline{k}^\dagger)$ is the global (3×3) covariance matrix of \underline{k} .

The likelihood of a target vector \underline{k} given a cluster Θ_i is then given by

$$P(\underline{k}|\Theta_i) = \frac{\exp(-\underline{k}^\dagger [\Sigma_i]^{-1} \underline{k})}{\pi^q |\Sigma_i|} \quad (8)$$

With $[\Sigma_i]$ the covariance matrix of cluster Θ_i computed during the learning phase. In practice the actual value of $[\Sigma_i]$ remains unknown and the covariance matrix is replaced by its maximum likelihood estimate $[\hat{\Sigma}_i]$ defined as

$$[\hat{\Sigma}_i] = \frac{1}{n_i} \sum_{p \in \Theta_i} \underline{k}\underline{k}^\dagger \quad (9)$$

With n_i the number of pixels belonging to the training cluster Θ_i .

From a computational point of view, it is generally preferable to deal with log-likelihood function, $L(\underline{k}|\Theta_i)$ instead of the expression given in (8)

$$L(\underline{k}|\Theta_i) = -\ln|\hat{\Sigma}_i| - \text{tr}([\hat{\Sigma}_i]^{-1} \underline{k}\underline{k}^\dagger) - q \ln \pi \quad (10)$$

Where $\text{tr}([\hat{\Sigma}_i]^{-1} \underline{k}\underline{k}^\dagger)$ stands for the trace of $[\hat{\Sigma}_i]^{-1} \underline{k}\underline{k}^\dagger$ and equals $\underline{k}^\dagger [\hat{\Sigma}_i]^{-1} \underline{k}$.

The logarithm function being strictly increasing with its argument, the optimal decision rule is transformed, in the case of [S] matrix segmentation, to

$$\begin{aligned} \text{Decide } p \in \Theta_i \text{ if } \Theta_i = \text{Arg min } d(\underline{k}|\hat{\Sigma}_i) \text{ with} \\ d(\underline{k}|\hat{\Sigma}_i) = +\ln|\hat{\Sigma}_i| + \text{tr}([\hat{\Sigma}_i]^{-1} \underline{k}\underline{k}^\dagger) \end{aligned} \quad (11)$$

Where the variable $d(\underline{k}|\hat{\Sigma}_i)$ may be assimilated to a statistical distance, derived from the opposite of log-likelihood function without constant terms.

During the segmentation process a pixel is then assigned to the minimum distance cluster, having the closest statistics.

The supervised segmentation algorithm may be summarized as follows

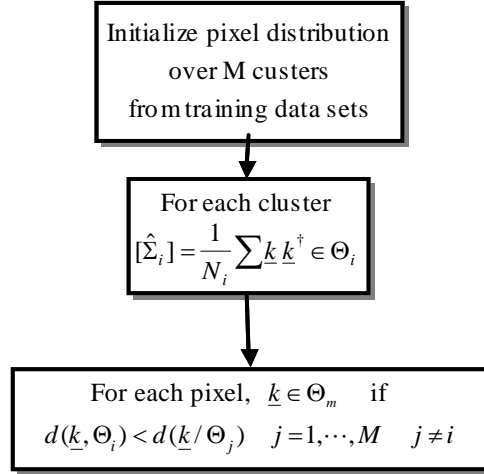


Figure 5 Supervised ML segmentation scheme.

5.2.1.4 Supervised ML segmentation of [T] or [C] matrix images

It has been shown that assuming that target vectors have a $N_c(0, [\Sigma])$ distribution, a sample n -look covariance matrix $[C] = 1/n \sum_n k k^\dagger$, follows a complex Wishart distribution with n degrees of freedom, $W_c(n, [\Sigma])$, given by

$$p([T]) = \frac{n^{qn} |[T]|^{n-q} \exp(-tr(n[\Sigma]^{-1}[T]))}{K(n, q) |[\Sigma]|^n}$$

$$\text{with } K(n, q) = \pi^{q(q-1)/2} \prod_{i=1}^q \Gamma(n-i+1)$$

Where $\Gamma(\cdot)$ represents the Gamma function.

A development similar to the one presented in the case of scattering matrices leads to the following decision rule

$$\text{Decide } p \in \Theta_i \text{ if } \Theta_i = \text{Arg max } L([T] | [\hat{\Sigma}_i]) \text{ with}$$

$$L([T] | [\hat{\Sigma}_i]) = -n \ln |[\hat{\Sigma}_i]| - n \text{tr}([\hat{\Sigma}_i]^{-1} [T]) + qn \ln n + (n-q) \ln |[T]| - \ln K(n, q) \quad (13)$$

With $[\hat{\Sigma}_i]$, the maximum likelihood estimate of the coherency matrix defined as

$$[\hat{\Sigma}_i] = \frac{1}{n_i} \sum_{p \in \Theta_i} [T] \quad (14)$$

Taking the opposite of the lower expression of relation (13) and removing terms that do not depend on the cluster under test, the optimal decision rule becomes

$$\text{Decide } p \in \Theta_i \text{ if } \Theta_i = \text{Arg min } d([T] | [\hat{\Sigma}_i]) \text{ with}$$

$$d([T] | [\hat{\Sigma}_i]) = + \ln |[\hat{\Sigma}_i]| + \text{tr}([\hat{\Sigma}_i]^{-1} [T]) \quad (15)$$

It may be noted that, since covariance and coherence matrices are related by a similarity transformation, the form of the statistical distance expressed in (15) is invariant under a change from coherency matrix form to covariance matrix form and (15) is adapted to both $[T]$ and $[C]$ representations.

The supervised segmentation algorithm may be summarized as follows

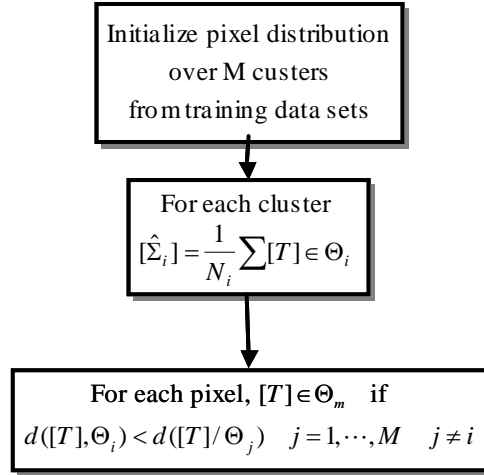


Figure 6 Supervised ML segmentation scheme.

5.2.1.5 Supervised segmentation performance

The performance of a supervised segmentation process is generally estimated via a confusion matrix computed from the training data.

Each element of the training data set is processed through the supervised classifier. The confusion matrix summarizes the percentage of pixels belonging to an original cluster, Θ_i , effectively affected to one of the M possible clusters, Θ_j , with $j = 1 \dots M$. The confusion matrix of an ideal, error-free, segmentation is diagonal.

User-defined training data sets are generally oriented towards the discrimination of different kinds of medium and may not represent all scattering behaviors encountered over a natural scene. Therefore, some parts of a SAR image may have statistics that do not fit with any of the proposed cluster ones. In this case, the assignment of the corresponding pixels may lead to irrelevant interpretation of the segmentation results. One possible solution to this problem consists in creating a reject class. A pixel is assigned to this reject class if it is too much distant, from a statistical point of view, from all the learning clusters.

A possible metric may be defined from the statistical distances defined in (11) and (15). A pixel belongs to the reject class if its distance to the closest cluster exceeds a given factor of the within-class distance standard deviation.

This decision process may be summarized as follows

$$\begin{aligned}
 &\text{Decide } p \in \Theta_{rej} \text{ if } d([T] | [\hat{\Sigma}_i]) > \sigma_{d_i} . r \\
 &\text{with } \Theta_i = \text{Arg min } d([T] | [\hat{\Sigma}_i])
 \end{aligned} \tag{16}$$

$$\text{and } \sigma_{d_i}^2 = \frac{1}{N_i} \sum_{p \in \Theta_i} d([T]|\hat{\Sigma}_i)^2 - \left(\frac{1}{N_i} \sum_{p \in \Theta_i} d([T]|\hat{\Sigma}_i) \right)^2$$

Where the variable r represents a user-defined rejection factor.

5.2.2 Maximum Likelihood unsupervised segmentation

It has been showed, in former paragraphs that a pixel can be assigned to a cluster in an optimal way, according to the following decision rule.

$$\text{Decide } p \in \Theta_i \text{ if } \Theta_i = \text{Arg max } P(x|\Theta_i) \quad (17)$$

In the case of an unsupervised segmentation scheme, the likelihood of a cluster, $P(x|\Theta_i)$, cannot be estimated since its calculation requires to compute the ML estimate of the cluster coherency matrix from the samples contained in the cluster.

An optimal solution to the unsupervised segmentation problem consists in distributing the pixels of an image over the set of M clusters so as to maximize the global likelihood defined as the product of all individual likelihood functions.

A rigorous resolution requires to test all the possible combinations and to select the one corresponding to the maximum joint likelihood value. This optimal solution cannot be applied due to the unrealistic computational load it involves.

Alternative solutions based on sub-optimal iterative optimization procedures are generally preferred.

5.2.3 Principle of statistical K-means clustering

The K-means procedure is an iterative optimization algorithm described by the following synopsis :

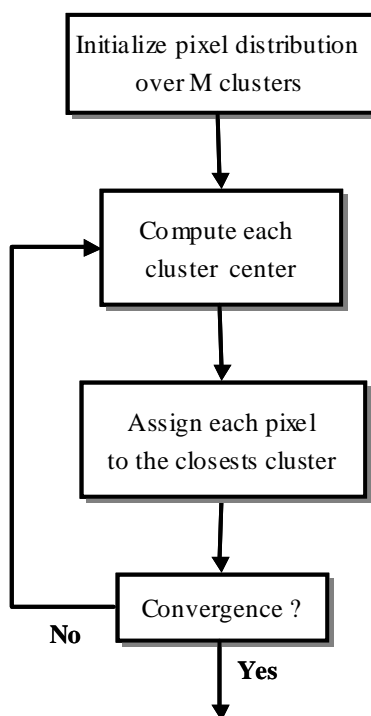


Figure 7 Synopsis of the K-means clustering algorithm.

The algorithm begins with the initialization of the image pixel distribution over the M clusters. This distribution may be done in a random way or according to user specifications. Once all pixels are affected, the different cluster centers are computed according to the processed data type. Each pixel is then affected to the closest cluster according to a distance measure. The convergence of the algorithm is then tested using stability metrics. If a termination criterion is met, the segmentation stops, otherwise a new iteration starts over from the class center computation step.

The K-means algorithm aims to optimize a global function by iteratively optimizing local expressions. It is known this type of techniques may get stuck into locally stable states and fail to determine the optimal pixel distribution. The initialization of the pixel distribution into N classes is then a critical stage of the K-mean clustering algorithm. An adequate initialization permits a fast convergence and provides correctly segmented clusters.

The convergence of the algorithm is evaluated by testing a condition of termination. Such a criterion may be defined from the estimation of the classification quality, or consist in a maximum number of iterations or in a sufficiently low number of pixels that are differently classified from one iteration to the other.

5.2.4 Unsupervised ML segmentation of [T] or [C] matrix images

The K-means clustering algorithm may be adapted to the segmentation of [T] or [C] polarimetric representations as shown on the following synopsis

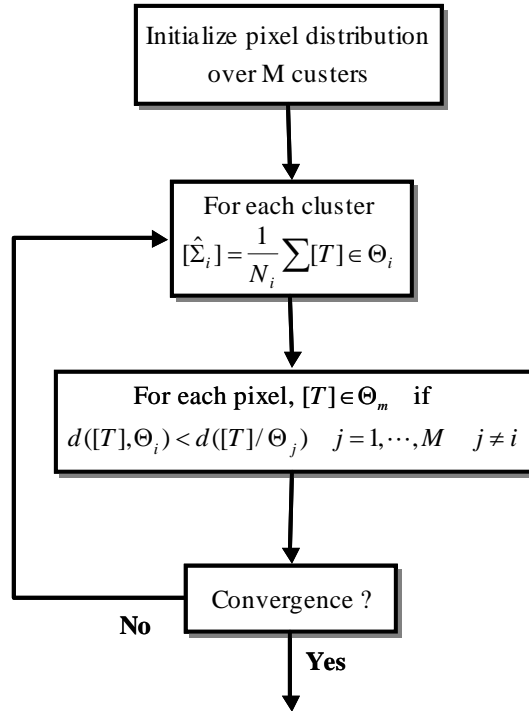


Figure 8 Unsupervised Maximum Likelihood segmentation scheme.

The class center initialization step consists in computing each cluster coherency matrix ML estimate. Pixels are then affected to a cluster according to the ML distance measure derived in (15).

5.2.4.1 Unsupervised Wishart H- α segmentation

The particularity of the identification procedure based on H and $\underline{\alpha}$, introduced in a former paragraph, resides in the estimation of the type of observed media from a physical interpretation of canonical scattering mechanisms using robust indicators. Nevertheless, the analysis of natural scenes using this unsupervised approach may reach some limitations:

- The arbitrarily fixed linear decision boundaries in the H- $\underline{\alpha}$ plane may not fit data distribution. A natural cluster corresponding to similar targets may lie across a frontier in the decision plane. In this case, pixels with very similar characteristics may be assigned, in an almost random way, to different classes due slightly different locations in the H- $\underline{\alpha}$ plane. This effect can be observed in **Figure 4** where the variability in natural media polarimetric features lead to noisy classification results.
- Even if the computation of H and $\underline{\alpha}$ requires fully polarimetric data, these two parameters do not represent the whole polarimetric information. The use of other indicators such as the span or specific correlations coefficients may improve the classification results in a significant way.

Segmentation procedures based on the whole coherency matrix statistics permit to overcome the limitations mentioned above. Nevertheless, it is shown in the following, that the physical interpretation of the scattering phenomenon permits to enhance in a significant way the performance of statistical segmentation schemes.

The unsupervised segmentation scheme mentioned above is initialized in an efficient way with the results of the unsupervised identification of a scattering mechanism, using H , A and $\underline{\alpha}$. This initialization provides 8 stable clusters relating to the underlying physical scattering mechanism.

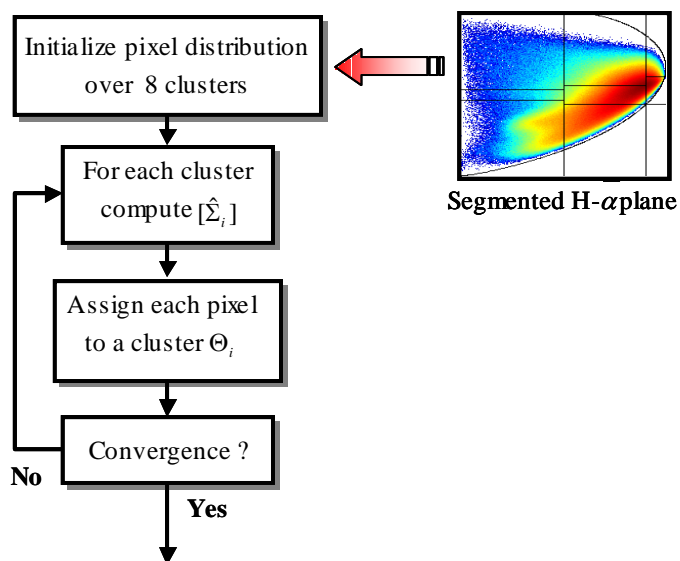


Figure 9 Unsupervised Wishart $H-\underline{\alpha}$ segmentation scheme.



Figure 10 Wishart $H-\underline{\alpha}$ segmentation results.

An important improvement in the segmentation accuracy can be observed in the image presented in **Figure 10**. The main kinds of natural media are clearly discriminated by the Wishart $H-\underline{\alpha}$ segmentation scheme. This unsupervised classification algorithm modifies the

decision boundaries in an adaptive way to better fit the natural distribution of the scattering mechanisms and takes into account information related to the back-scattered power.

5.2.4.2 Unsupervised Wishart H-A- α segmentation

The ML Wishart segmentation may be further improved by explicitly including the anisotropy information during the segmentation procedure. As mentioned previously, the anisotropy indicates the relative importance of secondary mechanisms obtained from the expansion of a coherency matrix. This polarimetric indicator is particularly useful to discriminate scattering mechanisms with different eigenvalue distributions but with similar intermediate entropy values. In such cases, a high anisotropy value indicates two dominant scattering mechanisms with equal probability and a less significant third mechanism, while a low anisotropy value corresponds to a dominant first scattering mechanism and two non-negligible secondary mechanisms with equal importance.

Among the different approaches tested, the best way to introduce the anisotropy information in the classification algorithm consists in implementing two successive segmentation procedures as shown in **Figure 11**.

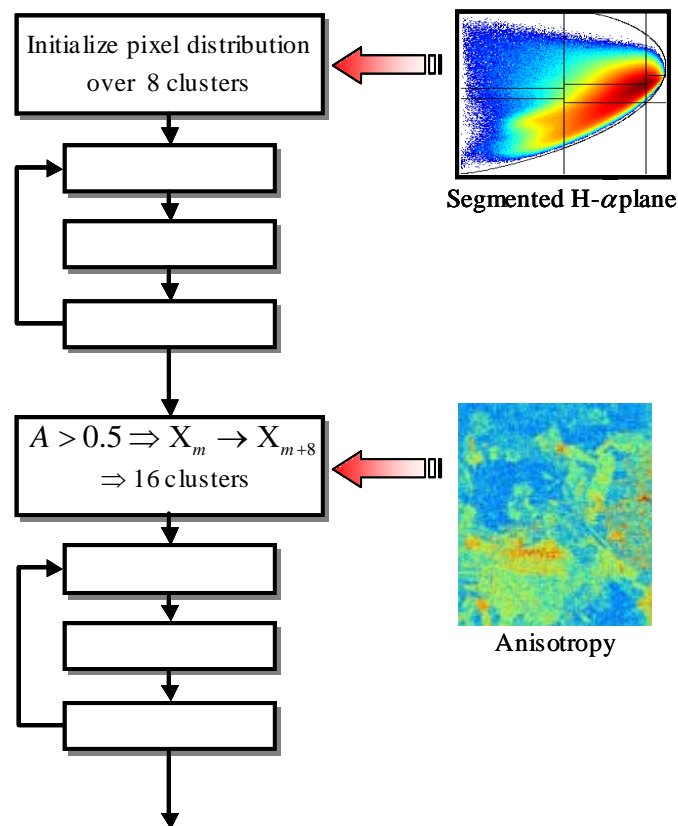


Figure 11 Unsupervised Wishart H-A- α segmentation scheme.

Polarimetric data are first segmented according to the algorithm presented in the former paragraph. Once this procedure has converged, the 8 resulting clusters are split into 16 ones

by comparing the anisotropy of each pixel to a threshold fixed to 0.5. The 16 segments are then used to initialize a second Wishart ML segmentation procedure.

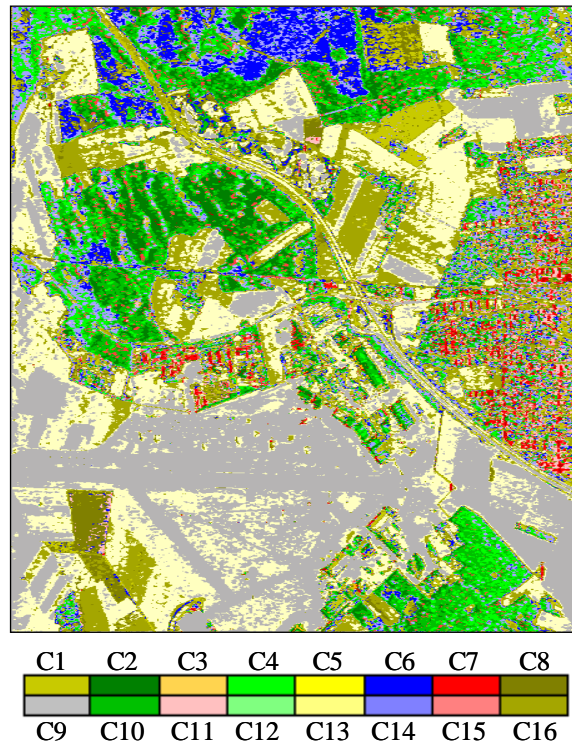


Figure 12 Wishart H-A- α segmentation results.

The segmentation results presented in **Figure 12** show an enhanced description of the Oberpfaffenhofen scene. The introduction of the anisotropy in the clustering process permits to split large segments into smaller clusters discriminating small disparities in a refined way. Several kinds of agricultural fields are separated. The runway and other low intensity targets are distinguished from other surfaces. Buildings are discriminated from other types of scatterers present in urban areas. The Wishart H-A- α classification scheme gathers into segments pixels with similar statistical properties, but does not provide any information concerning the nature of the scattering mechanism associated to each cluster.