Comparison of algorithms for the classification of polarimetric SAR data

V. Alberga[a,1], D. Borghys[a,2], G. Satalino[b,3], D. K. Staykova[c,4], A. Borghgraef[a,5], F. Lapierre [a,6], C. Perneel [d,7]

[a]Signal and Image Centre (SIC) - Royal Military Academy (RMA); Av. de la Renaissance 30, 1000 Brussels, Belgium

[b]Institute of Intelligent Systems for Automation (ISSIA) - National Research Council (CNR); Via G. Amendola 122/D-I, 70126 Bari, Italy

[c]Biophysics Group, Dept. of Chemistry - Göteborg University; Medicinaregatan 9C, SE-413 90 Göteborg, Sweden

[d]Royal Military Academy (RMA), Dept. of Applied Mathematics; Av. de la Renaissance 30, 1000 Brussels, Belgium

ABSTRACT

Most of the current SAR systems aquire fully polarimetric data where the obtained scattering information can be represented by various coherent and incoherent parameters. In previous contributions we reviewed these parameters in terms of their "utility" for landcover classification, here, we investigate their impact on several classification algoritms. Three classifiers: the minimum-distance classifier, a multi-layer perceptron (MLP) and one based on logistic regression (LR) were applied on an L-Band scene acquired by the E-SAR sensor. MLP and LR were chosen because they are robust w.r.t. the data statistics. An interesting result is that MLP gives better results on the coherent parameters while LR gives better results on the incoherent parameters.

Keywords: landcover classification, earth observation, remote sensing, image classification, synthetic aperture radar (SAR) polarimetry.

1. INTRODUCTION

Classification of land cover/land use constitutes one of the main applications of synthetic aperture radars (SAR); these remote sensing systems offer also the possibility of fully polarimetric acquisitions, i.e., they can operate ensuring that the full scattering information carried by the electromagnetic waves is measured and used.

In previous works,[1–4] the different possible representations of polarimetric SAR data were reviewed in terms of their "utility" when used for supervised landcover classification. For this purpose, classification tests were performed and measures of their accuracy used to quantify such "utility". It was possible to show that the various representations provide similar results, in terms of classification performance, making the choice of one rather than another practically equivalent. Here, we study more in detail the various approaches to their classification. In other words, we move our interest from the data to the methods to classify them.

The classifiers considered here are: a classical one, the *minimum distance* (*MD*) algorithm, a neural network, the *multi-layer perceptron* (*MLP*) trained by the *back-propagation* (*BP*) learning rule and one performing a *logistic regression* (*LR*). The reason for the choice of these algorithms is that none of them need any *a priori* knowledge on the statistics of the data to be classified. Hence, they can operate on data characterized by different distributions and the results compared.

E-mail contacts:
[1] lupo_ubriaco@yahoo.it, [2] dirk.borghys@elec.rma.ac.be, [3] satalino@ba.issia.cnr.it, [4] dkstaykova@gmail.com, [5] aborghgr@elec.rma.ac.be, [6] flapierr@elec.rma.ac.be, [7] Christiaan.Perneel@rma.ac.be

A data set from an airborne sensor, the German E-SAR, operating at L-band was used. The analysis procedure consists of the accuracy estimation of classification experiments and then of the quantitative comparison of the algorithms and their efficiency based on the accuracy estimates. An interesting result is, for instance, the variation of the overall performance of MLP and LR when dealing with different representations.

After introducing in section 2 the classifiers under analysis and in section 3 the polarimetric parameters considered for the tests, in section 4 details on the experimental data and on the approach to their classification will be given. Finally, in section 5 we will carry out data analysis and interpretation leaving conclusions to section 6.

2. CLASSIFICATION ALGORITHMS

2.1 Minimum distance

In classification applications, it is common practice to represent remotely sensed data through column vectors like:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \tag{1}$$

where $x_1, x_2, \cdots, x_n$ are the spectral values of the pixel vector $\mathbf{x}$ in bands 1 to n, respectively.

A classifier which needs only the mean vectors of the various classes to be estimated is the minimum distance, which should be more precisely indicated as *minimum distance to class means* classifier. Indeed, this algorithm operates by placing a pixel in the class of the nearest mean in the multispectral measurement space.[5]

An advantage of this method with respect, for example, to the maximum likelihood classification, is its computational speed. Moreover, the MD algorithm is efficient when the number of training samples per classes is limited, since it is based only on the use of the class mean vector $\mathbf{m}$ whose evaluation is still accurate even with few samples. The main disadvantage stems from the fact that it can model only symmetric classes in the multispectral space; hence, classes with a well-defined data spread in a particular spectral direction cannot be represented correctly.

Classification is performed on the basis of:

$$\mathbf{x} \in \omega_i \quad \text{if} \quad \mathrm{d}(\mathbf{x}, \mathbf{m}_i)^2 < \mathrm{d}(\mathbf{x}, \mathbf{m}_j)^2 \quad \forall\, j \neq i\,, \tag{2}$$

where $\mathrm{d}(\mathbf{x}, \mathbf{m}_i)^2$ is the squared Euclidean distance between the position of the generic pixel $\mathbf{x}$ to be classified and the mean value $\mathbf{m}_i$ of class ω_i. Since expanding the condition of (2) leads to a linear expression with respect to $\mathbf{x}$, the decision surfaces for this classifier are planes in the n-dimensional spectral space. The possibility of adopting a definition for the distance other than the Euclidean one is also given.

2.2 Multi-layer perceptron

The neural network classifier used in this work is the Multi-Layer Perceptron architecture, with one hidden layer, trained by the Back-Propagation learning rule. The MLP is a fully connected feed-forward neural network, composed of nodes arranged in layers. It can perform every non-linear input-output mapping, such as classification functions,[6] or more complex tasks such as the approximation of continuous functions.[7] For this purpose, it is necessary to submit the MLP to a training phase that searches the optimum set of weights minimizing a cost measure, usually given by the mean square error between estimated and expected outputs. This training phase, performed by the well-known BP learning rule,[6] requires a set of input-output examples. In this context, the input examples are obtained by using a supervised procedure that identifies on the mono- or multi-band image, and for each class of interest, a corresponding region of points. The inputs are the pixels themselves that contain the values corresponding to the measured parameters. Consequently, the outputs examples are given by the class labels j, $j = 1, .., n$ (being n the number of classes), represented by n-dimensional Boolean vectors.

The MLP dimension to be used is related to the current classification problem. In particular, the number of MLP input nodes corresponds to the number of polarimetric features under investigation, whereas the number

of output nodes is set equal to the number n of classes identified on the images. The performance of each trained MLP is estimated on an independent test data set.

2.3 Logistic regression

Logistic regression[8] was developed for dichotomous problems where a target class has to be distinguished from the background; Borghys et al.[9] applied it, for instance, for classification of multi-channel SAR data. Practically, LR's task is the search of the β_i's coefficients optimizing a non-linear function, the *logistic function*, defined as:

$$p\left(target \mid \mathbf{x}\right) = \frac{exp\left[\beta_o + \sum_i x_i \beta_i\right]}{1 + exp\left[\beta_o + \sum_i x_i \beta_i\right]} \,. \tag{3}$$

$p\left(target \mid \mathbf{x}\right)$ is the conditional probability that a pixel belongs to the considered class (*target* class) given the vector of input features $\mathbf{x}$ of a pixel. Dealing with a dichotomous problem, it follows that the probability of having a background pixel is given by:

$$p(background \mid \mathbf{x}) = 1 - p(target \mid \mathbf{x}) \,. \tag{4}$$

Hence, logistic regression finds a combination of the input features that optimises the log-likelihood defined as:

$$L(\beta) = \sum_{i=1}^{n} \left\{ y_i \ln\left[p(target \mid \mathbf{x})\right] + (1 - y_i) \ln\left[1 - p(target \mid \mathbf{x})\right] \right\} , \tag{5}$$

where y_i is equal to 1 for targets and to 0 for background pixels.

The logistic regression (i.e., the search for the β_i's) is carried out with the Wald's forward step-wise method using the commercial statistics software SPSS. In Wald's method, at each step, the most discriminant feature is added and the significance of adding it to the model is verified. This means that only the features that contribute significantly to the discrimination between the foreground and the background class are added to the model. LR thus implicitly performs a feature selection. Optimisation is performed on the pixels of the learning set and then the resulting logistic function (Equation (3)) is applied to the complete feature images; this results in detection images per class which are then combined into a classification map by assigning to each pixel the class for which the value in the detection image is the largest.

3. INVESTIGATED POLARIMETRIC PARAMETERS

As already mentioned, different representations exist of the scattering interactions of electromagnetic waves. For fully polarimetric radars, a substantial division can be made between the 2×2 *scattering matrix* $[\mathbf{S}]$ and the higher order ones, the *coherency* and *covariance* matrices. The former, and the parameters extracted from it, directly describe the scattering properties of a given resolution element in an imaged scene. The latter take into account the random nature of natural environments and try to render also this aspect in the parameters derived from the data. This implies averaging processes that, in the absence of regular data sampling, are performed spatially over neighboring scene elements assuming ergodicity of the data themselves.

Let us list the observables investigated in this work:

Elements of the covariance matrix. Second order matrices are necessary in order to take into account the statistical variations of the scattering interactions. The elements of these matrices are usually considered for practical applications and provided as input to classification algorithms according to various analysis strategies. Here, we took into account the covariance matrix.

Entropy/α parameters. This approach involves the generation of the coherency matrix and the calculation of its eigenvalues and of some related quantities.[10] The parameters derived in this way describe general characteristics and relevance of the elemental targets present in a SAR resolution cell.

Parameters of the target decomposition theorems. *Target decomposition* (*TD*) *theorems* provide an interpretation of the signals scattered from a target by considering them as superpositions of several contributions.[11] The application of such theorems leads to the evaluation of parameters expressing the intensity of the scattering mechanisms defining the signal detected by the sensor. Higher or lower intensity values of one component are connected to the predominance of a mechanism with respect to the others. In particular, *coherent* theorems decompose the scattering matrix [**S**], which is considered as a linear combination of several others; *incoherent* theorems operate, instead, on second order matrices. We used for our experiments the principal decomposition theorems: the Pauli decomposition and that of Cameron, as examples of coherent methods, and the Freeman decomposition as an example of the incoherent ones.

Scattering matrix elements. Different polarization bases providing complementary information may be taken into account; in the present study we referred to the horizontal-vertical (*hv*) linear basis.

4. DATA AND METHODOLOGY

For our tests, we had at disposal single-look complex data of the area of Oberpfaffenhofen, Germany, acquired by the E-SAR airborne sensor of the German Aerospace Centre (DLR) during a measurement campaign in October '99. They consist of L-band scattering matrices measured in the *hv*-basis. In terms of pixels, the data set has the following dimensions: 1000×4050, in range and azimuth, respectively.

Noise removal and speckle reduction were obtained by means of simple boxcar filtering with a fixed window dimensions. An averaging window of 5×11 pixels (range × azimuth) was adopted for all parameters. In this way, data processing could be performed with a consistent approach, i.e., choosing the same averaging window size to define the covariance and coherency matrices or to filter other data derived from either incoherent or coherent methods.

(a) (b)

Figure 1. Oberpfaffenhofen test site: (a) optical sensor picture (the bright rectangle indicates the area corresponding to the SAR data); (b) SAR backscattered intensity image and regions of interest: [1] "water", [2] "houses", [3] "roads", [4] "trees", [5] "grass", [6] "field 1", [7] "field 2".

The imaged area is situated approximately 25 km South West of the city of Munich and includes several interesting features: the DLR establishment, the former Fairchild Dornier aeroplane factory and the airfield shared by the two firms (upper left part of Figure 1 (a)). The village of Gilching is not far from them, on their right in the picture; a small lake is also present. Other important man-made structures are the motorway and the railway line stretching across the image. The vegetation patches consist of coniferous and mixed forests, meadows and crops.

A common analysis procedure was followed for all the polarimetric parameters:[4]

- at first, seven ground cover classes were defined: "water", "houses", "roads", "trees", "grass", "field 1" and "field 2", and for each of them, separated areas of training and test samples identified having a comparable number of pixels (for this purpose, aerial photographs and a topographic map were used as complementary sources of information);
- then, the training pixels from each class were fed to the classifiers to perform the training stage;
- as a following step, all the data were classified;
- finally, the test samples were used to measure the classification performance.

The accuracy of the classification tests was estimated by means of quantities typically used for this purpose, namely: *Kappa coefficient* (κ), *overall accuracy* (OA), *producer's accuracy* (PA) and *user's accuracy* (UA).[12] κ ranges from 0 to 1, the other three accuracy measures are expressed in percentage.

5. EXPERIMENTAL RESULTS

Let us report and comment here the results of our experiments. First some details and implementation issues regarding the different input parameter sets are given. In section 5.2 the actual results are presented and analysed. section 5.3 discusses some of the results and interpretes them.

5.1 Preliminary remarks regarding the input parameters

By definition, the **covariance matrix** is Hermitian positive semidefinite, hence, its symmetric elements are complex conjugates and only nine independent parameters are necessary in order to completely characterize it. For the classification tests, these nine parameters (the three real main diagonal elements and the real and imaginary parts of the three non-redundant off-diagonal elements) were given as input to the classifiers.

The **entropy, α and anisotropy ($H/\overline{\alpha}/A$)** parameters are, respectively, measures of the randomness of the scattering processes, of the relative importance of the scattering mechanisms and of their "type".[10] An average $\overline{\alpha}$ angle is normally used in polarimetric SAR data analysis.

As already said, target decomposition (TD) theorems permit to identify different scattering mechanisms corresponding to sets of theoretical models. This means that these methods "intrinsically" operate as classifiers since they recognize and weight the contributions of different ideal targets in a scene.

Referring as first to incoherent methods, we applied the **Freeman decomposition**. In this method the covariance matrix is decomposed in contributions of volume scattering, double bounce scattering and surface scattering. The total power of these three scattering contributions were given as input to the classifiers.

We studied also two coherent TD methods: the **Pauli decomposition** and that of Cameron.[11,13] The first method decomposes the measured [**S**] matrix into three contributions, three scattering matrices, due to a sphere and to two differently oriented diplanes, all of them weighted by complex coefficients. Their phase normalization was performed with respect to the phase of the sphere mechanism and the elements of the scattering matrices provided as input to the classifier (a similar procedure was followed also for the other coherent representations).

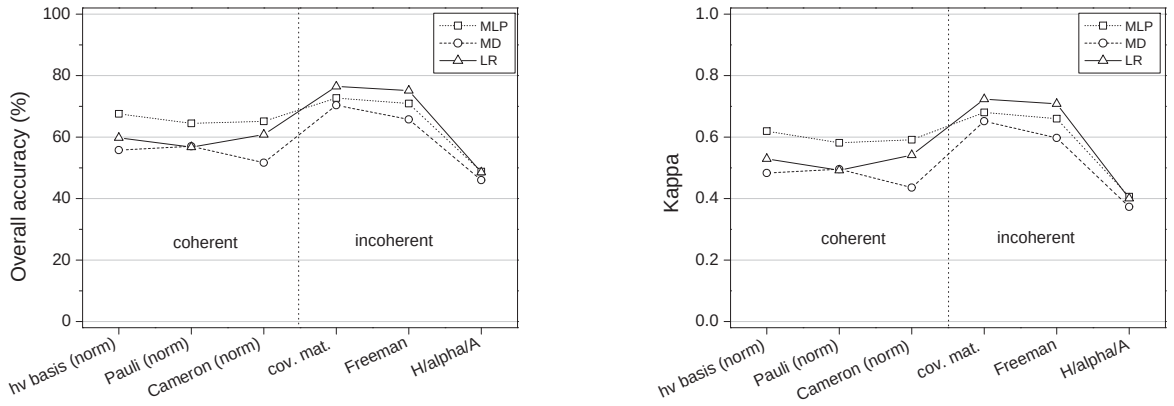
The **Cameron decomposition** considers both the form of the scattering matrix and the geometrical characteristics of the scatterer that it represents. Indeed, given a generic matrix [**S**] (not only in the monostatic case), one can characterize it by its tendency of being more or less symmetric according to the reciprocity rule and split it into two terms representing reciprocal and non-reciprocal scattering mechanisms. In turn, the reciprocal

term (the one usually available after calibrating monostatic radar data) represents a target which is more or less symmetric with respect to an axis in the plane orthogonal to the radar line-of-sight and again a distinction can be made between the most dominant and the least dominant symmetric target components.

The last series of tests was performed giving the elements of the **scattering matrix** (S-matrix), expressed in the hv linear polarization basis. Phase normalization was obtained with respect to the phase of the S_{hh} element and the normalized [**S**] matrix elements were then spatially averaged.

5.2 Results and Analyses

In Figure 2, the accuracies of the classification tests are summarized. For global comparison we considered the global κ-coefficient and overall accuracy. Different symbols in the figure denote the three different classifiers (MD: Circle, MLP:Square, LR:Triangle). Table 5.2 presents the same results numerically. The table also shows the average κ and OA as well as the standard deviation over the classifiers, in columns 8-12), and over the different parameter sets, in the 2 bottom rows of the table.



(a) (b)

Figure 2. Accuracy estimation of the classification tests on all the polarimetric parameters: (a) overall accuracy, (b) Kappa.

	MD		MLP		LR					
Parameters	κ	OA	κ	OA	κ	OA	$\overline{\kappa}$	$\sigma(\kappa)$	$\overline{OA}$	$\sigma(OA)$
S-matrix	0.48	55.8	0.62	67.6	0.53	59.8	0.54	0.07	61.1	6.00
Pauli	0.50	57.0	0.58	64.5	0.49	56.7	0.52	0.05	59.4	4.42
Cameron	0.44	51.7	0.60	65.1	0.54	60.8	0.53	0.08	59.2	6.84
Cov Mat	0.65	70.3	0.68	72.7	0.72	76.5	0.68	0.03	73.2	3.13
Freeman	0.60	65.7	0.66	70.9	0.71	75.1	0.66	0.05	70.6	4.71
$H/\overline{\alpha}/A$	0.37	46.0	0.41	48.8	0.40	48.6	0.39	0.02	47.8	1.56
Mean value	0.51	57.8	0.59	64.9	0.57	62.9				
St.Dev	0.10	8.94	0.09	8.52	0.13	10.9				

Table 1. κ-coefficient and overall accuracy overview.

From these global results it is interesting to notice that the multi-layer perceptron (MLP) gives the best results for the "coherent parameters", while the logistic regression (LR) gives better results for the classification based on the covariance matrix and the Freeman decomposition. The minimum distance classifiers, gives globally the least good results. The variation in global kappa or OA over the three classifiers is between 5 and 20%.

Of the six input parameter sets, the $H/\overline{\alpha}/A$ parameters give the lest good results for all classifiers. In evaluating the $H/\overline{\alpha}/A$ classification results one must be careful since a direct comparison of these observables with the other studied here is critical. Indeed, none of the $H/\overline{\alpha}/A$ parameters expresses an intensity or a power measurement: H and A are normalized quantities whereas α is an angle ranging from 0 to 90 degrees. For this reason, the literature often reports their combined use with the backscattered power.[14,15]

The overall best results are obtained by LR when the covariance matrix elements and the Freeman decomposition parameters are used as inputs.

The variation of results over the classifier is smallest with the $H/\overline{\alpha}/A$ parameters ($\sigma(\kappa) = 0.02$ on an average of 0.39, which is 5%) and largest with the cameron decomposition parameters (15%). The variation over the different parameter sets is larger than the variation over the different classifiers. This underlines the importance of the choice of input parameters in PolSAR landcover classification.

In order to obtain a more detailed overview of the results, the user's and producer's accuracies (UA and PA) are also determined per class. Ideally both UA and PA should be high. Figure 3 shows plots of UA versus PA. This allows to see both parameters at the same time. The figures are given for each set of input parameters seperately. The different symbols denote the three different classifiers. In each graph the results for the different landcover classes are presented. The results for each landcover type, obtained by the three classifiers, are connected by different types of lines. This allows to see the influence of the classifier choice on the classification results for each of the landcover classes.

UA and PA are interesting as seperate measures of quality of a classification because they each represent a different quality measure and for some applications one of the two is more important than the other. In our case, concerned with general landcover classification and without having a given application of the classification in mind, we consider UA and PA to be equally important. In this case it is possible to construct a single quality measure by combining UA and PA. Several ways to achieve this exist. In this paper we consider the arithmetic mean of UA and PA. Fig. 4 shows the average of UA and PA versus the land cover classes for the different input parameter sets.

A first observation from the figures is that, while the overall accuracy does not depend a lot of the choice of classifier, the results for each seperate class do change a lot as a function of the chosen classifier. In most cases both MLP and LR give better results than MD. From the global analysis based on κ and OA, it was concluded that MLP gives better results on the coherent parameter sets and LR on the non-coherent sets. The results shown in fig. 3 and fig. 4 allow to investigate this in more detail. The difference between both classifiers depends on the polarimetric parameters used as input as well as on the considered landcover class. When the S-matrix is used as input, the MLP gives better results for the classes Field1, Field2, Trees, Roads and Water, but LR gives the best results for Grass and Houses. For the Pauli parameters the relative results are similar. However, for the Cameron decomposition parameters the MLP gives the best results for all classes except water and roads.

For the non-coherent parameter sets globally LR gives better results than MLP, as shown in fig. 2. Analysis of fig. 3 and fig. 4 confirm this global result but also show some exceptions. For the elements of the covariance matrix LR gives better results than MLP for all classes except water and field2. For the Freeman parameters the exceptions are water, field2 and roads and for the $H/\overline{\alpha}/A$ parameters LR only gives the best results for water and grass. For this case the results of the LR and MD for grass are very similar.

The fact that the relative behaviour of the different classifiers with the different landcover classes seems to vary a lot with the choice of input parameters, is again an indication of the importance of the choice of the parameter set.

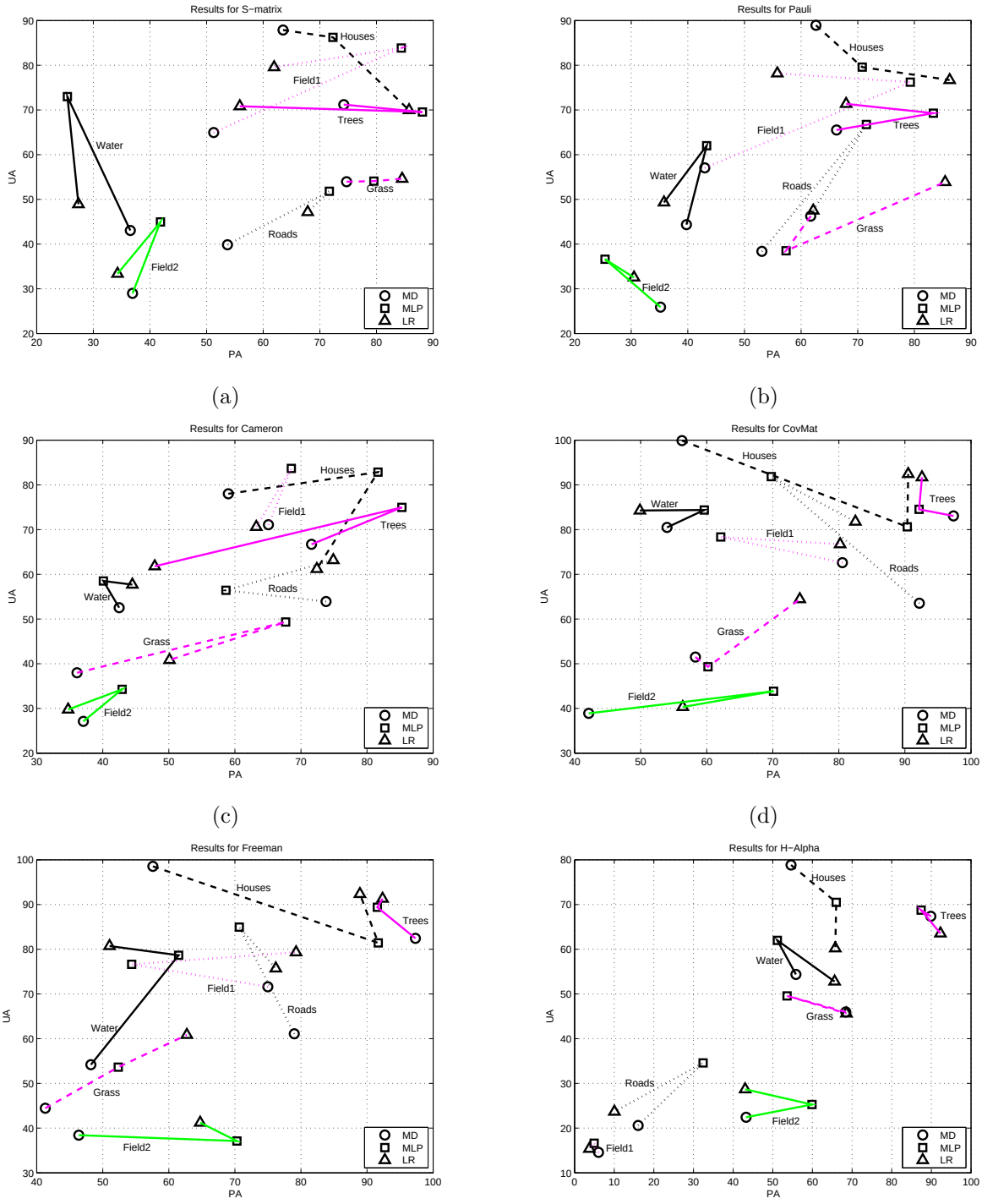
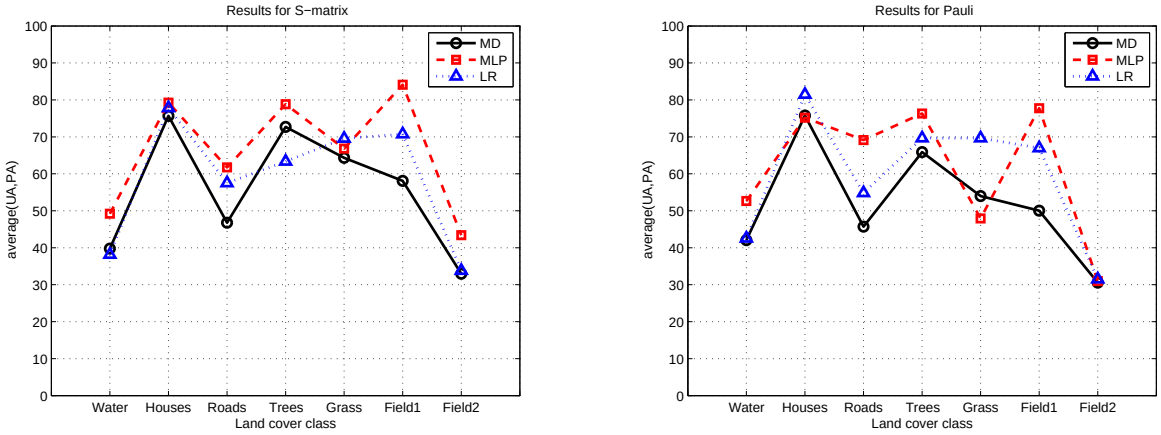
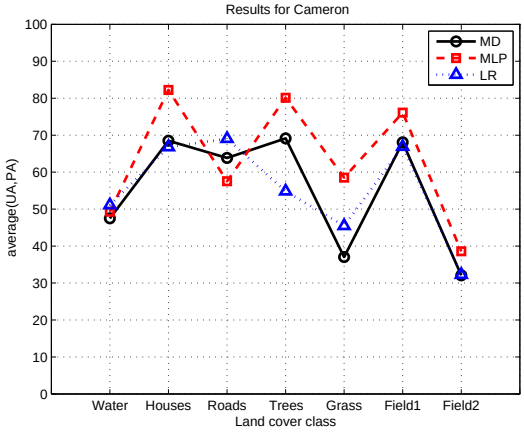


Figure 3. User's Accuracy vs. Producer's Accuracy: (a) [**S**] matrix, (b) Pauli decomposition, (c) Cameron decomposition, (d) Covariance matrix, (e) Freeman decomposition,(f) $H/\overline{\alpha}/A$.

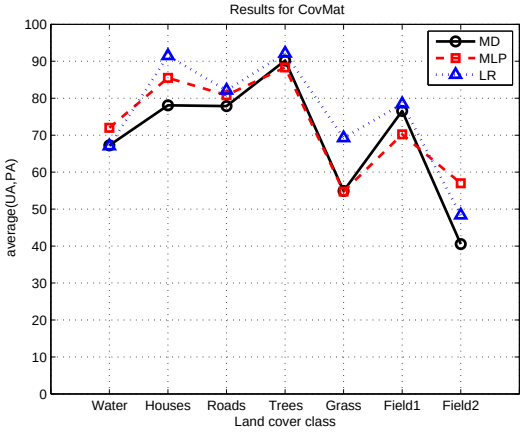


(a)

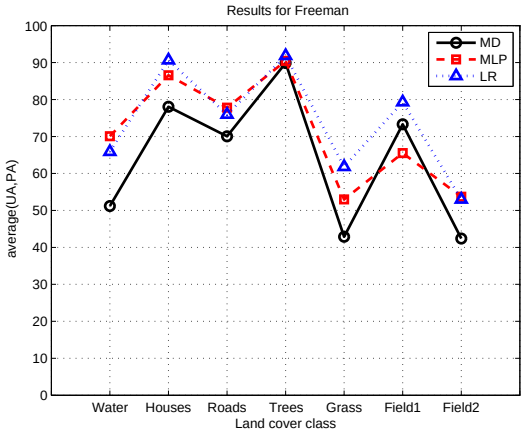
(b)



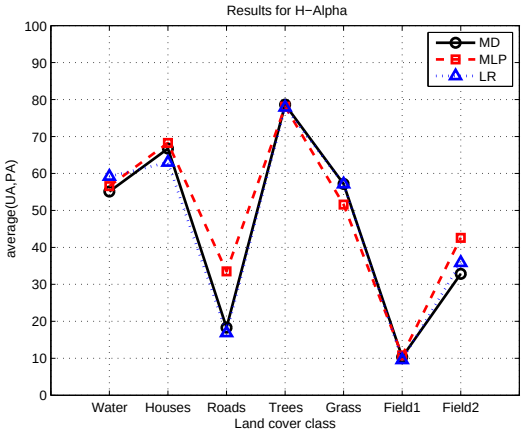
(c)



(d)



(e)



(f)

Figure 4. Average of UA and PA vs. the landcover class for: (a) [**S**] matrix, (b) Pauli decomposition, (c) Cameron decomposition, (d) Covariance matrix, (e) Freeman decomposition,(f) $H/\overline{\alpha}/A$.

5.3 Discussion and interpretation

Some of the studied representations provide comparable results at the global level (based on OA and global κ); only the entropy/α parameters clearly deviate from the trend of the others.

Indeed, the overall accuracy for the tests on the entropy/α parameters is always the worst. A possible interpretation is the following: the entropy and anisotropy parameters are derived by the eigenvalues of the coherency matrix; no use is made of the corresponding eigenvectors. One may then suppose that part of the information originally contained in the data has been "neglected" leading to poorer classification results. Another aspect to consider is the compression operated by eigenvalue analysis into fewer, dominant, components that could reduce detailed interpretation. Finally, H, $\overline{\alpha}$ and A do not express intensity or power measurements. However, also the unsupervised classifier by Cloude and Pottier,[10] which operates a division of the H/α plane into regions corresponding to different scattering characteristics, considers only these values disregarding part of the information content of the coherency matrix so that our approach does not diverge too much from established methods.

The choice of the classes and of the corresponding training data plays a relevant role: some of the ground cover types always give rise to problems in the classification processes and poor accuracy, whereas, for others, these are constantly acceptable. Hence, a detailed ground truth knowledge remains fundamental, in the sense that the selection of the training data and of the classification items may affect the final results sometimes as deeply as the choice of the type of the inputs (how the data should be represented). In almost all the tests with a given classifier, "houses" and "trees" were better recognized than the other classes. Referring to "houses", it is however interesting to observe that the LR and the MLP yield much better producer's accuracy than the MD when classifying the covariance matrix elements and the Freeman and Cameron decompositions terms (see fig. 3). The performance of the first two classifiers differs clearly from that of the minimum distance one when operating on these three sets of observables.

In the case of "water", there is a correlation between poor classification accuracy and low intensity of the signal typically returned by this class. Indeed, it usually shows weak backscattering, hence, a proportionally high noise contribution. This yields poor characterization of the class and limits the capability of the classifiers to recognize its members. In L-band SAR data, a similar reasoning can be made for the class "roads". The selected members of this class in our experiment are part of the runway of the airfield. This is a very smooth surface with a low radar return. When investigating the confusion matrices we see that in all classification results, the classes roads and water are very confused.

In general, the MLP and LR classifiers perform efficiently for all the parameters with the exception just of H, $\overline{\alpha}$ and A. Looking at Figure 2, one may note that the multi-layer perceptron is better than the logistic regression when classifying coherent observables, whereas it is worse when operating on incoherent parameters.

6. CONCLUSIONS

We have presented the results of a study aiming at a quantitative comparison classification algorithms operated on polarimetric SAR images. Since the tests were carried out on a relatively limited amount of data, our considerations do not presume to have an absolute validity but should indicate a tendency that is reasonable to assume valid for similar experimental conditions. We could observe that the performance of each classifier depends on the chosen input, i.e., the data representation. However, since the performance variations are relatively limited and not showing a precise trend, it is not possible at this stage to indicate an "absolute" best set of polarimetric observables.

It is important to remember that we used a straightforward approach to speckle reduction and that the classification accuracy may be improved using any of the techniques, which are currently intensively investigated, that actively re-define the averaging region boundaries. These algorithms present the advantage of effectively averaging homogeneous areas and reducing the smearing of the parameters due to the straddling of the fixed window on two or more different regions; our observations would remain valid and the use of each set of observables could be in this way further refined.

In the same way, we did not investigate here the possibilities of improvement offered by the spatial regularization of the classification maps. In fact, it is very unlikely that isolated pixels from one class appear in area

almost completely assigned to another class, so that local filters could be used to re-assign them according to homogeneity criteria of the image.

This work has further carried out a research on SAR polarimetry along previously traced guidelines.[1,2,4] The goal is a better understanding and a more efficient use of the different parameters derivable from fully polarimetric radar data. We mention finally that this paper extends and partially revises results reported in Alberga et al.[16] and in previous related works (e.g., the classification accuracy of the Pauli decomposition coefficients was there underestimated).

ACKNOWLEDGMENTS

Thanks to the German Aerospace Centre (DLR) for kindly proving the experimental data.

Part of the SAR data processing was performed using the free RAT software developed at the Department of Computer Vision and Remote Sensing of the Technical University of Berlin, Germany; the authors would like to acknowledge the excellent service rendered by the RAT team.

REFERENCES

[1] Alberga, V., Satalino, G., and Staykova, D. K., "Polarimetric sar observables for land cover classification: analyses and comparisons," in [*Proceedings of SPIE - SAR Image Analysis, Modeling, and Techniques VIII*], **6363** (September 2006). DOI: 636305.

[2] Alberga, V., "A study of land cover classification using polarimetric sar parameters," *Int. Journal of Remote Sensing* **28(17)**, 3851 – 3870 (September 2007).

[3] Borghys, D. and Perneel, C., "A supervised classification of multi-channel high-resolution SAR data," *Earsel eProceedings* **6(1)**, 26–37 (2007).

[4] Alberga, V., Satalino, G., and Staykova, D. K., "Comparison of polarimetric sar parameters in terms of classification performance," *Int. Journal of Remote Sensing* **29(14)**, 4129 – 4150 (July 2008).

[5] Richards, J. A., [*Remote sensing digital image analysis*], Springer-Verlag, Berlin, Germany (1994).

[6] Hertz, A., Krogh, A., and Palmer, R. G., [*Introduction to the theory of neural computation*], Addison Wesley, Redwood City, CA, USA (1991).

[7] Funahashi, K., "On the approximate realizations of continuous mappings by neural networks," *Neural Networks* **2(3)**, 183–192 (1989).

[8] Hosmer, D. W. and Lemeshow, S., [*Applied logistic regression*], John Wiley & Sons, New York, NY , USA (1989).

[9] Borghys, D., Yvinec, Y., Perneel, C., Pizurica, A., and Philips, W., "Supervised feature-based classification of multi-channel sar images," *Pattern Recognition Letters: Special Issue on Pattern Recognition for Remote Sensing* **27(4)**, 252–258 (March 2006).

[10] Cloude, S. R. and Pottier, E., "An entropy based classification scheme for land application of polarimetric sar," *IEEE Transactions on Geoscience and Remote Sensing* **35(1)**, 68–78 (January 1997).

[11] Cloude, S. R. and Pottier, E., "A review of target decomposition theorems in radar polarimetry," *IEEE Transactions on Geoscience and Remote Sensing* **34(2)**, 498–518 (March 1996).

[12] Congalton, R. G., "A review of assessing the accuracy of classification of remotely sensed data," *Remote Sensing of Environment,* **37**, 35–46 (1991).

[13] Cameron, W. L., Youssef, N. N., and Leung, L. K., "Simulated polarimetric signatures of primitive geometrical shapes," *IEEE Transactions on Geoscience and Remote Sensing* **34(3)**, 793–803 (May 1996).

[14] Kimura, K., Yamaguchi, Y., and Yamada, H., "Pi-sar image analysis using polarimetric scattering parameters and total power," in [*Proc. IGARSS*], 425–427 (July 2003).

[15] Spencer, P., Alberga, V., Chandra, M., Hounam, D., and Keydel, W., "Sar polarimetric parameters for land use classification," in [*Proc. EUSAR 2000*], 783–785 (May 2000).

[16] Alberga, V., Staykova, D., Krogager, E., Danklmayer, A., and Chandra, M., "Comparison of methods for extracting and utilizing radar target characteristic parameters," in [*Proc. IGARSS 2005*], **3**, 2019 – 2021 (July 2005).