CLASSIFICATION OF FULLY POLARIMETRIC SAR DATA FOR LAND USE CARTOGRAPHY

Cédric Lardeux [a], Pierre-Louis Frison [a], Jean-Paul. Rudant [a], Jean-Claude Souyris [b], Céline Tison [b], Benoît Stoll [c]

[a] Université de Marne-la-Vallée, 77245 Champs sur Marne, France
[b] CNES - DCT/SI/AR, 18 avenue Edouard Belin, 31 401 Toulouse Cedex 4, France
[c] Université de la Polynésie française, B.P. 6570 98702 FAA'A Aeroport Tahiti - Polynésie Française

KEY WORDS: polarimetry, land use, svm, classification

ABSTRACT:

French Polynesia islands are located at the middle of the South Pacific Ocean. They are thus subject to a strong environmental planning leading to landscape changes as well as to the introduction of invasive species. This study comes within the framework of the global cartography and inventory of the Polynesian landscape. An AIRSAR airborne mission took place in August 2000 over the main Polynesian islands. Polarimetric SAR data are particularly adapted to the cloudy conditions generally encountered over the South Pacific Islands. Fully polarimetric data allows the analysis of a geometrical and physical point of view. Different decompositions, such as H/A/α or based on the Pauli formalism have shown their potential for such applications. In order to apply these indicators and to produce a semi-automatic cartography of the Tubuai Island, we choose to use the SVM (Support Vector Machine) as supervised classifier. These results are also compared with the Wishart classifier based on the analysis of the polarimetric coherency matrix only. As our full polarimetric data are also available in P and L bands this study evaluates the contribution of the different wavelength. This study shown that the combination of SVM and full polarimetric data, associates with different wavelength, gives promising results.

1. INTRODUCTION

Radar data are of particular interest over tropical areas such the French Polynesian Islands due to the cloudy conditions generally persistent. Fully polarimetric SAR data were acquired in L and P bands over the main Polynesian islands. The overall goal of this study is to assess the potential of such fully polarimetric SAR data for land-use cartography. When dealing with classification methods applied to full polarimetric data, the Wishart classification (Lee et al., 2004) or other, based for example on the H/A/α decomposition (Cloude and Pottier, 1995) are generally used. In order to integrate different polarimetric descriptors, not only the elements defining the coherence matrix used in Wishart classification, but also other polarimetric descriptors, such the H/A/α parameters, it is proposed in this study to investigate the SVM (Support Vector Machine) classification method. It is especially well suited to handle linearly non separable case by using the Kernel theory (Burges, 1998), and has been mostly applied to hyperspectral remote sensed data. However few studies has also been conducted with SAR data (Fukuda and Hirosawa, 2001), (Mercier and Girard-Ardhuin, 2005). The study area and radar data are detailed in the second part of this paper. The third part describes the polarimetric parameters involved in the classification method and briefly presents the principle of the SVM method. Then, results of the SVM classification are discussed in relation with the definition of the Support Vector that has been made.

2. STUDY AREA AND DATASET

2.1 Study area

French Polynesia islands are located at the middle of the South Pacific Ocean. They are quickly evolving in the tourism industry, and from the economic and geostrategic points of view. They are thus subject to a strong environmental planning leading to landscape changes as well as to the introduction of invasive species. This study comes within the framework of the global cartography and inventory of the Polynesian landscape. We focus on data acquired over the Tubuai island, in the Australes Archipelago at the South of French polynesia. Tubuai is a 45 km^2 island with a population of about 6000 inhabitants. It is particularly relevant because of its great landscape diversity: several types of forests, agricultural fields, and residential areas.

In our application we would discriminate different kind of landscape. We choose to produce a map with seven classes of four types. The first and the more difficult to discriminate is the forested area with four classes :*Hibiscus tiliaceus* (also called Purau), *Pinus Caribeae* (also called Pinus), *Paraserianthes Falcataria* (also called Falcata) and *Psidium cattleianum*. The second type is the "Low Vegetation" class that includes vegetation up to approximately one-meter height: fern lands, swamps vegetation, and crops. The other type of class is "No Vegetation" class which includes the bare fields and low grass fields. The last class is the sea.

Several ground surveys has been carried out, the last one in July 2005, which, combined to a Quickbird image acquired in August 2004 shown in fig 1, allowed to give a ground truth over the entire island.

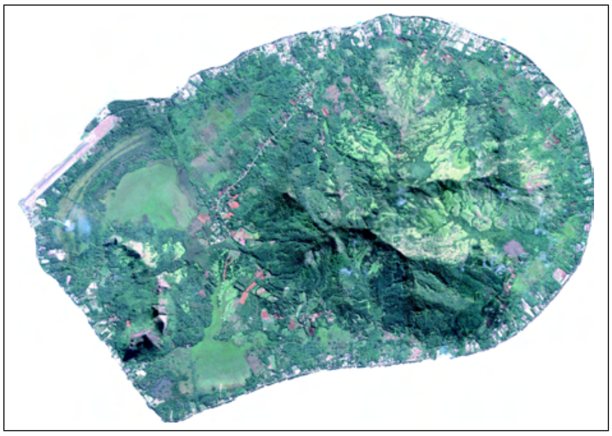

Figure 1. Tubuai Island Quickbird image

2.2 Airsar data

An AIRSAR airborne mission took place in August 2000 over the main Polynesian islands. The AIRSAR data were acquired over Tubuai along 2 passes in reverse path, in Polsar mode. Consequently, the data set consists in full polarimetric data in L (λ=23cm) and P (λ=67cm) bands, with an additional TopSAR C (λ=5.7cm) band channel in VV polarisation.

Data are delivered in MLC (Multi Look Complex) format, corresponding to about 9 looks, with a resolution of 5 meters.

The fig 2 shown an Airsar composite.

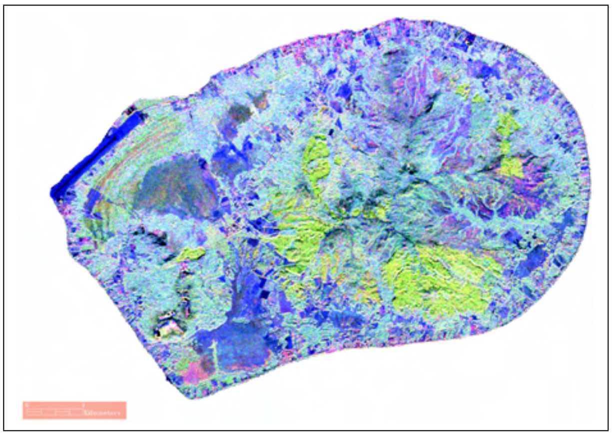

Figure 2. Tubuai AIRSAR Composite (R : L-HH G: L-HV B: C-VV)

3. METHODOLOGY

3.1 Polarimetric indicators

For each pixel, the coherency matrix T is derived from the Stokes parameters given by AIRSAR data. Even if the elements of the T matrix give the entire polarimetric information, additional polarimetric parameters are derived: these are the H/A/α coefficients (Cloude and Pottier, 1995), the polarimetric coherence between co-polarized circular polarizations, ρ_{rrll} (Mattia et al., 1997), where LL (resp. RR) stands for Left-Left (resp. Right-Right) polarization. Other indicators such as the Span, and the intensities of circular polarizations have also been investigated. A more detailed description of these parameters is given below:

3.11 T The coherency matrix is constructed from a scattering vector in the base of Pauli that reflect geometrical properties (Cloude and Pottier, 1995).

$$k_p = \frac{1}{\sqrt{2}} \begin{pmatrix} S_{HH} + S_{VV} \\ S_{HH} - S_{VV} \\ 2S_{HV} \end{pmatrix}, \; [T] = k_p.k_p^{*T} \tag{1}$$

3.12 H/A/α This three parameters are generated by a decomposition in eigenvector of T matrix (Cloude and Pottier, 1995).

- The entropy H characterizes the wave depolarization and is very useful to discriminate for example the forest non forest area each characterized by high and low entropy respectivly.
- The α " The α parameter is particularly interesting because it gives the reflection mechanisms of the wave over the considered pixels. It characterizes the double bound, single bound, and volume scattering. It is meaningful for low entropy values, indicating that the polarimetric information is significant.
- The Anisotropy parameter permit to give a difference between the second and third eigenvalue (mechanism) and is meaningfull for $0.7 < entropy < 0.9$.

3.13 ρ_{rrll} It has been shown that this parameter is well suited for bare soil surfaces, as it allows soil roughness discrimination while giving a low sensitivity to soil moisture (Mattia et al., 1997).

$$|\rho_{rrll}| = \sqrt{\frac{\langle |S_{rr}.S_{ll}^{*}|^2 \rangle}{\langle |S_{rr}|^2 \rangle . \langle |S_{ll}|^2 \rangle}} \tag{2}$$

3.14 Span The Span characterizes the total backscattering power that are reflected.

$$SPAN = |S_{hh}|^2 + |S_{vv}|^2 + 2|S_{hv}|^2 \tag{3}$$

3.15 Intensities in circular basis We have computed the intensities of the co and cross polarization state in the circular basis (Left and Right) :

$$|S_{ll}|^2, |S_{rr}|^2, |S_{rl}|^2 \tag{4}$$

3.2 Support Vector Machine

A brief description of SVM is made here and more details can be found in (Burges, 1998).

3.21 Linear case We should now consider the case of two classes problem with N training samples. Each samples are described by a Support Vector (SV) X_i composed by the different "band" with n dimensions. The label of a sample is Y_i. For a two classes case we consider the label -1 for the first class and +1 for the other.

The SVM classifier consist in defining the function $f(x) = sign(\langle \omega, X \rangle + b)$ that found the optimum separating hyperplane as presented in Fig. 3.

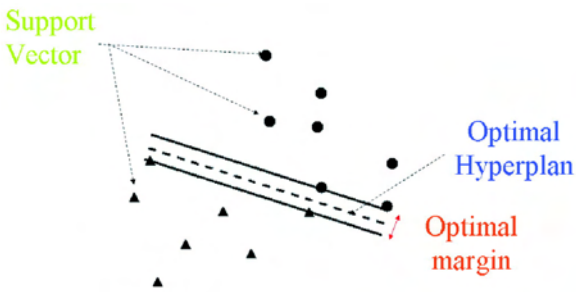

Figure 3. SVM Classifier-Linear case

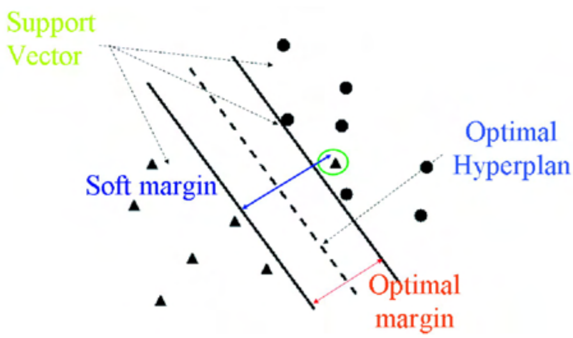

Figure 4. SVM Classifier-Nonlinear case

The sign of f(x) gives the label of the sample. The goal of the SVM is to maximize the margin between the optimal hyperplane and the support vector. So we search the $\min \frac{\|\omega\|}{2}$.

To do this, it is more easier to use the Lagrange multiplier. The problem comes to solve :

$$f(x) = Sign(\sum_{i=1}^{Ns} y_i.\alpha_i \langle x, x_i \rangle + b) \quad (5)$$

where α_i is the Lagrange multiplier

3.22 Nonlinear case If the case is nonlinear as the Fig. 4 the first solution is to make soft margin that is particularly adapted to noised data. The second solution that is the particularity of SVM is to use a kernel. The kernel is a function that simulates the projection of the initial data in a feature space with higher dimension $\phi : \Re^n \mapsto H$. In this new space the data are considered as linearly separable. To apply this, the dot product $\langle x_i, x_j \rangle$ is replaced by the function

$$K(x, x_i) = \langle \phi(x), \phi(x_i) \rangle$$

Then the new function to classify the data are :

$$f(x) = Sign(\sum_{i=1}^{Ns} y_i.\alpha_i.K(x, x_i) + b) \quad (6)$$

Three kernel are commonly used :

- The polynomial kernel $K(x, x_i) = (\langle x.x_i \rangle + 1)^p$
- The sigmoid kernel $K(x, x_i) = \tanh(\langle x.x_i \rangle + 1)$
- The RBF kernel $K(x, x_i) = \exp^{-\frac{|x-x_i|^2}{2\sigma^2}}$

Due to the best results given for the present study, the RBF kernel has been retained here. A future work would be to develop a new kernel accounting for the distribution of the data, such the one is due to the presence of speckle in SAR data.

3.23 Multiclass case The principle of SVM was described for a binary classification, but many problems have more than two classes problem. There exists different algorithms to multiclass problem as "One Against All" (OAA) and "One Against One" (OAO).

If we consider a problem with K class :

OAA algorithm consists in the construction of k hyperplane that separate respectively one class and the (k-1) other classes.

OAO algorithm consists in the construction of $\frac{k(k-1)}{2}$ hyperplane which separate each pair of classes.

In the two cases the final label is that mainly chosen.

3.24 Wishart classification The Wishart classification involved only the T matrix elements especially dedicated to SAR data as it accounts for the Wishart distribution observed due to the presence of speckle noise.

For the monostatic case, the polarimetric information is define by the target vector h :

$$h = \begin{pmatrix} S_{HH} \\ \sqrt{2}.S_{HV} \\ S_{VV} \end{pmatrix} \quad (7)$$

For multilook data that is the most currently case we represent the data by a polarimetric covariance matrix Z

$$Z = \frac{1}{n}.\sum_{k=1}^{n} h_k h_k^{*T} \quad (8)$$

Where h_k is the kth sample of h, the superscript * denote the complex conjugate and n is the number of looks (samples). The covariance matrix could be describe by the Wishart distribution as :

$$p(Z) = \frac{n^{qn}|Z|^{n-q} \exp^{-tr(n\sum^{-1} Z)}}{K(n,q)|\sum|^n} \quad (9)$$

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

Where $\Gamma()$ is the gamma function and tr() is the trace of the matrix. q represent the number of elements of the target vector h (3 for monostatic case and 4 for the bistatic case). n represent the number of looks.

The Wishart classification consist in a maximum likelihood classification based on a Wishart distribution.

4. RESULTS AND DISCUSSION

4.1 Methodology

We now present the schema of the classification and the different parameters that are used to compute the classification. Table 1 presents the training and control samples number made with the ground truth data.

Table 1. Training and testing samples number used for the Tubuai Island classification

Sort	Training sample	Control sample
Pinus	5854	5854
Falcata	2779	2779
Purau	6382	6382
Psidium cattleianum	367	367
Low Vegetation	9014	9014
No Vegetation	4608	24.45
Sea	98343	98343

The SVM classification was produced with the Libsvm library (1). In a first time, results obtained with the SVM method when considering only the T matrix elements are compared to Wishart supervised classification. This later has been performed using PolSarPro software (2). In a second time, several Support Vector corresponding to the combination of the different polarimetric descriptors given in § 3.1 have been tested for L and P band tested individually. Finally, different combination of Support Vectors merging the L, P, and C bands parameters were evaluated. Concerning the multiclass case considered for the SVM method, after several tests, the OAO algorithm has been retained as well as the RBF kernel with σ = 0.5 and the cost parameter equal to 1000 (soft margin).

The results of the different classification algorithms have been evaluated using the Producer's Accuracy (PA) and the User's Accuracy (UA), as described in (Oruc, Marangoz and Buyuksalih, 2004):

- $PA = \frac{Pixel\ correctly\ classify}{Pixel\ of\ the\ class}$
- $UA = \frac{Pixel\ correctly\ classify}{Pixel\ labeled\ as\ this\ class}$

An ideal classification would give PA=1 and UA =1.

- PA<1 means that some pixels of the ground truth are not correctly classify .
- UA<1 is a criterion of overclassification.

  It means that they are more pixels in the other classes of the ground truth than in the concerned class.

Due to the different pixel numbers involved in the different control classes, the mean of the PA (MPA) of the classes is preferred to the commonly used overall accuracy.

4.2 Discussion

4.21 SVM vs Wishart comparison The comparison between SVM classification with Support Vector consisting only of the T matrix elements and the Wishart classification gives similar results for L band (MPA= 57.50 vs 56.83 resp.). At P band, a slighty better result for Wishart classifier can be observed (MPA = 60.63 vs 56.44 resp.). This is due a confusion by SVM classifier between "Low vegetation" and "Sea" that does not occur with Wishart classifier.

4.22 Influence of the Support Vector The influence of the Support Vector configuration when regarding L or P band separately, as well as merging them together with C band has been assessed. 3 different Support Vector configurations were tested:

- C1: only the T matrix elements
- C2: all the parameters presented in 3.1

Overall results are given in Table 2.

When only one band is considered, the C2 with respect to C1 configuration enhances the MPA of 9% and 16% for L and P band respectively.

Table 2. Accuracy results of different Support Vector configurations

Bands	L		P		L+P	L+P+ Cvv
SV conf.	C1	C2	C1	C2	C2	C2
MPA	56.50	65.94	52.75	68.75	84.03	91.11

Results obtained over the different classes for L band are detailed in Table 3

As expected, the best results are obtained over low and no vegetated classes, at the exception of Purau one, which is the dominant of the 4 sub-classes belonging to the "densely vegetated" class. The poor results obtained for the *Psidium cattleianum* class may be due to the small relative pixel number of this class.

Table 3. Accuracy results of the SVM classifier for some configurations in L band

Band	L			
SV Conf.	C1		C2	
Accuracy	PA	UA	PA	UA
Pinus	47.75	52.18	58.54	59.58
Falcata	16.98	50.32	28.21	57.69
Purau	72.50	51.95	72.83	57.75
Psidium cattleianum	0	0	19.07	70.00
Low Vegetation	91.45	90.32	96.62	93.69
No Vegetation	66.86	81.94	89.37	91.74
Sea	99.96	99.16	99.87	99.91
MPA	56.50		66.36	

On the other hand, as expected, the combination of the 3 bands gives the best results as shown in table 4. The addition of the single VV channel of the C band is significant as it allows the discrimination between Pinus and Falcata species. Indeed, the crown between Pinus, Falcata and Purau are very different that seems enhanced the Cvv band.

The corresponding classification image is showed in Fig. 5, confirming the quality of the results with respect to ground truth data.

Table 4. Accuracy results of SVM classification with L+P+Cvv in C2 configuration

Sort	PA	UA
Pinus	98.43	99.17
Falcata	76.83	82.34
Purau	93.37	89.78
Psidium cattleianum	73.84	80.65
Low Vegetation	99.01	97.86
No Vegetation	96.27	98.47
Sea	99.99	100.00
MPA	91.11	

5. CONCLUSION

SVM classification is applied to full polarimetric SAR data over the Tubuai Island, French Polynesia, which is mainly constituted of dense vegetation. When only polarimetric coherency matrix elements are considered, SVM algorithm give similar results than Wishart classifier for L band, and slightly less accurate results for P band. The best configuration for the Support Vector consists in the T matrix elements, in addition to other polarimetric descriptors such H, A, α, circular polarisation intensities, the SPAN, and the copolarised circular polarisation correlation coefficient ρ_{rrll}. The combination of the 3 bands improve significantly the results, even the C band single VV channel that allows to discriminate

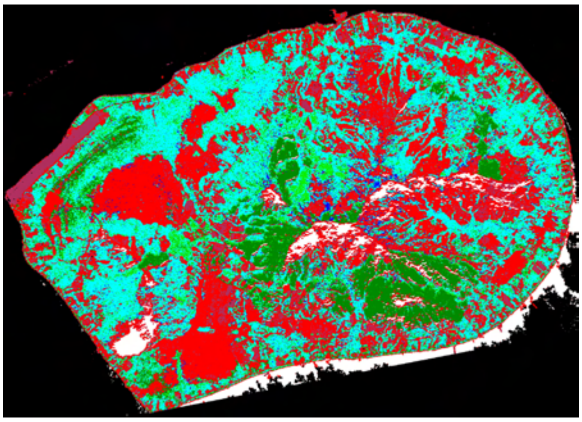

Figure 5. Result of SVM classification with L+P+Cvv in C2 configuration

Pinus	Falcata	Purau	Psidium cattleianum
Low Vegetation	No Vegetation	See	Unclassified

to dense vegetation classes, i.e. the Pinus and Falcata species. We see finally that the use of Polarimetric SAR data is relevant for land use classification. The presented methods might be of interest for investigations in agricultural field and urban areas as well.

ACKNOWLEDGEMENTS

The authors are very grateful to Jean-Yves Meyer for for the survey mission and we would like to thank the Government of French Polynesia and its Urbanism Department for providing the AirSAR, MASTER and Quickbird data required for this study.

REFERENCES

C. J. Burges, *A tutorial on support vector machines for pattern recognition*, in Data mining and knowledge discovery, U. Fayyad, Ed. Kluwer Academic, 1998, pp. 1-43.

S. R. Cloude, E. Pottier *A Review of Target Decomposition Theorems in Radar Polarimetry*, IEEE TGRS, vol. 34, no. 2, pp 498-518, Sept. 1995.

S. Fukuda and H. Hirosawa *Support Vector Machine Classification of Land Cover: Application to Polarimetric SAR Data* in Geoscience and Remote Sensing Symposium, 2001. IGARSS '01. IEEE 2001 International , 9-13 July 2001 Page(s):187 - 189 vol.1

J-S.Lee, Mitchell R. Grunes, E.Pottier, L.Ferro-Famil *Unsupervised Terrain Classification Preserving Polarimetric Scattering Characteristics*, IEEE TTGRS, vol. 42, no. 4, April 2004

Libsvm is available at http://www.csie.ntu.edu.tw/cjlin/libsvm

F.Mattia, T.Le Toan, J-C.Souyris, G.De Carolis, N.Floury, F.Posa, G.Pasquariello, *The Effect of Surface Roughness on Multifrequency Polarimetric SAR Data*, IEEE TGRS, Volume 35, NO. 4, July 1997

G.Mercier and F.Girard-Ardhuin, *Unsupervised Oil Slick Detection by SAR Imagery using Kernel Expansion*, Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International vol.1, 25-29 July 2005 Page(s):494 - 497

M.Oruc, A.M.Marangoz, G.Buyuksalih, *Comparison of pixel-based and object-oriented classification approaches using Landsat-7 Etm spectral bands*, ISPRS July 2004

ESA,*PolSarPro*, available at http://earth.esa.int/polsarpro/