MULTI-SOURCE SVM FUSION FOR ENVIRONMENTAL MONITORING IN MARQUESAS ARCHIPELAGO

Robin Pouteau, Benoît Stoll, Sébastien Chabrier

South Pacific Geosciences (GePaSud) Laboratory - French Polynesia University (UPF)
BP 6570, 98702 FAA'A - TAHITI - French Polynesia

ABSTRACT

Mapping vegetal species in natural ecosystems needs the use of several complementary - structural as functional - information sources to be as accurate as possible. SVM fusion is an adapted tool for multi-source classification. Most of the contemporaneous studies on land use classification concern cultivated fields. This study shows that, with the increasing development of adapted statistical tools and various remotely sensed data, it is now possible to map natural ecosystems, a key stage for eco-environmental management.

Index Terms— Data fusion, multi-source imagery, support vector machines (SVM), optical data, synthetic aperture radar (SAR) data, topography

1. INTRODUCTION

Several researches state that the combined use of multi-source remote sensing data improves accuracy of land cover classification, using different methods [1]-[5]. Most of them are compared in [6], namely maximum likelihood, decision trees and support vector machines (SVM), showing that SVM gives the best accuracy.

Multi-source SVM fusion is certainly a key research subject in remote sensing sciences for the future, potentially allowing to downscale class sets and to handle the most complex structures. The aim of this work is to explore the contribution of multi-source SVM fusion for mapping and monitoring the Marquesas islands landscapes.

2. MATERIAL AND METHODS

Nuku Hiva island is a good study model for the Marquesas archipelago in term of alien species invasion which is arguably one of the major threat to native ecosystems [7]. There is a particular need for better quality and more information on the distribution and impact of invasive species in order to improve policy, legislation and implementation procedures against these aliens.

SVM [8] are chosen because they perform more accurately than other classifier in mono-source, e.g. [9]-[11],
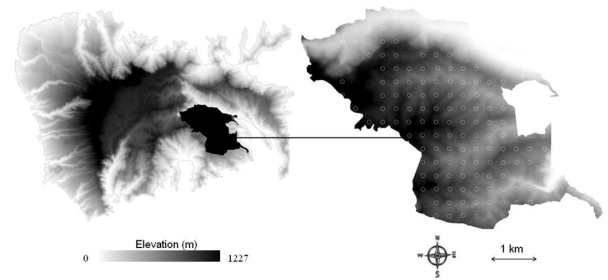

Figure 1: The study site: the Nuku Hiva Island (left) and its Baie du Contrôleur Domain (right), an ecologically disturbed area. Each circle in the right map represents a sampling area.

and multi-source cases [6]. They are frequently used with three kinds of nonlinear kernels: the polynomial, the Gaussian radial basis function and the sigmoid. Since these kernels have rarely been studied in a multi-source remote sensing problem, we compare their outputted accuracy.

Remote sensing is a useful tool for ecosystem mapping for three main reasons. First, in mountainous areas such as Pacific volcanic islands, access is often limited and resources are difficult to evaluate in situ. Secondly, vegetation structural complexity needs an integrative approach (as pixels) to be understood. Finally, affecting ecological parameters such as luminosity, nitrogen availability or water resources, the above vegetation stratum structures the underlying ones; remote sensing just informs synoptically about the above vegetation stratum. Thus, remotely sensed data and ground truth can be efficiently linked emphasizing main remotely sensible vegetation characteristics.

A first ground truth campaign is carried out to look for a representative class set. Vegetal community is commonly divided by phyto-sociologists into 3 strata: herbaceous (<1m), shrubs (1-5m) and trees (>5m) [12]. To characterize vegetation composing the study area, 143 inventories in the commonly used surfaces - 100m² for herbaceous plant community and 450m² for shrub and tree ones - are sampled systematically in a mesh network from an initial random point (figure 1). Distance between two sampling area is 300m. For each inventoried species in each stratum, an
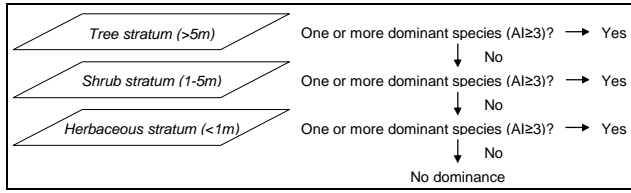
Figure 2: Process used to reach the dominant highest species visible on the remotely sensed scenes; AI=abundance index.

abundance index (from 0 to 5) is inputted. Then, we compute the process presented in figure 2 aiming to select the dominant highest vegetation stratum only for each plant community i.e. emergent species in the remotely sensed images (Table 1).

In a second ground truth campaign, 36 training plots of 450m² (~1‰ of the site study), three per class, are selected and geolocalised with a GeoXH Trimble GPS. Such balanced datasets are used to avoid class over- or under-representation problems [13]. For classification assessment, 36 validation plots are sampled.

3. MULTISOURCE CLASSIFICATION

Three complementary multi-sensor structural and functional information sources are used for the analyses:

- Optical data such as IKONOS satellite scenes from 2005 inform about vegetation texture and passive absorption spectra. The 1 m-merged data set (3 bands multispectral) spectral resolution is $\lambda=0.45$-0.72 μm, i.e. the visible spectrum. The high spatial resolution of IKONOS imagery gives useful details for species discrimination by computing some gray level co-occurrence matrix (GLCM) texture metrics [14]. Four GLCM texture metrics among variance, contrast, dissimilarity and angular second moment are computed by using three window sizes of 3x3, 9x9 and 15x15 pixels which potentially correspond to intra-tree micro-texture, intra-tree macro-texture and inter-tree texture respectively. A 50 x 50 pixels image is extracted for each class and a r² matrix is built in each window size for each band to detect a possible redundancy. If the r² coefficient between two attributes is upper than 80, couples are considered as a single variable (Table 2).

Unfortunately, in tropical areas, remotely sensed images suffer from cloudy conditions and optical spectrum response does not contain enough information for species discrimination.

- The NASA PACRIM II AirSAR mission of 2000 over-flied Marquesas archipelago, providing 5 m-resolution SAR data in 3 bands: TopSAR CVV (λ =5.7 cm), and PolSAR L (λ =23 cm) and P (λ =67 cm) bands in full polarimetry. This dataset allows extracting polarimetric indices to reach land cover structural properties. Relying on the work of [15], the ten most relevant polarimetric indices

Table 1: The 11 natural land cover composing the site study and some characteristics; T=Tree, S=Shrub , H=Herbaceous, Inv.=Invasive, PI=Polynesian introduction and Ind.=Indigenous.

Species	Stratum	Status
Acacia farnesiana (A)	S	Inv.
Casuarina equisetifolia subsp *equisetifolia* (Ca)	T	Ind.
Dicranopteris linearis (D)	H	Ind.
Falcataria moluccana (Fa)	T	Inv.
Ficus prolixa var *prolixa* (Fi)	T	Ind.
Hibiscus tiliaceus subsp *tiliaceus* (H)	T	Ind.
Inocarpus fagifer (I)	T	PI
Pandanus tectorius var *tectorius* (Pa)	T	Ind.
Psidium guajava (Ps)	S	Inv.
Sapindus saponaria (Sa)	T	Ind.
Schizostachyum glaucifolium (Sc)	T	PI

Table 2: r² matrix of GLCM attributes for 3x3, 9x9 and 15x15 pixels window size; each value represents a mean for the 3 bands and its standard deviation. Values in bold are considered as significantly correlated; ASM=angular second moment.

3x3 window	Variance	Contrast	Dissimilarity
Contrast	60 ± 2	-	-
Dissimilarity	58 ± 3	**93 ± 1**	-
ASM	5 ± 2	5 ± 2	9 ± 2
9x9 window	Variance	Contrast	Dissimilarity
Contrast	**85 ± 1**	-	-
Dissimilarity	**81 ± 2**	**96 ± 1**	-
ASM	18 ± 16	20 ± 16	31 ± 20
15x15 window	Variance	Contrast	Dissimilarity
Contrast	**88 ± 1**	-	-
Dissimilarity	**83 ± 1**	**96 ± 1**	-
ASM	16 ± 14	18 ± 16	28 ± 21

for Polynesian vegetation classification are extracted. Active radar backscatters are dependant of vegetation structure, humidity or incidence angle and add thus evident supplementary information. Digital speckle is filtered with a Frost filter (damping factor=1; window size=5x5 pixels) showing good results in preserving edge information [16]. Unlike the optical data, SAR data is insensitive to cloud cover but we can find relief shadows due to the airborne sensor flying over high volcanic Marquesas Islands.

- Oro-topography is a third information source concerning vegetation spatial distribution. Climatic factor such as moisture and temperature are typically variable in mountainous areas, affecting vegetation distribution by controlling key ecological processes [17]. We use four

Table 3: Overall accuracy and Kappa coefficient as a function of used nonlinear kernel.

Accuracy	RBF	Polynomial (d=3)	Sigmoid
OA (%)	70	66	22
Kappa	0.67	0.63	0.15

Table 4: Overall accuracy and Kappa coefficient as a function of the considered sources.

Accuracy	1m-IKONOS	DEM	AirSAR	5m-IKONOS +DEM+AirSAR
OA (%)	54	30	20	70
Kappa	0.50	0.25	0.13	0.67

indices well known to affect - directly or not - patterns of climate zonation: elevation (m.a.s.l.), slope steepness (°), eastness (dimensionless) i.e. exposition to the trade winds and compound topographic index (CTI, dimensionless) quantifying fluid drainage [18]-[19].

The chosen multi-source decision scheme is the most relevant one in [6] All SVM are trained on each individual data. Their outputs are then used for a SVM-based decision fusion to predict the final class membership of each sample.

4. FUSED-SVM RESULTS

As shown before by [19] in mono-source case, RBF and polynomial kernels produce similar results with a perceptible superiority for the RBF one (Table 3). Likewise, [20] denote that the RBF kernel has less numerical difficulties than others. Our results corroborate these observations in a multi-source case.

With an OA=70% (Table 4), fusion results are fairly good for such a complex problem, the site study landscape being a complex system of numerous intrusive plant communities. Multi-source fusion has two effects. The first one is a synergic effect between each complementary mono-source successful classifications whereas the second one is based on fruitless classification: SVM decision fusion is able to use mono-source classification confusion pattern as information. For example, the *Falcataria moluccana* class is strongly and partially confused with *Casuarina equisetifolia* subsp *equisetifolia* on the 1m-IKONOS and the AirSAR-based classification respectively and with *Dicranopteris linearis* on the DEM-based classification. With a producer accuracy of 91% and a user accuracy of 95% for the *Falcataria moluccana* class, fused-SVM uses this confusion pattern as information to class efficiently these species (figure 3).

Due to its spatial and spectral resolutions, AirSAR data is not adapted to the detailed species-based class set we used
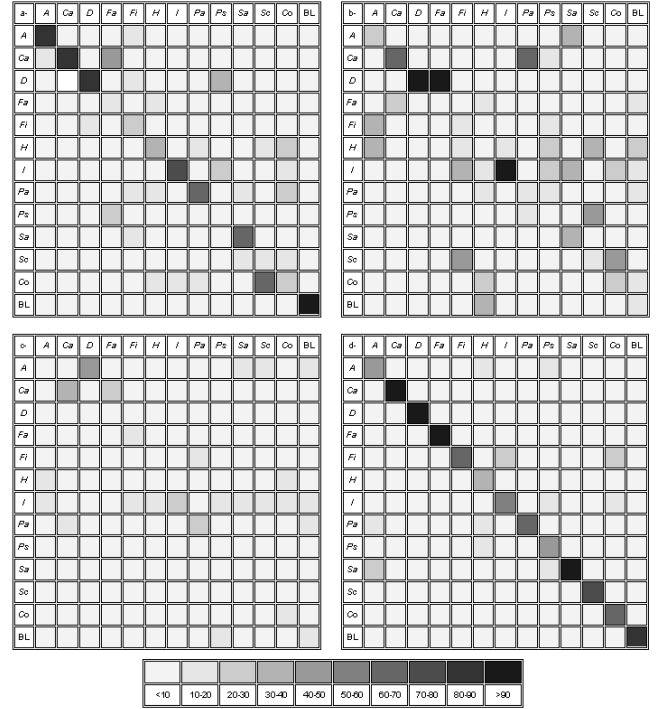
Figure 3: Confusion matrix of the RBF-SVM classification of the (a-) 1m-IKONOS, (b-) DEM, (c-) AirSAR and (d-) three classifications (5m-IKONOS+AirSAR+DEM) merged in an additive RBF-SVM classification. ROIs were merged in accordance with the species stratum.
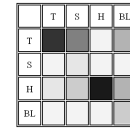

Figure 4: Confusion matrix of the RBF-SVM classification of the AirSAR data. ROIs were merged in accordance with the species stratum. Then, a further stratified random sampling was computed; T=Tree, S=Shrub , H=Herbaceous ; greyscale is the same than figure 3.

figure 3-c. Conversely, and by nature, it is adapted to structural class sets such as vegetation strata (figure 4).

Results on the DEM and summarized in figure 3 prove that some species are ecologically generalists whereas certain species are specialists. Four species have a clear oro-topographical determinism. *Casuarina equisetifolia* subsp *equisetifolia* and *Dicranopteris linearis*, distributed on rocky outcrops and ridges respectively i.e. areas with high elevation (500 ± 13 m in average and 470 ± 120 m respectively) and low CTI (2.0 ± 1.0 and 2.0 ± 1.2), *Inocarpus fagifer* living in riparian sites, where CTI is high (3.8 ± 2.4) while *Sapindus saponaria* is a typical component of semi-xerophilous forests with low CTI (2.4 ± 0.9) and located on the strongest slopes with a mean of 36 ± 8.4°.

5. MAPPING RESULTS

In the Marquesas archipelago, multi-source SVM fusion allows classifying fine scale class set as dominant species. Alien invasive species are dominant in 14% of the total study site (234 ha). As illustrated in figure 5, the invasion is generally spread but often concentrated near bare lands as areas disturbed by landslides or road construction. Alien invasive species seem to take advantage of human perturbations and landscape fragmentation, facilitating flux of their propagules. Some alien invasive species are elsewhere well known to modify ecological condition as aggravating the soil erosion hazard. Bare lands, as erosion prone areas, are already covering 24 ha i.e. 1.4% of the total study site!
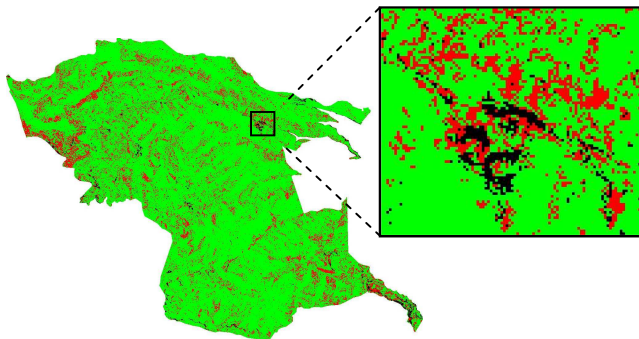
Figure 5: Map of alien invasive species; red=alien invasive are dominant, green=native dominant; black: bare lands and communication infrastructures.

AKNOWLEDGEMENTS

REFERENCES

[1] J. A. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," IEEE Trans. Geosci. Remote Sens., vol. 37, pp. 1367-1377, 1999

[2] J. H. Halldorsson, J. A. Benediktsson, and J. R. Sveinsson, "Support vector machines in multisource classification," *IEEE International Geosci. Remote Sens. Symposium*, IGARSS '03, vol. 3, pp. 2054-2056, 2003.

[3] D. B. Michelson, B. M. Liljeberg, and P. Pilesjo, "Comparison of algorithms for classifying Swedish landcover using Landsat TM and ERS-1 SAR data," *Remote Sens. Environ.*, vol. 71, no.1, pp. 1-15, 2000.

[4] G. Chust, D. Ducrot, and J. L. Pretus, "Land cover discrimination potential of radar multitemporal series and optical multispectral images in a Mediterranean cultural landscape," *Int. J. Remote Sens.*, vol.25, no.17, pp. 540-552, 1990.

[5] X. Blaes, L. Vanhalle, and P. Defourny, "Efficiency of crop identification based on optical and SAR image time series," *Remote Sens. Environ.*, vol.96, no. 3/4, pp. 352-365, 2005.

[6] B. Waske and J. A. Benediktsson, "Fusion of Support Vector Machines for Classification of Multisensor Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, pp. 3858-3866, 2007.

[7] J. Florence and D. Lorence, "Introduction to the flora and vegetation of the Marquesas Islands", *Allertonia*, vol. 7, pp. 226-237, 1997.

[8] V. Vapnik and A. Chervonenkis, *Statistical learning theory*. New York: Wiley, 1998.

[9] S. Fukuda and H. Hirosawa, "Support vector machine classification of land cover: application to polarimetric SAR data," *IEEE International Geosci. Remote Sens. Symposium*, vol. 1, pp.187-189 2001.

[10] G. M. Foody and A. Mathur, "A relative evaluation of multiclass image classification of support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, pp. 1335-1343, 2004.

[11] S. Jin, D. Li, and J. Wang, "A comparison of support vector machine with maximum likelihood classification algorithms on texture features," *IEEE International Geosci. Remote Sens. Symposium*, vol. 5, pp. 3717-3720, 2005.

[12] F. Gillet, *La phytosociologie synusiale intégrée. Guide méthodologique*, Documents of the Laboratoire d'Ecologie végétale, Institut de Botanique, Neuchâtel Univ., 2000.

[13] B. Waske, J. A. Benediktsson, and J. R. Sveinsson "Classifying remote sensing data with support vector machines and imbalanced training data," in *Multiple Classifiers Systems* Heidelberg: Springer Berlin, 2009.

[14] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification", *IEEE Trans. Systems Man Cybernet.*, vol. 3, pp. 610-621, 1973.

[15] C. Lardeux, "Apport des données radar polarimétriques pour la cartographie de la végétation naturelle," PhD thesis, Paris-Est Marne-la-Vallée Univ., France, 2008.

[16] M. R. de Leeuw, and L. M. Tavares de Carvalho, "Performance evaluation of several adaptive speckle filters for SAR imaging," *Anais XIV Simpósio Brasileiro de Sensoriamento Remoto*, pp. 7299-7305, 2009.

[17] A.D. Richardson, "Foliar chemistry of balsam fir and red spruce in relation to elevation and the canopy light gradient in the mountains of the northeastern United States," *Plant and Soil*, vol. 260, pp. 291-299, 2004.

[18] I. D. Moore, P. E. Gessler, G. A. Nielsen, and G. A. perterson, "Soil attribute prediction using terrain analysis," *Soil Sci. Soc. Am. J.*, vol. 57, pp. 443-452, 1993.

[19] P.E. Gessler, O.A. Chadwick, F. Chamran, L. Althouse, and K. Holmes, "Modeling soil-landscape and ecosystem properties using terrain attributes," *Soil Sci. Soc. Am. J.*, vol. 64,pp. 2046-2056, 2000.

[20] B. Schölkopf and A. Smola, Learning with kernels. Cambridge: MIT Press, 2002.

[21] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," Technical report, Department of Computer Science & Information Engineering, National Taiwan Univ., Taiwan, 2009.