MULTI-SOURCE SVM FUSION FOR ENVIRONMENTAL MONITORING IN THE MARQUESAS ARCHIPELAGO

Robin Pouteau, Benoît Stoll and Sébastien Chabrier

South Pacific Geosciences (GePaSud) Laboratory - University of French Polynesia (UPF)
BP 6570, 98702 FAA'A - TAHITI - French Polynesia. e-mail: {firstname.lastname}@upf.pf

1. INTRODUCTION

Several past works state that the combined use of multi-source remote sensing data improves accuracy for land cover classification, e.g. [1] [2], using different methods. Most of them are compared in [3], namely maximum likelihood, decision trees and support vector machines (SVM), showing that SVM give 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 threats to native ecosystems [4]. 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 [5] are chosen because they perform more accurately than other classifiers in mono-source cases, e.g. [6] [7] [8], and in multi-source cases [3]. They are frequently used with three kinds of nonlinear kernels: the polynomial, the Gaussian radial basis function and the sigmoid ones. Since these kernels have rarely been studied in a multi-source problem in remote sensing applications, we compare their output accuracy.

Remote sensing is a useful tool for ecosystems 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 (such as pixels one) to be optimally 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 stratum characteristics.

A first ground truth campaign is carried out to look for a representative class set. Vegetal community is commonly divided by phytosociologists into 3 strata: herbaceous (<1m), shrubs (1-5m) and trees (>5m). To characterize vegetation composing the study area, 143 inventories in the commonly used surfaces - 100m² for herbaceous synusiae and 450m² for shrub and tree ones - are sampled systematically in a mesh network from an initial random point. The distance between two sampling area is 300m. For each inventoried species in each stratum, an abundance index (from 0 to 5) is inputted. Then, we compute a simple process to select the dominant highest vegetation stratum only for each synusiae i.e. emergent species in the remotely sensed images.

In a second ground truth campaign, 36 training spots of 450m² (~1‰ of the site study), three per class, are selected and geolocalized with a GeoXH Trimble GPS. Such balanced datasets are used to avoid class over- or under-representation problems [9]. For validation, 36 others spots 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's scenes from 2005 informs us about vegetation texture and passive absorption spectra. The 1m-merged data set (3 bands multispectral) spectral resolution is λ=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 [10] for example. Four GLCM texture metrics - variance, contrast, dissimilarity and angular second moment - are computed by using three window sizes of 3x3, 9x9 and 15x15 pixels which visually correspond to intra-tree micro-texture, intra-tree macro-texture and inter-tree texture respectively.

- The NASA PACRIM II AirSAR mission of 2000 over-flied the Marquesas archipelago, providing 5m-resolution SAR data in 3 bands: TopSAR C_{VV} (λ =5.7 cm), and PolSAR L (λ =23 cm) and P (λ =67 cm) bands in full polarimetry. Relying on the work of [11], the ten most relevant polarimetric indices for Polynesian vegetation classification are extracted. Active radar backscatters are dependant of vegetation structure, humidity or incidence angle and add thus evident supplementary information. Unlike the optical data, SAR data is insensitive to cloud cover but we can get relief shadows due to the airborne sensor flying over the 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 patterns by controlling key ecological processes. We use four topographical indices well known to affect - directly or not - patterns of climate zonation: elevation (meters above sea level), slope steepness (degrees), eastness (dimensionless) i.e. exposition to the trade winds and a compound topographic index (CTI, dimensionless) quantifying fluid drainage [12].

The chosen multi-source decision scheme is the most relevant one in [3]. 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. RESULTS AND DISCUSSION

As shown by [13] in a mono-source case, RBF and polynomial kernels produce similar results with a significative superiority for the RBF one. Likewise, [14] denote that the RBF kernel has less numerical difficulties than others. Our results corroborate these observations in a multi-source case.

With an overall accuracy of 70%, 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 confusion patterns as information.

Due to its spatial and spectral resolutions, AirSAR data is not adapted to the detailed species-based class set we used. Conversely, and by nature, it is adapted to structural class sets such as vegetation strata.

Results on the DEM prove that some species have a higher ecological valence than other more specialist species. Four species have a clear spatial determinism. *Casuarina equisetifolia* subsp *equisetifolia* and *Dicranopteris linearis*, distributed on rocky outcrops and ridges respectively i.e. areas with high elevation and low CTI, *Inocarpus fagifer* living in riparian sites, where CTI is high while *Sapindus saponaria* is a typical component of semi-xerophilous forests with low CTI and localized on the strongest slopes.

In the Marquesas archipelago, multi-source SVM fusion allows classifying fine-scale class sets as dominant species. Alien invasive species are dominant in 14% of the total study site (234 hectares). The invasion is generally spread but often concentrated near disturbed areas. Alien invasive species seem to take advantage of anthropogenic perturbations and landscape fragmentation, facilitating flux of propagules and encouraging the invasion process. Some alien invasive species are elsewhere well known to modify ecological condition as aggravating soil erosion hazard. Bare lands, as erosion prone areas, are already covering 24 hectares i.e. 1.4% of the total study site!

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] 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.

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

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

[6] S. Fukuda and H. Hirosawa, “Support vector machine classification of land cover: application to polarimetric SAR data,” *IEEE International Geosci. Remote Sens. Symposium, IGARSS’01* vol. 1, pp.187-189 2001.

[7] 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.

[8] 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, IGARSS '05,* vol. 5, pp. 3717-3720, 2005.

[9] 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.

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

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

[12] 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.

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

[14] 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.