Lobo, A., P. Legendre, J. L. González Rebollar, J. Carreras & J.-M. Ninot. 2002. Land cover classification at a regional scale in Iberia: separability in a multi-temporal and multi-spectral data set of satellite images. Pp. 348–353 *in*: Proceedings of the Symposium "Recent Advances in Quantitative Remote Sensing", Auditori of Torrent, Spain, 16-20 September 2002.

Land Cover Classification at a Regional scale in Iberia: separability in a multi-temporal and multi-spectral data set of satellite images

Agustín Lobo[1], Pierre Legendre[2], José Luis González Rebollar[3], Jordi Carreras[4] and Josep-Maria Ninot[4]

[1]*Institut de Ciències de la Terra "Jaume Almera" (CSIC), Lluis Solé Sabarís s/n, 08028 Barcelona, Spain* (alobo@ija.csic.es)

[2]*Département des Sciences Biologiques, Université de Montréal, C.P. 6128 succ. A, Montréal, Québec, Canada H3C 3J7*

[3]*Estación Experimental del Zaidín (CSIC), Prof. Albareda 1, 18008 Granada, Spain*

[4]*Departament de Biologia Vegetal, Facultat de Biologia, Universitat de Barcelona, Diagonal, 645, 08028 Barcelona, Spain*

ABSTRACT. *Earth Observation at regional scales, such are the Iberian Peninsula or the Mediterranean Basin, is an important tool to understand the relationships between climate and surface properties. Among the different layers of information that can be derived from satellite imagery, Land Cover maps are important by themselves and as an aid to infer other variables. Land Cover legends at regional scales require finer categories than those used at a global scale, which implies processing multi-spectral imagery acquired by Earth Observing systems with daily acquisition rates. In this article we discuss several alternatives to analyze satellite image data sets that are both multi-temporal and multi-spectral, with spatial resolution of 1 km^2. In order to facilitate the interpretation of our results, we restrict our analysis to pixels that correspond to cells with a uniform and known cover on the ground, as described by a detailed vegetation map, in Catalonia (NE Spain). Our results indicate that canonical Redundancy Analysis is efficient at reducing the multi-spectral and multi-temporal space while keeping high statistical separability among habitat types. The small fraction of uniform pixels (~2 %) suggests that, at least for the Mediterranean Region, data fusion techniques would be convenient to increase spatial resolution in the data set, and that instruments keeping daily acquisition rates but with higher spatial resolution (~1 ha) should be considered.*

1 INTRODUCTION

Earth observation through remotely sensed imagery has an important role in Global Change research. Satellite images are used to assess the state of the surface, to parameterize some models and to validate results. Considering that these images are the only observations done at a scale close to the one intended for models, their analysis should suggest new approaches for modeling.

Among the applications of Remote Sensing on this field, Land Cover (LC) mapping holds a central place. Global Land Cover maps are critical information for monitoring changes of the Earth surface and also have an auxiliary role to estimate other surface variables. Historically, Land Cover mapping has used either multi-temporal or multi-spectral imagery depending on whether the study had global or continental extent with coarse resolution (typically using AVHRR data sets), or covered smaller areas with higher resolution (typically using LANDSAT and SPOT images).

Most work done with AVHRR data sets for global and continental Land Cover mapping reduce each multi-spectral image to one single layer of Normalized Difference Vegetation Index (NDVI), which is proportional, in a statistical sense, to the fraction of photosynthetically-active radiation that is intercepted by green tissue (fPAR). The multi-temporal and multi-spectral data set is thus simplified into a time sequence of NDVI layers. An early and important finding of Remote Sensing is that time series of NDVI are very good descriptors of vegetative phenology. Global and continental charts produced from AVHRR data sets are essentially based on the information provided by time series of NDVI, although different authors used different techniques for classification (Tucker et al. 1985, Townshend et al 1987, Lloyd 1990, Loveland 1991, Eidenshink 1992, Running et al. 1994, 1995, DeFries et al 1995, Ehrlich and Lambin, 1996, Loveland et al. 2000). A similar approach was used at a regional scale by Lloyd (1989) and Lobo et al. (1997).

While most work on Global Change has been conducted at a global scale, interest on modeling and assessing the impact at a regional scale is growing. A regional scale facilitates evaluation of results, and its detail is more appropriate to study the implications of Global Change for human populations. Current legend schemes of Land Cover classes are relevant information for studies at global and continental scales, but regional applications require more detailed classifications, which in turn require more spectral information. Newer recent satellite Earth observation systems with daily acquisition rates are equipped with more spectral bands than earlier NOAA systems.

A methodological problem arises when we decide to analyze time series of multi-spectral images and want to use more than one single index across time, as data from each cell becomes a multi-variate time series. In an analogous way as done with the older AVHRR imagery, it is, in principle, possible to identify specific features in time profiles of several indices, but little is known on the phenological behavior of indices other than NDVI and alike. Another approach is to stack spectral bands from images of successive dates, as if they were bands from other regions in the electromagnetic spectrum, to create a huge multi-spectral image). This approach simplifies the problem by ignoring its temporal aspect and subsuming it into its multi-spectral aspect. Such an approach would, however, severely affect the essence of the problem, since the time axis (the arrow of time) is of a different nature than the axes of the variables.

In this study, we describe the use of Canonical Redundancy Analysis (RDA) to transform the original space of *n (cells) x t (times) x p (spectral variables)* to a reduced space of *n (cells) x p (scores)* which is subsequently submitted to discriminant analysis. As a first step and in order to facilitate the interpretation of our results, we restrict our analysis to pixels that correspond to cells with a uniform and known cover on the ground, using an annual set of SPOT-4 VEGETATION images and land-cover maps of Catalonia (NE Spain).

2 METHODS

2.1 Data

We have processed an annual (*1999*) set of 36 S10-VEGETATION images and 44 digital maps at the scale 1:50,000 from the series of Maps of Habitats of Europe for Catalonia (NE Spain). Each map covering an area of 28 km x 18.5 km, we have screened a total of 22,792 km^2. S10-VEGETATION images are 10-day syntheses of daily calibrated and atmospherically corrected images produced through the Maximum Value Compositing method of Holben (1986). SPOT-4 VEGETATION has four spectral bands: 430 – 470, 610 – 680, 780 – 890 and 1580 – 1750 nm. The legend of the maps is based on the Corine Biotopes Manual (Devillers *et al.*, 1991) and the Directive 92/43 of the European Union with specific improvements for Catalonia (NE Spain) (Carreras & Vigo 1997). This system is very close to the more recent and comprehensive European Nature Information System (EUNIS) of the European Environment Agency, developed and maintained by the European Topic Center on Nature Protection and Biodiversity of the European Union. Rather than to *habitats* in the ecological sense sense, the legend corresponds to that of a land-cover map with emphasis on vegetation categories.

2.2 Analysis

We selected those 1-km^2 cells that had at least 90% of their surface in a single habitat patch (a total of 495 cells, 2.17% of all screened cells), and extracted the reflectance data and ancillary information from the sequence of S10-VEGETATION pixels that matched those cells. We organized the reflectance values as one data matrix for each selected pixel. The data matrices had time observations as rows (one observation for each S10 composite), while, as columns, we included the four spectral bands and two normalized difference (ND) indices: NDVI (which we call here ND(nir,red) for consistency) and ND(nir,mir).

We ran three approaches of analysis, aiming to discriminate the selected cells according to their habitat type. In all three approaches, we calculated the statistical separability of the different habitat categories in the transformed spaces using the Jeffries-Matutsita distance (Richards 1999). The use of the Jeffries-Matutsita distance as a measure of statistical separability in all cases let us compare the discriminant power of each approach.

The first approach used only the (univariate) time series of ND(nir,red) for each selected cell. Therefore, this approach focused on the temporal aspect of the data set. We selected the column of NDVI profiles from the data matrix of each selected cell, assembled a global data matrix with these profiles as row vectors, ran a PCA on the global data set, and selected the first four scores (accounting for > 96% of the total variance).

The second approach used the spectral space defined by the bands of the imagery for each period of synthesis, hence focusing on the spectral characteristics of the data set. We made a multi-variate table with the selected cells and their reflectance values in the four bands of the VEGETATION image for each period of synthesis, calculated the statistical separability for each table, and, finally, calculated a combined statistical separability matrix by selecting the highest separability value across time for each pair of habitats.

The third approach used the (multi-variate) time course of the spectral responses for each cell, attempting to combine both sources of information: temporal and spectral. We ran a RDA (Legendre and Legendre, 1998) on the multivariate table of all selected cells using bands B2, B3 and MIR plus the two ND indices as matrix of response variables, and a matrix of dummy variables coding for the cells as matrix of explanatory variables.

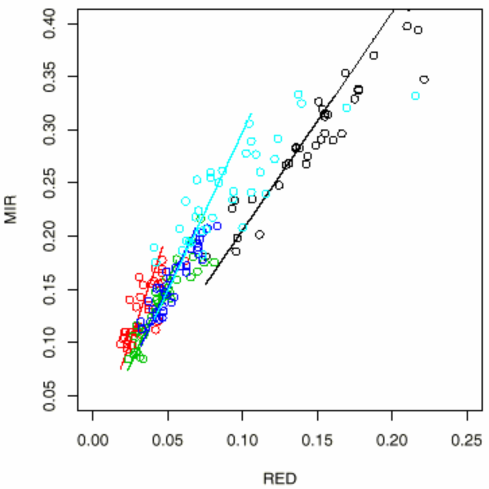

Figure 1. Examples of four robust fits of a regression line to the mir vs. red reflectance values of four selected cells, constraining the intercept to a 0 value. Black symbols, a non-irrigated cereal crop (code 30); red, evergreen oak forest (code 25) Catalonia (NE Spain).

2.3 Slope of the MIR *vs*. RED regression.

The relationship between values of red and medium-infrared reflectance is dependent upon both the structure and the humidity of the target. Most plots of the values of MIR *vs*. RED bands for a given target along an annual cycle are a cloud in which a linear trend accounts for most of the variance (Fig. 1). Some outliers located well below the linear trend are likely due to recent rainfall events. We used a robust linear regression method to fit a line constraining the intercept to be 0 for each selected cell, and test the use of the slope as a characteristc feature of different habitats.

3 RESULTS

A plot of the selected cells in the plane of the first two principal components (PCs) of the ND(nir,red) time series (Fig. 2) indicates that the main source of variation is along an axis from the evergreen oak community (code 25, relatively flat plots of ND(nir,red), see Fig. 3) to the non-irrigated cereal fields (code 30, wave of ND(nir,red) peaking in early spring). Therefore, the main source of variation is due to the "degree of deciduoucity" of the plants in the cell. A second axis of variation ordinates the ND(nir,red) time profiles from those with a narrow wave picking in spring to those with a narrow wave picking in summer (code 31, rice fields), with the flat profiles of evergreen communities and the wide waves of the *Genista purgans* shrubland (a pattern of discontinuous evergreen shrubs over an herbaceous background) at intermediate positions. The second main source of variation is thus related to the timing of the greenness peak.

Values of the Jeffries-Matusita index indicate that several habitat types cannot be adequately discriminated using the PCs of the ND(nir,red) time profiles. Particularly low values are found among the different Mediterranean shrub communities (codes 5, 6 and 7) and between those and *Pinus halepensis* woodlands (code 12). In some cases the low separability is a consequence of the definition of the habitat types. This is the case for *Rosmarinus* garrigue and *Pinus halepensis* woodlands over *Rosmarinus* garrigue (codes 12 and 13) and between the two types of *P. halepensis* woodlands (codes 12 and 13). As these woodlands can be very open, a substantial part of the reflectance actually comes from the understory.

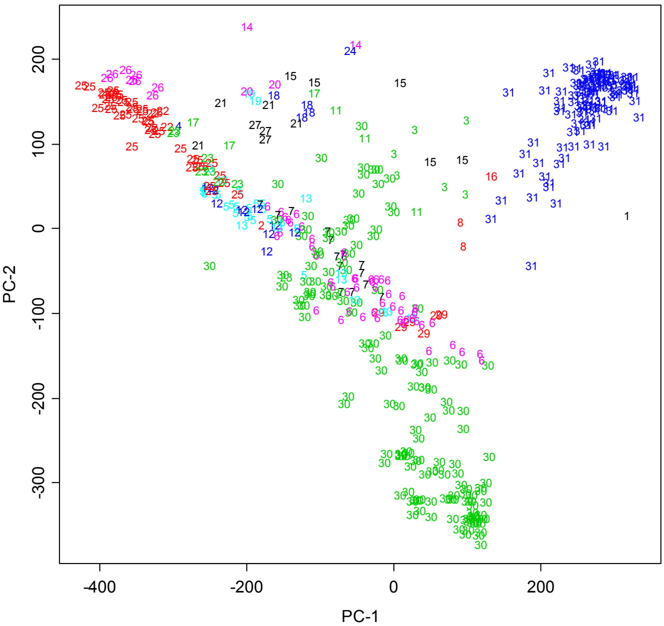

Figure 2. Projection of the selected cells (identified by their habitat codes) on the plane of the first two PCs. See Fig. 3 for time profiles of representative cells with habitats coded as 25 (forest of evergreen oak), 30 (not-irrigated creal fields), 3 (montane fields of *Genista purgans*) and 31 (rice fields). Other relevant codes: 5 (bushland of *Cistus sp*. on siliceous soils), 6 (garrigues of *Rosmarinus officinalis*), 7 (garrigues of *Cistus clusii* and *Anthyllis cytisoides* on calcareous soils), 14 (forests of *Abies alba*), 15 (forests of *Pinus uncinata*).

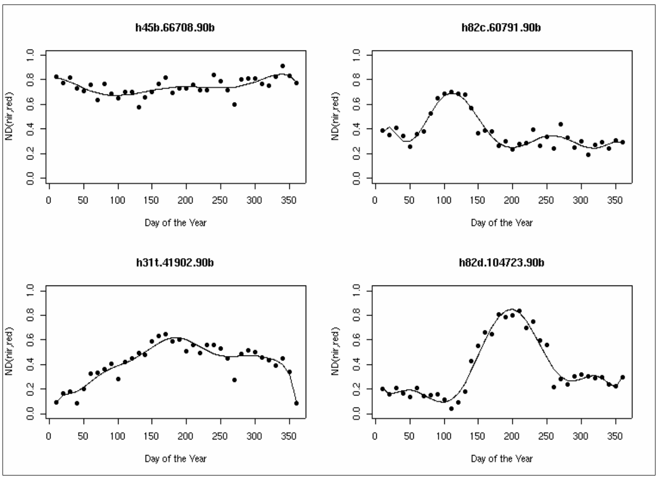

Figure 3. Time profiles of ND(nir,red) for four selected pixels with codes 25, 30, 3 and 31 (forest of evergreen oak, non-irrigated cereal fields, sub-alpine shrubland and rice fields). Note the position of these habitats in Fig. 2. *(25 -> h45b; 30->h82c; 3->h31t; 31->h82d)*

The RDA ordination is essentially the same as the PCA ordination. Clustering within class 30 is more evident and, in general, cells tend to cluster by habitat type more than in the PCA ordination. As a consequence, discrimination among habitat types is enhanced in the RDA ordination . Separability reaches high values for classes 5 *vs*. 6 (1.96) and 5 *vs*. 7 (1.99), while classes 6 *vs*. 7 remain undiscriminated. Statistical separability of the Mediterranean shrublands (codes 5, 6 and 7) *vs*. the *P. halepensis* woodlands (codes 12 and 13) also achieves high values in this data set. Discrimination between the two evergreen oak forests (codes 25 and 26) is, however, even more difficult here, although the values of separability in the PCA-transformed data set were already too low. Despite a small increase, discrimination between *P. halepensis* (codes 12 and 13) and *P. uncinata* forests (code 23) is weak. An important increase of separability occurs between the Mediterranean shrublands (codes 5, 6 and 7) and the *P. uncinata* forests (code 23).

MIR *vs*. RED slopes calculated by robust linear regression with the constrain of null intercept show an interesting pattern (Fig. 4). Lowest values are found for rice fields, increase for cereal fields and attain a maximum for evergreen forests. Values for garrigue habitats are intermediate. Several other facts deserve being highlighted. First, while MIR *vs*. RED slopes do not discriminate between the two evergreen oak forests (*Quercus suber* and *Q. ilex*), they show a clear bimodal distribution for the case of *Q. suber* forests, with two distinct clusters. Cereal fields show also two clusters. Second, values from cells of *Cistus* garrigue are higher than those from cells of the other two garrigues. Third, conifer forests show intermediate values between the M-R slopes of evergreen forests and garrigues. Including M-R slopes as an additional variable for RDA increases separability, particularly between the *Rosmarinus* and *Cistus* garrigues. There is an increase between the two evergreen oak forests (*Quercus suber* and *Q. ilex*) as well, but this is a consequence of the presence of lower cluster of *Q. suber*. However, in both cases the increase is not sufficient to reach a good discrimination in these two pairs of habitats.

Despite the enhanced discrimination power, some habitat types still show poor separability. Two of the three Mediterranean shrublands cannot be discriminated between them, neither can be the two types of woodlands of *P. halepensis*. Furthermore, there is a weak separability between one of these two woodlands and one type of *P. uncinata* forest.

Combining the best spectral separability of each period of time does not improve discrimination within any of the pair of habitats with weak separability in the RDA-transformed data set. Searching the best discriminant period for each pair of habitat types also has the problem of selecting an eventually high separability merely due to transient conditions (i.e., a recent rainfall) rather than to more stable characteristics. This risk is much reduced if the entire annual cycle is considered, as in the RDA approach.

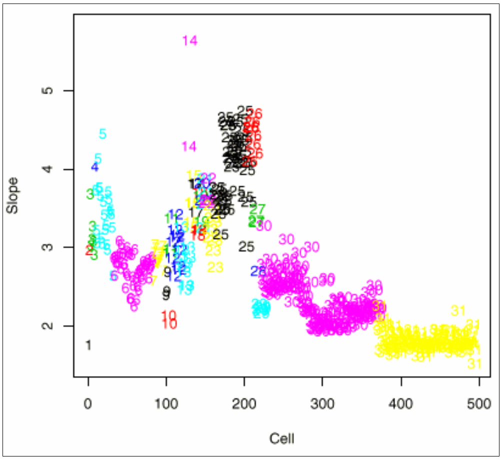

Figure 4. Values of the slope of the MIR vs. RED regression line with 0 intercept. Cells coded, and ordered, by habitat.

4 CONCLUSIONS

Direct RDA reduces a very complex multi-spectral and multi-temporal space and still keeps a high statistical separability among habitat types. Results using the RDA-transformed data set are better than those using NDVI time series and than those combining the best discriminant dates using all bands.

Some habitat types cannot be sorted out in the RDA-transformed data set. In some cases, the problem can be solved by going up one step in the hierarchical structure of the legend, and merging the indistinguishable types into a broader category, while still keeping an acceptable thematic detail. This is the case, for example, of the Mediterranean shrublands "Rosmarinus garrigue" and "Cistus and Anthyllis cytisoides garrigue" that can be lumped together into the category "Western garrigue" of the EUNIS classification scheme. However, in other cases, the indistinguishable types are not related in the hierarchical legend, and merging them together would create incoherent or too broader categories (i.e., "P. halepensis woodlands" and some "P. uncinata forests"). More research is required to discriminate these habitat types, with particular attention to including angular information.

The fraction of 1 km^2 pixels that are completely included within one single habitat patch is very low in our region of study (2.17%). Data fusion of imagery acquired by complementary systems might result into products with the required intensity in both time and space. Also, Remote Sensing systems designed for the Mediterranean region and featuring spatial resolution close to 1 ha and daily acquisition should be contemplated.

ACKNOWLEDGEMENTS

This research has been conducted within the framework of project AMFIBER (REN2001-1841/GLO, *Programa Nacional de Ciencia y Tecnología* of Spain) and the Programa Nacional Ramón y Cajal (*Ministerio de Ciencia y Tecnología* of Spain), using imagery provided by the VEGETATION Preparatory Program.

REFERENCES

Bartalev, S., Achard F., Erchov D., Gond. V., 2000. The potential contribution of SPOT 4/VEGETATION data for mapping Siberian forest cover at continental scale. Proceedings of the VEGETATION 2000 conference, 3-6 April 2000 Belgirate (Italia). http://vegetation.cnes.fr:8080/

Braun-Blanquet, J., 1951, Pflanzensoziologie. 2ond. Ed. (Springer-Verlag. Viena).

Carreras, J. and Vigo, J. 1997. Projecte de Cartografia dels Hàbitats a Catalunya. Report of the "Grup de Geobotànica i Cartografia de la Vegetació (Universitat de Barcelona)" for the "Departament de Medi Ambient de la Generalitat de Catalunya". Unpublished.

Casanova, D. 1998. Quantifying the effects of land conditions on rice growth: a case study in the Ebro delta (Spain) using remote sensing. Thesis Landbouwuniversiteit Wageningen, 219 pp.

DeFries R.S. and Townshend, J.R.G. 1995, Global discrimination of land cover types from metrics derived from AVHRR Pathfinder data. . *Remote Sensing of Environment* **54,** 209-222.

Devillers, P., Devillers-Terschuren, J. & Ledant, J.P. 1991. Corine Biotopes Manual. Habitats of the European Community. Commission of the European Communities. Luxenbourg.

Eidenshink, J.C.,1992. The 1990 conterminous US AVHRR data set. *Photogrammetric Engineering and Remote Sensig*, **58**, 809-813.

Ehrlich, D. and Lambin, E.F., 1996. Broad scale land-cover classification and international climatic variability. . *International Journal of Remote of Remote Sensing*, **17**, 845-862

Gao, B.-C., 1996, NDWI - a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment,* **58**, 257-266.

Holben, B.N., 1986. Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote of Remote Sensing*, **7**, 1417-1434.

Legendre, P. and Legendre, L. 1998. Numerical Ecology. 2nd edition (Amsterdam, Elsevier).

Lloyd, D. 1989. A phenological description of Iberian vegetation using short wave vegetation index imagery. *International Journal of Remote of Remote Sensing* **10**, 827-833

Lobo, A., Ibáñez-Martí, J.J. and Carrera Giménez-Cassina, C. 1997. Regional scale hierarchical classification of temporal series of AVHRR vegetation index. *International Journal of Remote Sensin*g, **18**, 3167-3193.

Loveland, T.R. and Belward, A.S. 1997. The IGBP-DIS Global 1 km Land Cover Data Set, DISCover First Results. *International Journal of Remote Sensin*g, **18**, pp. 3289-3295

Loveland, T., B. Reed, J. Brown, D. Ohlen, Z. Zhu, L. Yang and J. Merchant. 2000. Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1-km AVHRR Data. *International Journal of Remote Sensing*, **21**, 1303-1330.

Loveland, T.R., Merchant, J.M., Ohlen, D.O. and Brown, J.F. 1991. Development of a land cover characteristics for te conterminous US. *Photogrammetric Engineering and Remote Sensig*, **57**, 1453-1453.

Richards, J. and Jia, X., 1999. Remote Sensing Digital Image Analysis. An Introduction. (Berlin, Springer).

Running, S.W., Loveland, T.R., and Pierce L.L., 1994. A vegetation classification logic based in global biogeochemical models. *AMBIO*, **23**, 77-81

Running, S.W., Loveland, T.R., Pierce L.L., Nemani, R.R. and Hunt, E.R., 1995. A vegetation classification logic for global land cover analysis. *Remote Sensing of Environment* **51**, 39-48.

Spot Image 1998. *VEGETATION User´s Guid*e. http://www.spotimage.fr/data/images/vege/VEGETAT/book_1/e_frame.htm

Townshend, J.R.G., Justice, C. and Kalb, V., 1987, Characterization and classification of Sout American land cover types using satellite data. . *International Journal of Remote Sensin*g, **8**, 1189-1207

Tucker, C.J., Townshend, J., and Goff, T.E., 1985. African land-cover classification using satellite data. *Science*, **227**, 369-375.