Taylor & Francis
Taylor & Francis Group

Land cover classification at a regional scale in Iberia: separability in a multi-temporal and multi-spectral data set of satellite images

A. LOBO

Institut de Ciències de la Terra 'Jaume Almera' (CSIC), Lluis Solé Sabarís s/n, 08028 Barcelona, Spain; e-mail: alobo@ija.csic.es

P. LEGENDRE

Département de Sciences Biologiques, Université de Montréal, C.P. 6128, succ. Centre Ville, Montréal, Québec, Canada H3C 3J7

J. L. G. REBOLLAR

Estación Experimental del Zaidín (CSIC), Prof. Albareda 1, 18008 Granada, Spain

J. CARRERAS and J.-M. NINOT

Departament de Biologia Vegetal, Facultat de Biologia, Universitat de Barcelona, Diagonal, 645, 08028 Barcelona, Spain

Abstract. Earth observation at regional scales, such as of the Iberian Peninsula or 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 analyse satellite image datasets 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 ($\sim 2\%$) suggests that, at least for the Mediterranean Region, data fusion techniques would be convenient to increase spatial resolution in the dataset, and that instruments keeping daily acquisition rates but with higher spatial resolution ($\sim 1\,ha$) 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, parameterize some models and validate results. Considering that these images are

the only observations done at a scale close to the one used in models of global change, their analysis should guide new approaches for modelling.

Among the applications of remote sensing in this field, land cover (LC) mapping holds a central place. Global land cover maps present critical information for monitoring changes of the Earth surface. They 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 datasets), or covered smaller areas with higher resolution (typically using Landsat and SPOT images).

Most work done with AVHRR datasets for global and continental land cover mapping reduces 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) (see Myneni *et al.* 1995 for a summary). The multi-temporal and multi-spectral dataset 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 datasets 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 *et al.* 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).

Interest in modelling and assessing the impact of global change at a regional scale is growing. The regional scale facilitates evaluation of results, and the 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 than just NDVI.

Recent satellite Earth observation systems with daily acquisition rates have responded to such requirements and are equipped with more spectral bands than earlier NOAA systems. But while the analysis of time series of NDVI images deals with one single univariate time series for each cell, a problem with the analysis of time series of multi-spectral images is that data from each cell is a multi-variate time series on its own. One approach is to stack spectral bands from images of successive dates, as if they were bands from other regions in the electromagnetic spectrum, in order to create a huge multi-spectral image, in which the problem is simplified 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 paper, we use canonical redundancy analysis (RDA) to reduce the complex data space of time series of multi-spectral SPOT4-VEGEATION images and evaluate its discriminant power in terms of vegetation categories that have been previously and independently set by interactive photo-interpretation and field inspection.

2. Methods

2.1. *Data*

We have processed an annual (1999) time series of 36 S10-VEGETATION images. SPOT4-VEGETATION has four spectral bands: 430–470, 610–680, 780–890 and 1580–1750 nm with spatial resolution of 1 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) (see Maisongrande *et al.* in this issue for further details on SPOT4-VEGETATION products). Particularly important for the analysis of time series of satellite images is the fact that geometric features have been very much improved in VEGETATION products compared to earlier NOAA-AVHRR products. Multi-spectral and multi-temporal registration errors are close to, respectively, 100 m and 300 m.

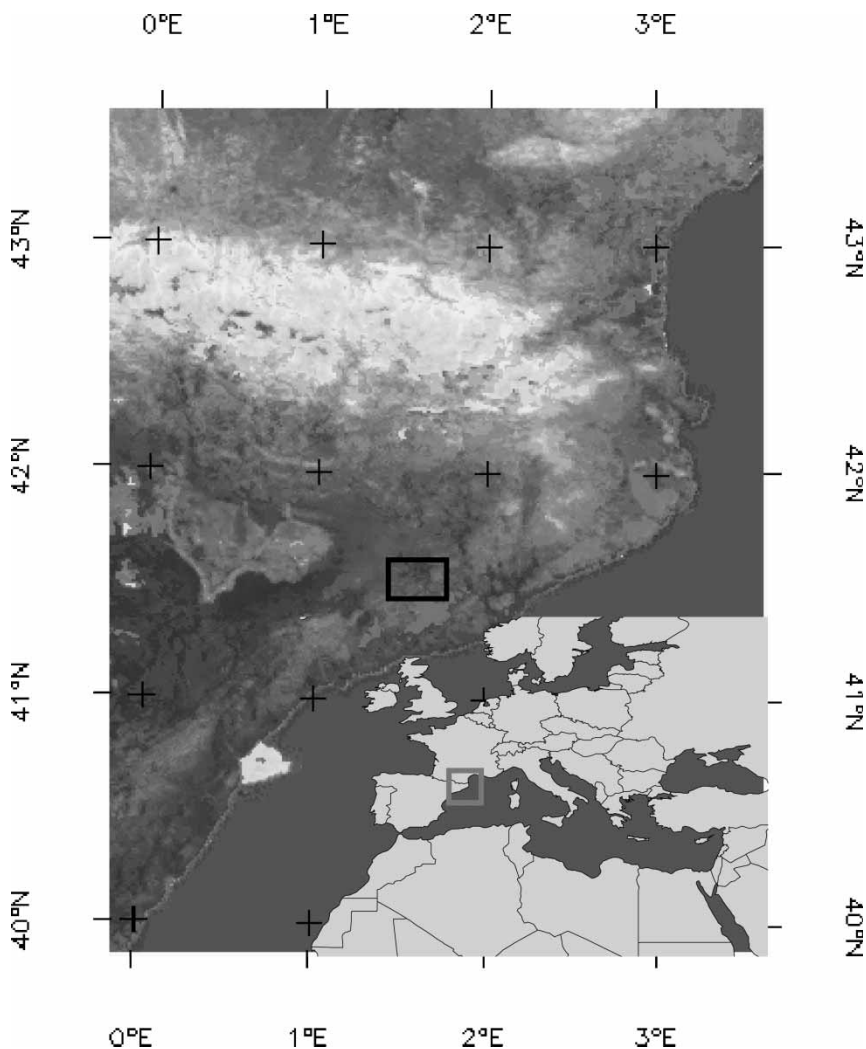We used 44 digital maps at the scale 1:50 000 (28 km × 18.5 km) from the series

Figure 1.   Study area. Colour composite of three S10 VEGETATION images of NDVI (red, early spring; green, mid summer; blue, early autumn). The rectangle represents one of the 1:50 000 vegetation maps used in this study.

of Maps of Habitats of Catalonia (north-east Spain) (figure 1) as explanatory variables. 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 (north-east Spain) (Carreras and Vigo 1997). The term habitat here must be understood as a land-cover unit with emphasis on vegetation categories.

2.2. *Analysis*

We selected 495 1-km^2 pixels that corresponded to a homogeneous cover in terms of the maps of habitats, extracted the reflectance data from the time sequence of S10-VEGETATION, and organized the reflectance values as one data matrix for each selected pixel. The data matrices recorded, as columns, the four spectral bands plus NDVI and a normalized difference (ND) index between reflectance values in the nir and mir bands (ND(nir,mir)). Each row in the data matrices is an observation extracted from an S10-VEGETATION image (figure 2) representing a 10-day interval.

We compared two approaches of analysis, both aiming to discriminate the selected pixels according to their habitat types. In both approaches, we calculated the statistical separability of the different habitat categories in the transformed spaces using the Jeffries-Matutsita distance (Richards and Jia 1999).

Selected values for pixel "cell_1" in the imagery:

Cell_1 t_1 b1 b2 b3 b4 nd1 nd2
Cell_1 t_2 b1 b2 b3 b4 nd1 nd2
...
Cell_1 t36 b1 b2 b3 b4 nd1 nd2

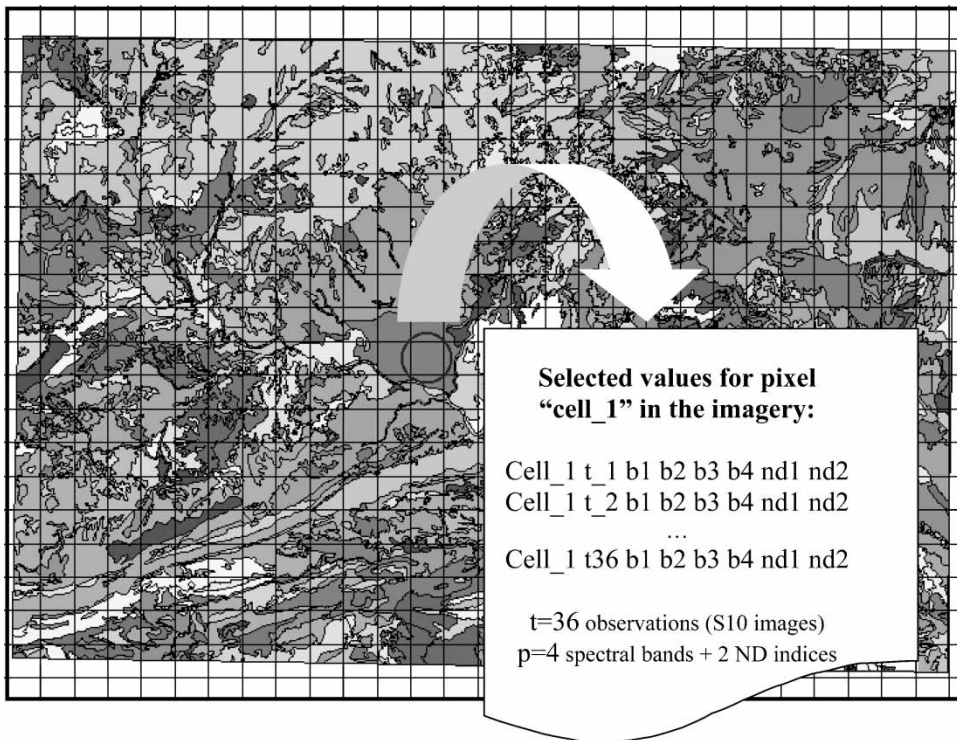t=36 observations (S10 images)
p=4 spectral bands + 2 ND indices

Figure 2. One of the 1 : 50 000 vegetation maps used in this study. The overlaid grid represents the grid of 1-km^2 S10-VEGETATION pixels. The encircled cell represents one of the selected cells, for which the annual imagery data have been extracted and saved to a data matrix.

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

The second 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 canonical redundancy analysis (RDA, Rao 1964; this method is described in Legendre and Legendre 1998) on the multivariate table of all selected cells and all 36 times, using bands B2, B3 and MIR plus the two ND indices as the columns of the matrix of response variables, and dummy variables coding for the cells as the matrix of explanatory variables.

3. Results

The information contained in the time series of NDVI is sufficient to provide a certain discrimination among habitat types. Most observations tend to cluster according to their habitat type in the discriminant plot (figure 3). This is particularly the case of rice fields (code 31), non-irrigated cereal fields (code 30), evergreen oak forests (codes 25 and 26), and *Pinus uncinata* forests with *Rhododendrum ferrugineum* (code 15). However, a significant number of cells are not clustered, and this dispersed distribution remains in higher discriminant axes. Values of the Jeffries-Matusita index (table 1) indicate that several habitat types cannot be discriminated using the PCs of the ND(nir,red) time profiles. Low values of the index are found among the different Mediterranean shrub communities (codes 5, 6 and 7) and between those and *Pinus halepensis* woodlands (code 12).

Discrimination among habitat types is enhanced in the RDA ordination (table 2). Among the three types of Mediterranean shrublands present in the data, 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. An important increase of separability occurs between the Mediterranean shrubland (codes 5, 6 and 7) and the *P. uncinata* forests (code 23). Curiously, discrimination between the two evergreen oak forests (codes 25 and 26) is weaker in this dataset, although the value of separability in the PCA-transformed dataset was already too low. Despite a small increase, discrimination between *P. halepensis* (codes 12 and 13) and *P. uncinata* forests (code 23) is weak. Two of the three Mediterranean shrublands cannot be discriminated, and neither can the two types of woodlands of *P. halepensis*. Furthermore, there is weak separability between one of these two woodlands and one type of *P. uncinata* forest.

4. Discussion and conclusions

Results using the RDA-transformed dataset are clearly better than those using NDVI time series, but some habitat types still cannot be separated. In some cases, the weak discrimination is a consequence of an inherent ambiguity in the definition of some of the habitat types, which prevents a neat distinction even at a conceptual level. This is, in turn, a consequence of the fact that in many cases there are no clear discontinuities on the ground. But in several other cases our methods and/or data simply fail to discriminate vegetation types. A number of complementary methods can be explored in searching for improvements. Among them, the inclusion of
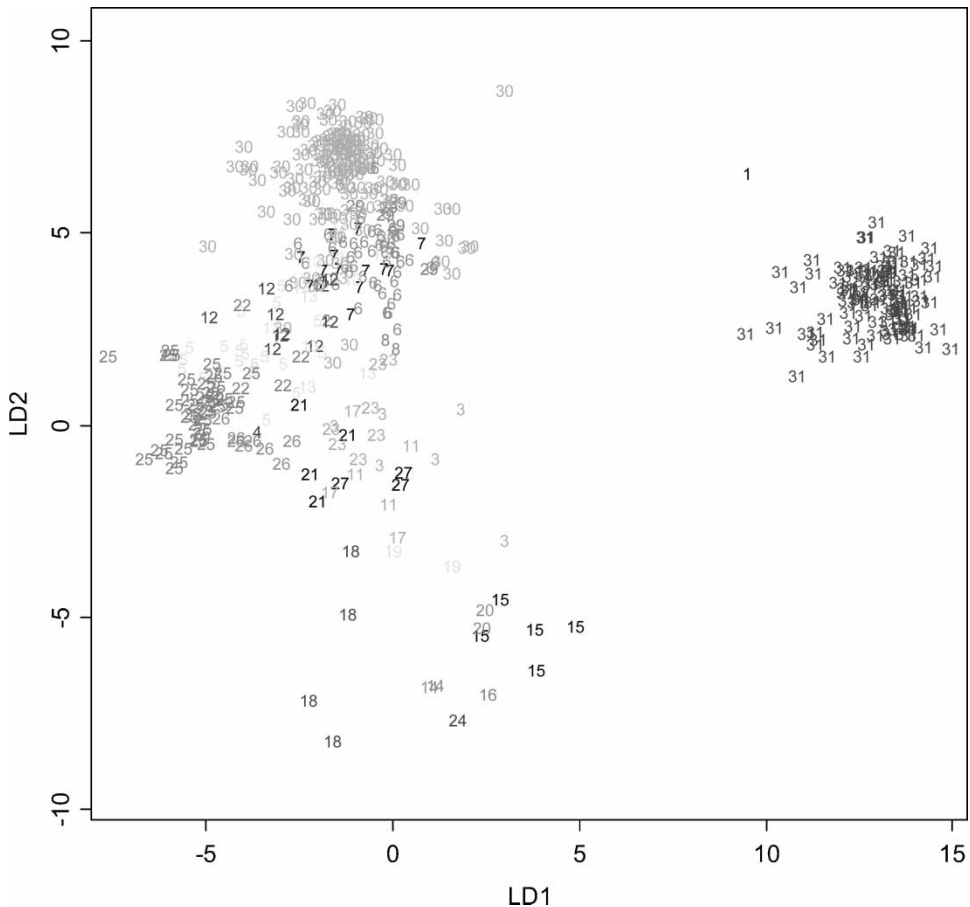
Figure 3. Projection of the selected cells (identified by their habitat codes) on the plane of the first two linear discriminat axes calculated on the NDVI dataset. Some relevant codes: 25 (forest of evergreen oak), 30 (non-irrigated cereal fields), 3 (montane fields of *Genista purgans*), 31 (rice fields), 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*). LD1, LD2—arbitrary units.

angular information in our analysis deserves particular attention, since adding information of a different nature often results in important enhancements in classification. Angular information has been shown to be very useful at discriminating forest types (Bartalev *et al*. 2000).

Despite the fact that, most likely, we will not be able to discriminate the many nuances of such a complex map legend as the one used here, our results indicate that we can reasonably expect to produce, at a regional level, a much finer legend than the one currently available in global products. However, an important caveat is the fact that the fraction of $1\,km^2$ pixels that are completely included within one single habitat patch is very low in our study area (2.17%). An obvious next step is the analysis of mixed (in terms of the habitats legend) pixels to determine if mixtures keep sufficient coherency to allow recovery of an appropriate legend for $1\,km^2$ pixels. If this is not the case, remote sensing systems designed for the Mediterranean region and featuring spatial resolution close to 1 ha and daily

Table 1. Statistical separability (Jeffries-Matusita index) using the PCA ordination of NDVI profiles for those habitat types with a number of selected cells sufficient to calculate the within-class covariance matrix. Codes: 3: Montane fields of Pyrenean broom (*Genista purgans*); 5: Low maquis of *Cistus* sp.; 6: Western garrigues of *Rosmarinus officinalis*; 7: *Cistus* sp. and *Anthyllis cytisoides* shrubland; 12: *P. halepensis* woodland over evergreen oak maquia; 13: P. *halepensis* woodland over *Rosmarinus garigue*; 15: Forest of *P. uncinata*; 23: Aforestation of *P. nigra*; 25: Forests of *Quercus suber*; 26: Forests of *Quercus ilex*; 29: Cereal crops (1) (non-irrigated); 30: Cereal crops (2) (non-irrigated); 31: Rice fields.

Code	3	5	6	7	12	13	15	23	25	26	29	30	31
3	0.00	2.00	2.00	2.00	2.00	2.00	1.58	2.00	2.00	2.00	2.00	1.98	2.00
5	2.00	0.00	1.72	1.82	1.59	1.54	2.00	1.39	1.85	2.00	2.00	1.97	2.00
6	2.00	1.72	0.00	0.89	1.61	0.51	2.00	1.83	1.94	2.00	1.81	1.80	2.00
7	2.00	1.82	0.89	0.00	1.70	1.05	2.00	1.97	2.00	2.00	1.87	1.79	2.00
12	2.00	1.59	1.61	1.70	0.00	1.28	2.00	1.62	1.95	2.00	2.00	1.93	2.00
13	2.00	1.54	0.51	1.05	1.28	0.00	2.00	1.71	1.74	1.97	1.81	1.83	2.00
15	1.58	2.00	2.00	2.00	2.00	2.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00
23	2.00	1.39	1.83	1.97	1.62	1.71	2.00	0.00	1.61	2.00	2.00	1.97	2.00
25	2.00	1.85	1.94	2.00	1.95	1.74	2.00	1.61	0.00	1.95	2.00	1.99	2.00
26	2.00	2.00	2.00	2.00	2.00	1.97	2.00	2.00	1.95	0.00	2.00	2.00	2.00
29	2.00	2.00	1.81	1.87	2.00	1.81	2.00	2.00	2.00	2.00	0.00	1.99	2.00
30	1.98	1.97	1.80	1.79	1.93	1.83	2.00	1.97	1.99	2.00	1.99	0.00	2.00
31	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00

Table 2. Statistical separability (Jeffries-Matusita index) using the RDA ordination for those habitat types with a number of selected cells sufficient to calculate the within-class covariance matrix. Figures in italics are differences with values of the same statistic for the PCA-transformed data set. Only those differences whose absolute value is higher than 0.1 are included.

Code	3	5	6	7	12	13	15	23	25	26	29	30	31
3	0.00	2.00	1.99	2.00	2.00	1.99	2.00	2.00	2.00	2.00	2.00	1.97	2.00
							0.42					*−0.02*	
5	2.00	0.00	1.96	1.99	1.96	1.89	2.00	1.90	1.92	2.00	2.00	2.00	2.00
			0.25	*0.17*	*0.37*	*0.35*		*0.50*	*0.07*			*0.03*	
6	1.99	1.96	0.00	1.09	1.98	1.22	2.00	1.98	1.99	2.00	2.00	1.93	2.00
		0.25		*0.20*	*0.37*	*0.72*		*0.15*	*0.04*		*0.19*	*0.13*	
7	2.00	1.99	1.09	0.00	1.99	1.70	2.00	1.99	1.97	2.00	2.00	1.97	2.00
		0.17	*0.20*		*0.29*	*0.65*		*0.01*	*−0.02*		*0.13*	*0.18*	
12	2.00	1.96	1.98	1.99	0.00	1.79	2.00	1.73	1.98	2.00	2.00	2.00	2.00
		0.37	*0.37*	*0.29*		*0.50*		*0.10*	*0.02*			*0.07*	
13	1.99	1.89	1.22	1.70	1.79	0.00	2.00	1.81	1.94	2.00	2.00	1.91	2.00
		0.35	*0.72*	*0.65*	*0.50*			*0.10*	*0.19*	*0.03*	*0.19*	*0.08*	
15	2.00	2.00	2.00	2.00	2.00	2.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00
	0.42												
23	2.00	1.90	1.98	1.99	1.73	1.81	2.00	0.00	1.91	1.99	2.00	2.00	2.00
		0.50	*0.15*	*0.01*	*0.10*	*0.10*			*0.31*	*−0.01*	*0.00*	*0.03*	
25	2.00	1.92	1.99	1.97	1.98	1.94	2.00	1.91	0.00	1.54	2.00	2.00	2.00
		0.07	*0.04*	*−0.02*	*0.02*	*0.19*		*0.31*		*−0.41*		*0.01*	
26	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.99	1.54	0.00	2.00	2.00	2.00
						0.03		*−0.01*	*−0.41*				
29	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00	1.95	2.00
			0.19	*0.13*		*0.19*						*−0.04*	
30	1.97	2.00	1.93	1.97	2.00	1.91	2.00	2.00	2.00	2.00	1.95	0.00	2.00
	−0.02	*0.03*	*0.13*	*0.18*	*0.07*	*0.08*		*0.03*	*0.01*	*0.00*	*−0.04*		
31	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00

acquisition should be contemplated. In the meantime, data fusion of imagery acquired by complementary systems might result in products having the required resolution in both time and space.

Acknowledgments

References

BARTALEV, S., ACHARD, F., ERCHOV, D., and 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*, Belgirate, Italy, 3–6 April 2000, http://vegetation.cnes.fr:8080.

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.

DEFRIES, R. S., HANSEN, M., and TOWNSHEND, J. R. G., 1995, Global discrimination of land cover types from metrics derived from AVHRR Pathfinder data. *International Journal of Remote Sensing*, **54**, 209–222.

DEVILLERS, P., DEVILLERS-TERSCHUREN, J., and LEDANT, J. P., 1991, *Corine Biotopes Manual. Habitats of the European Community*. (Luxembourg: Commission of the European Communities).

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

EIDENSHINK, J. C., 1992, The 1990 conterminous US AVHRR data set. *Photogrammetric Engineering and Remote Sensing*, **58**, 809–813.

HOLBEN, B. N., 1986, Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal 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 Sensing*, **10**, 827–833.

LLOYD, D., 1990, A phenological classification of terrestrial vegetation cover using short wave vegetation index imagery. *International Journal of Remote Sensing*, **11**, 2269–2279.

LOBO, A., IBÁÑEZ-MARTÍ, J. J., and CARRERA GIMÉNEZ-CASSINA, C. *International Journal of Remote Sensing*, **18**, 3167–3193.

LOVELAND, T. R., and BELWARD, A. S., 1997, The IGBP-DIS global 1 km land cover dataset, DISCover first results. *International Journal of Remote Sensing*, **18**, 3289–3295.

LOVELAND, T. R., MERCHANT, J. M., OHLEN, D. O., and BROWN, J. F., 1991, Development of a land cover characteristics for the conterminous US. *Photogrammetric Engineering and Remote Sensing*, **57**, 1453–1453.

LOVELAND, T., REED, B., BROWN, J., OHLEN, D., ZHU, Z., YANG, L., and MERCHANT, J., 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.

MAISONGRANDE, P., DUCHEMIN, B., and DEDIEU, G., 2002, VEGETATION/SPOT: An operational mission for the Earth monitoring. Presentation of new standard products. In *Recent Advances in Quantitative Remote Sensing*, edited by J. A. Sobrino (Valencia: Universitat de València), pp. 880–884.

MYNENI, R. B., HALL, F. G., SELLERS, P. J., and MARSHAK, A. L., 1995, The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 481–486.

RAO, C. R., 1964, The use and interpretation of principal component analysis in applied research. *Sankhyaá, Series A*, **26**, 329–358.

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.

TOWNSHEND, J. R. G., JUSTICE, C., and KALB, V., 1987, Characterization and classification of South American land cover types using satellite data. *International Journal of Remote Sensing*, **8**, 1189–1207.

TUCKER, C. J., TOWNSHEND, J., and GOFF, T. E., 1985, African land-cover classification using satellite data. *Science*, **227**, 369–375.