Land Cover Map for Portugal using VEGETATION data

Célia Gouveia [a,b] and Carlos C. DaCamara[b]

[a] Escola Superior de Tecnologia, Instituto Politécnico de Setúbal, Campus do IPS, Estefanilha, Setúbal, Portugal – cgouveia@est.ips.pt
[b] Centro de Geofísica da Universidade de Lisboa, Ed. C8, Campo Grande, 1749-016 Lisboa, Portugal – carlos.dacamara@meteo.pt

Abstract – In this work we begin by performing a Principal Component Analysis (PCA) on NDVI computed over clear-sky pixels from VEGETATION data. Two techniques (i.e. ISODATA and Fuzzy Clustering) are applied to the two retained Principal Components and used to produce a land cover map over Portugal. Quality of performed classifications is assessed based on a confusion matrix, using as reference the Corine Land Cover Map. Obtained results seem to be quite reasonable, namely in the case of forests and crops and show to be comparable with other classifications such as GLC2000.

Keywords: Land cover, satellite imagery, fuzzy algorithm, Principal Components.

1. INTRODUCTION

The main purpose of the VEGETATION system is to provide accurate measurements of basic characteristics of vegetation canopies on an operational basis. These measurements are especially useful in scientific studies involving both regional and global scale experiments over long time periods, as well as in systems designed to monitor important vegetation resources, like crops, pastures and forests. One especially important application of the data is to build up land cover maps that will constitute the thematic data required by several algorithms, such as air pollution, numerical weather prediction and land cover change among others.

The VEGETATION sensor on-board the SPOT 4 platform is specifically prepared to monitor the surface, and makes available daily images at a 1 km resolution.

Among the applications of VEGETATION, GLC2000 is worth mentioning since its main objective is to develop a Global Land Cover Map, using data that were collected by the sensor during the year 2000. A map for Europe has already been produced in the framework of the program (Bartholomé, 2002).

The goal of this work is to produce land-classified images of land cover over Portugal, using a method based on *fuzzy logic clustering* and compare the results with those from a land-classified image obtained using a standard *clustering* procedure (*ISODATA*).

2. DATA AND RESULTS

The data consist of B2, B3, and SWIR (i.e. red, near infrared, and short-wave infrared) images from the SPOT-4 VEGETATION satellite, covering the period from May to October.

The reference map used is the Corine Land Cover Map (CLC90), available on a 250m by 250m grid, which has been aggregated from the original vector data at 1:100,000. We have reprojected the Corine Land Cover Map to Geographic Coordinates and the pixels were geocoded to the size of 1000 m, applying the nearest neighbor scheme. The Corine Land Cover is a key database for integrated environmental assessment and provides a pan-European inventory of biophysical land cover, using a 44 class-nomenclature. This version is available 'on line' in the European Environmental Agency web site and corresponds to an old version from 1990. Currently it is possible to find a new version of Corine for the year 2000, which is not yet accessible in the case of Portugal. The 44 classes were compared with the GLC2000 legend and then regrouped into the 13 GLC2000 classes. (Table A).

Table A. Codes and legend of GLC2000 classes

Codes	GLC2000 classes
1	Closed deciduous broadleaved forest
2	Closed evergreen needleaved forest
3	Mixed needleaved and broadleved forest
5	Mixed closed forest and shrubland
6	Closed shrubland
7	Cultivated and managed areas, hebaceous crops, non-irrigated
8	Cultivated and managed areas, hebaceous crops, irrigated
9	Permanently cropped area with rainfed shrubs crops
11	Grassland
12	Wetland
13	Bare soil and sparsely vegetated area
14	Outside regions

A fuzzy algorithm developed by Calado and DaCamara (2002) for NOAA-AVHRR sensor was adapted to VEGETATION and then applied to the dataset (Calado et al., 2005). Regularly distributed days in the study period with less than 10% of cloudy pixels (27 days) were selected and the Normalized Difference Vegetation Index (NDVI) was computed over clear-sky pixels using B2 and B3 reflectance values.

A Principal Component Analysis was then performed on NDVI. In order to improve the classification efficiency, we have decided to rely on the N Rule (Preisendorff, 1988), and accordingly we have retained the first 2 PCs.

The Empirical Orthogonal Functions (EOFs) corresponding to the selected PCs are shown in Figure 1. PC1 is roughly an accumulation of NDVI in the period whereas PC2 reflects the difference of NDVI between winter and summer. Figure 2 presents the spatial representation of the first 2 PCs. Dark regions on the left panel of Figure 2 mean high values of PC1 and high vegetation during all year and light regions mean low vegetation. Dark regions on the right panel mean high values of PC2 in winter and therefore more vegetation in winter then in summer.

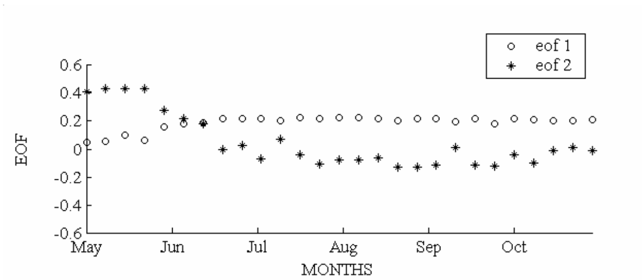

Figure 1. Empirical Orthogonal Functions for the first two Principal Components.

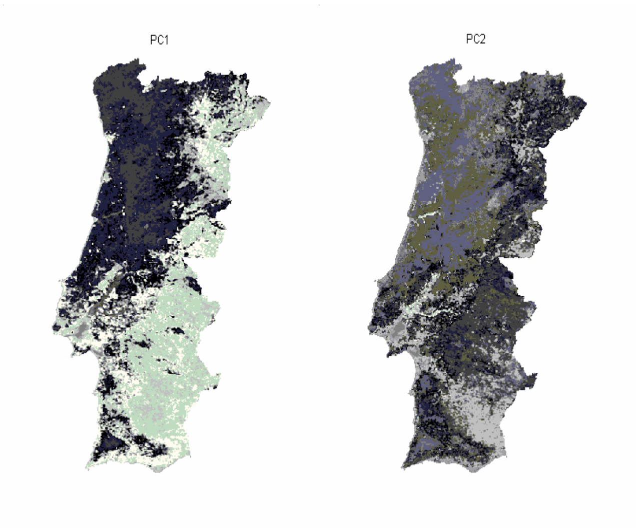

Figure 2. Images of the first two Principal Components.

Whilst urban areas may often be visually detected on low-resolution satellite imagery, they are usually difficult to identify from digital data. This problem arises because their spectral signature is similar to bare soil (Eva et al., 2004). On the other hand, there may be overlaps with other classes such as arable land or even forest (Mucher et al., 2002). In order to overcome this problem it is useful to mask these classes e.g. by relying on ancillary data, such as the Corine Land Cover Map (Latifovic et al., 2004, Agrawal et al., 2003). In this work, masks for urban areas and wetland were applied before performing the clustering algorithms.

The Corine Land Cover Map for Portugal with 13 classes is presented in Figure 3. It is worth noting that the CORINE map, which is more than ten years old, was based on LandSat imagery, which has a much higher resolution (30m) then the VEGETATION sensor.

A fuzzy logic classification using *C-MEANS* was performed for the study area using the retained two principal components and specifying 44 clusters. In this type of classification each pixel might belong to more then one class and the respective membership functions give the possibility to belong to each class. A confusion matrix was then computed between the 44 clusters and the 11 remaining classes. Histograms for each cluster were plotted and inspected and then used to define a set of rules to attribute a pixel to a given class. Table B shows the defined rules for each class and the clusters assigned to each of class. Figure 4 shows the obtained histograms for clusters 7 and 28, which were both assigned to class code 3. The result of the performed classification is presented in Figure 5.

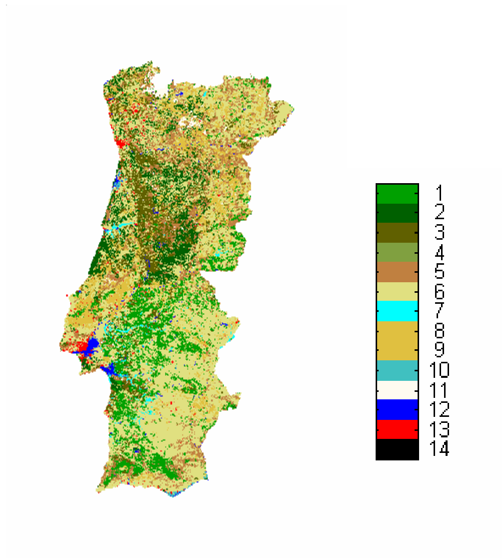

Figure 3. Corine Land Cover Map, (CLC90).

Table B. Defined rules to assigned the clusters to the grouped Corine Map classes presented in the classified image and the clusters selected to each class.

Code	Rules	Selected Clusters
1	C1 > 15% C6 < 50% C5 > 15%	4,8,15,20,27,30,37,43
2	C2 > 15% C6 < 40%	6,11,12,17,24,29
3	C3 > 15% C2 < 20% C6 < 50%	7,28
5	C5 > 10% C6 < 50% C1 < 10%	14
6	Rules for remaining codes are not fulfilled	1,2,3,9,10,13,16, 18,19,22,23,26 31,32,33,34,35, 38,39,40,42,44
7	C7 > 15%	25
11	C11 > 5% C12 < 29%	41
12	C13 > 10%	5,21,36

We have also applied to the same data the so-called Iterative Self-Organizing Data Analysis Technique ISODATA (again with 44 clusters). A confusion matrix was then computed between the 44 clusters obtained by ISODATA and the remaining 11 classes. If a cluster has more then 60 % pixels belonging to a given class then it is allocated to this class. If not we choose the next percentage for assigning the cluster to a given class. Obtained results are presented in Figure 6 and a comparison with those in Figures 4 and 5 allows to conclude that there seems to be a general good agreement with the reference data, namely in the case of first method (C-means).

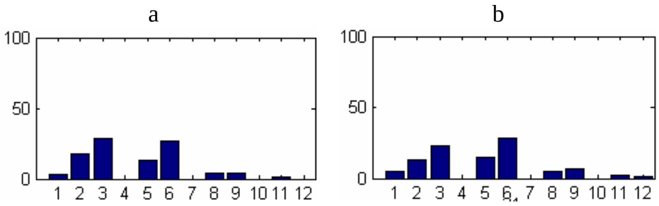

Figure 4. Clusters histograms assigned to class code 3 a) cluster 7 b) cluster 28.

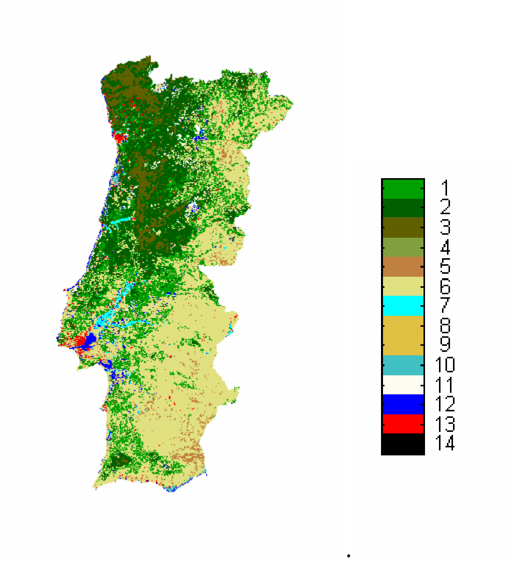

Figure 5. Fuzzy Clustering Classification, using 2 Principal Components and 44 Clusters.

In order to assess the quality of the classifications performed we have computed some accuracy measures. A global measure of classification quality is given by the overall accuracy (A) and the quality of each individual class may be assessed based on three different measures; the Producer's accuracy (PA), the User's accuracy (UA) and the Comparison Index (CI). These measures are defined as follows:

$$A = \frac{\text{Correctly Classifieded Pixels}}{\text{Total Number of Pixels}} \quad (1)$$

$$PA = \frac{\text{Correctly Classified ed Pixels}}{\text{Total Detected Pixels from Reference}} \quad (2)$$

$$UA = \frac{\text{Correctly Classifieded Pixels}}{\text{Total Detected Pixels from Classified data}} \quad (3)$$

$$CI = \sqrt{UA \times PA} \quad (4)$$

Obtained results are given in Table C and suggest that the fuzzy technique revealed to be the best in certain cases, namely in identifying arable land (non irrigated and irrigated) and bare soil. Furthermore, the overall accuracy reveals is higher.

It is worth comparing our results with those from other classifications such as GLC2000. Results respecting to GLC2000 for Portugal are presented in Figure 6 and the respective quality measures against the Corine Land Cover are given in Table D. Results generally point to a better skill in the case of fuzzy clustering. A possible exception is the shrubland class (code 5), but even this class is clearly underestimated in the case of the GLC2000 classification.

The different resolutions between VEGETATION and Land Sat are expected to lead to classifications capturing different land cover characteristics. Furthermore, the observed scene is different due to the different periods of acquisition data. In this way, these results should be regarded as a comparison with those obtained in GLC2000.

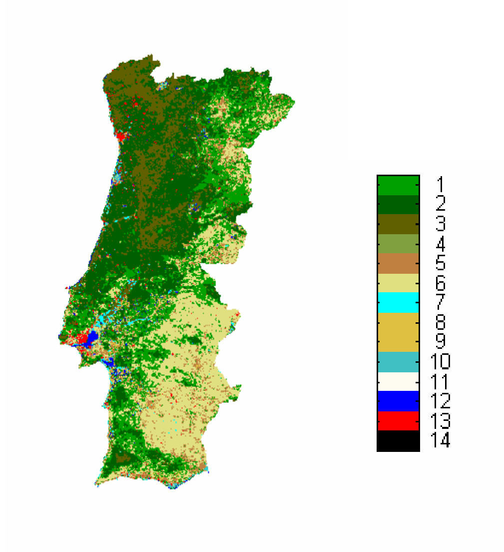

Figure 6. ISODATA Classification, using 2 Principal Components and 44 Clusters.

Table C. Quality parameters of the performed classifications: A - Accuracy, PA - Producer's Accuracy, UA - User's Accuracy and CI - Comparison Index

	Fuzzy Clustering				ISODATA			
Classes	PA	UA	CI	A	PA	UA	CI	A
Forest	65	41	51		84	37	56	
Shrubland	4	19	8		8	14	10	
Arable land (non irrigated)	61	57	59	47	38	69	51	43
Arable land (irrigated)	26	20	23		13	10	11	
Bare soil	8	8	8		2	13	5	
Water	37	23	29		29	38	33	

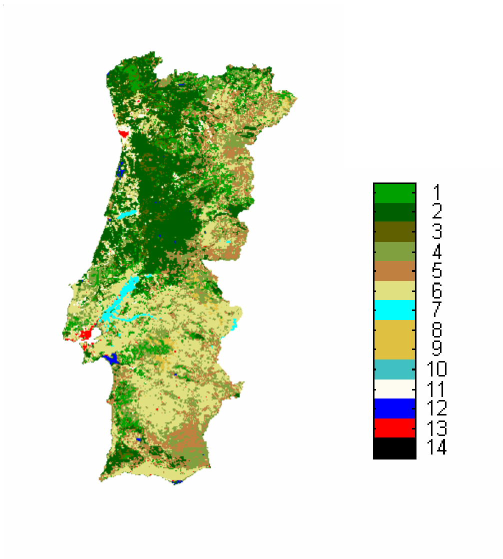

Figure 6. GLC2000 map for Portugal

Table D. Quality parameters respecting to the GLC2000 classification.

<table>
<tr><td></td><td colspan="4">GLC2000</td></tr>
<tr><td>Classes</td><td>PA</td><td>UA</td><td>CI</td><td>A</td></tr>
<tr><td>Forest</td><td>59</td><td>46</td><td>52</td><td rowspan="6">39</td></tr>
<tr><td>Shrubland</td><td>21</td><td>20</td><td>21</td></tr>
<tr><td>Arable land (non irrigated)</td><td>41</td><td>63</td><td>51</td></tr>
<tr><td>Arable land (irrigated)</td><td>27</td><td>20</td><td>23</td></tr>
<tr><td>Bare soil</td><td></td><td></td><td></td></tr>
<tr><td>Water</td><td>23</td><td>84</td><td>44</td></tr>
</table>

3. CONCLUSIONS

We have performed a Principal Component Analysis (PCA) on NDVI computed over clear-sky pixels from VEGETATION data. Two techniques, namely a fuzzy logic clustering technique (C-means) and a k-means iterative clustering technique (ISODATA), were applied to the two retained Principal Components and used to produce a land cover map over Portugal. The two different techniques for land use classification have results of reasonable quality, if we take into account the different resolution and the different acquisition data period between the reference map and our data. When comparing our results with those from the GLC2000 classification, it appears that the fuzzy clustering technique is an adequate method for Land Cover Classification.

We think that our results might have been improved if we had used as reference the new Corine Land Cover Map build (based on data for the year 2000). Improvements might also have been achieved by using a bi-directional effect correction method (e.g. Roujean, 1992, Champeaux, 2002). We believe that this would especially improve the identification of the low represented classes, namely those corresponding to sparsely vegetated areas and bare soil. Accordingly it is anticipated that the fuzzy method that was described in this work will be applied to other vegetation indices obtained from geometrically corrected data.

4. ACKNOWLEDGEMENTS

The authors are grateful to Dr. Etienne Bartholomé from JRC and to Prof. José Cardoso Pereira from ISA (Portugal) who kindly provided the VEGETATION raw data.

5. REFERENCES

Agrawal, S., Joshi, P. K., Shukla Y. and Roy, P. S. (2003). SPOT VEGETATION multi temporal data for classifying vegetation in south central Asia. Current Science, Vol 84, No 11, 10 June 2003.

Bartholomé E., 2002. GLC 2000 European window. Status of partnership – first results. GLC 2000 "first results"workshop, Ispra 18-22 March 2002.

Calado, T.J. and DaCamara, C.C., 2002: A fuzzy approach to masking cloud systems over land. LSA SAF Training Workshop, EUMETSAT, Lisbon, Portugal, July 2002.

Calado,T.J., Gouveia C. and DaCamara, C.C, 2005.A fuzzy approach to masking cloud systems over land. 31st International Symposium on Remote Sensing of Environment Global Monitoring for Sustainability and Security, Saint Petersburg Russian Federation, June 20-24, 2005

Champeaux J-L, Garrihues, S. and Gouveia, C. (2002) Correction of bi-directional effects using BDC algorithm over France and his impacts on Land Cover Classifications. Land SAF Workshop, 8-10 July 2002, Lisbon, Portugal.

Eva, H.D., Belward, A.S., DeMiranda E.E., DiBella C.M., Gond V., Huber, O., Jones, S., Sgrenzaroli, M. and Fritz S., (2004). A land cover map of South America. Global Change Biology, 10, 731–744.

Mucher, C.A., de Badts, E.P.J., 2002. Global Land Cover 2000: Evaluation of the SPOT VEGTATION sensor for land use mapping. Wageningen, Alterra, Green World Research pp 49

Latifovic, R., Zhub, Z-L., Cihlara, J., Girib, C., Olthof, I, (2004). Land cover mapping of North and Central America—Global Land Cover 2000. Remote Sensing of Environment, 89, 116-127

Preisendorff, R.W. (1998) Principal Componente analysis in Meteorology and Oceanography. Vol 17, Developments in Atmospheric Science, Elsevier, 425 pp.

Roujean, J. L. and Leroy, M. and Deschamps, P. Y. (1992) A Bidirectional Reflectance Model of the Earth's Surface for the Correction of Remote Sensing Data. Journal of Geophysical Research, 97:20.455-20.468.