INTEGRATION AND USAGE OF INDICES, FEATURE COMPONENTS AND TOPOGRAPHY IN VEGETATION CLASSIFICATION FOR REGIONAL BIODIVERSITY ASSESSMENT

Ayşegül Domaç[1], Ugur Zeydanli[2], Ertan Yeşilnacar[3], M. Lütfi Süzen[4]

[1] Conservation GIS Laboratory, Department of Biology, Middle East Technical University, Ankara, TURKEY
aysegul@kafkaskoruma.net

[2] Conservation GIS Laboratory, Department of Biology, Middle East Technical University, Ankara, TURKEY
ugur@kafkaskoruma.net

[3] Department of Geomatics, University of Melbourne, 3010, Melbourne, Victoria, AUSTRALIA
ertan@sunrise.sli.unimelb.edu.au

[4] RS –GIS Lab. Department of Geology, Middle East Technical University, Ankara, TURKEY suzen@metu.edu.tr

KEY WORDS: Remote Sensing, Vegetation, Classification, Accuracy, Ecology, Spectral

ABSTRACT:

The classification of vegetation has been an important research subject in botany, ecology, geography, and other disciplines to map the differences in vegetation types. Classifying vegetation by remote sensing is valuable because it can determine vegetation distribution and occurrence for very large areas in a short time. Advances in technology have led to developments in methods of vegetation classification, leading to the creation of new and more sophisticated components and powerful techniques. Classifying original bands and/or image components may cause unsatisfactory results in spectrally chaotic fields. In such cases, the demand for accurate land-use, land-cover, vegetation, and forestry information may require more explanatory components those components should represent specific information for target land-covers and not contain redundant knowledge.

In this study, spectral bands of Landsat Thematic Mapper and topographic data were used as an input. Different image components and indices were produced and then used in the Maximum Likelihood Classification method. In order to find out proper inputs for our case, newly produced components and indices were statistically compared and the bands that include the information about vegetation are selected. Overall accuracy parameter that is obtained from the Error Matrix helped to evaluate the results of the classification. Results obtained in this study suggest that using these spectrally improved bands and indices; the accuracy of the classification could be increased up to 10-15 percent.

1. Introduction

The need to map wide areas with limited resources forced the improvement of vegetation classification methods by using satellite images. Conservation agencies use these images to extract variety of vegetation types in order to assess the biodiversity of a region. Among the possible commercial satellite systems, Landsat images have got some serious advantages over other systems such that: 30 m ground resolution yields as a convenient resolution for regional vegetation studies with a minimum mapping unit of 100 ha., the spectral coverage fits well to the vegetation spectra, and the wide swath width yields in less number of images to process which maintains the coherence of the imagery. Furthermore Landsat system is a mature system dating back to early 1970's, hence plenty of researchers have exploited many mapping methods. However still the classification results are way off the desired levels, only major homogenized groups of forest can be discriminated, yet the conservation measures require a more detailed legend. Subdivision of this multi-spectral continuum into meaningful vegetation classes is a major challenge that requires careful consideration (Brook and Kenkel, 2002). For instance, visual analysis of different bands/colour composites from a multispectral dataset with constant pixel resolution still reflects the same spatial structure, even if the contrast between different scene elements (i.e. forest patches versus non-forest patches) might considerably vary for the different band combinations (Bryan,1988).

With the aid of classical vegetation indices or raw input bands hardly any classification can fulfill this need; hence some improvements should have to be made either by post classification sorting or by adding new components derived from the original input bands to the classification process. The extraction of spectral information related to this type of target from Landsat TM imagery has been achieved through the use of image processing techniques such as band rationing and principal component analysis (Sabine1999). The major fact behind this new component adding is to create a spectral subset of the data itself and to create more explanatory variables which can be used to exploit the variance of the vegetation types that are desired to be mapped from imagery. Due to the high similarity among individual bands of a multispectral image, statistical data compression tools like principal component analysis (PCA) are often applied in image analysis and image classification to reduce the amount of redundant information (Ricotta *et al.*, 1999). The objective of the study was to improve the accuracy of vegetation classification by using future components which were constituted by using raw bands and various vegetation indices.

2. Study Area

Study area is located in the Southern part of Turkey in the Mediterranean region and covers approximately 235km² (Figure 1). Elevation values of the region vary between

300 m and 2500 m above sea level. The area is characterized by a variety of landcover types, including; forest areas, open areas and farmland which were suitable for the purpose of this study.

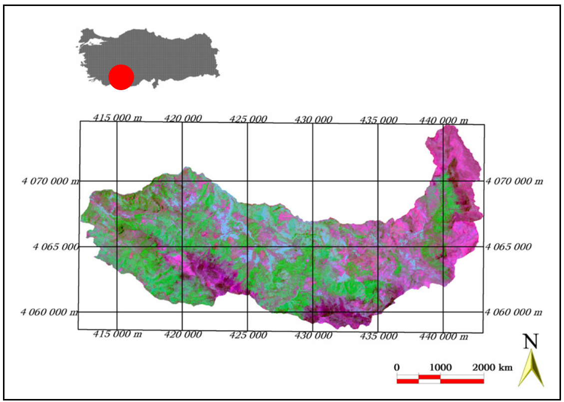


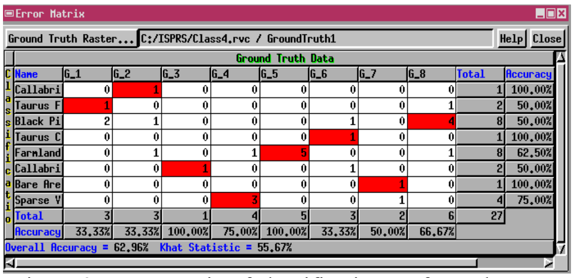
Figure 1. Location of the study area

3. Methodology

Using raw bands of Landsat in the classification process is a widely used way of extracting vegetation maps. But the statistical similarities of vegetation spectral responses, spatial resolution of data and presence of similar species sometimes do not allow obtaining the desired results from the original bands. As a first step of the study raw bands of Landsat ETM belonging to Mediterranean region were classified. The maximum likelihood method was used to classify the image because, unlike the minimum distance and the parallelepiped classifiers, this technique takes into account both the spectral variability within and between classes (Fahsi *et al*,, 2000).

The classification legend was determined by using the available data such as forest management maps and reconnaissance field survey results. While forming the training data, this legend was taken into account and eight different vegetation classes; Callabrian Pine, Black Pine, Taurus Fir, Taurus Cedar, Farmland, Sparse vegetation were discriminated.

To check the accuracy of the results, ground truth data set with 26 reference point were determined using the 1/25 000 scaled forest management map of the region. When the training set was applied on the classified image, an overall accuracy of 62.96% was obtained, which is not satisfactory for this kind of studies.



Error Matrix

Ground Truth Raster... C:/ISPRS/Class4.rvc / GroundTruth1 Help Close

Ground Truth Data

Classification Name	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	Total	Accuracy
Callabri	0	1	0	0	0	0	0	0	1	100.00%
Taurus F	1	0	0	0	0	0	0	1	2	50.00%
Black Pi	2	1	0	0	0	1	0	4	8	50.00%
Taurus C	0	0	0	0	0	1	0	0	1	100.00%
Farmland	0	1	0	1	5	0	0	1	8	62.50%
Callabri	0	0	1	0	0	1	0	0	2	50.00%
Bare Are	0	0	0	0	0	0	1	0	1	100.00%
Sparse V	0	0	0	3	0	0	1	0	4	75.00%
Total	3	3	1	4	5	3	2	6	27	
Accuracy	33.33%	33.33%	100.00%	75.00%	100.00%	33.33%	50.00%	66.67%		

Overall Accuracy = 62.96% Khat Statistic = 55.67%

Figure 3. Error matrix of classification performed on raw bands.

At this step a new method was implemented, in order to increase the accuracy of the result. Suitable vegetation indices and image components were produced by using Principal Component Analysis (PCA) which is a technique for removing or reducing the duplication or redundancy in multispectral images and for compressing all of the information that is contained in an original *n*-channel set of multispectral images into less than *n* channels or, more specifically, to their principal components (Ricotta *et al*., 1999).
In this study the main inputs of the feature components are the indices. Two sets of indices were used; the first set includes the vegetation indices which directly give the spectral response of chlorophyll by using the ratio between red and NIR bands. The second set was used to remove the soil noise by changing slope value of red and NIR bands.

First set of indices are most commonly used remote sensing tools for extracting green vegetation cover that employ red and near infrared vegetation such as Normalized Difference Vegetation Index (NDVI) (Drake *et al*., 1999). In addition to NDVI, Global Vegetation Index (GVI), Infrared Percentage Vegetation Index (IPVI), Transformed Vegetation Index (TVI), and Tasseled Cap Greenness Index were used. Equations of these indices are given in Table 1.

Normalized Difference Vegetation Index	$NDVI = \frac{NIR - red}{NIR + red} \times 255$
Global Vegetation Index	GVI=-0,2848*TM1-0,2435*TM2-0,5439*TM3+0,7243*TM4+0,0840*TM5-0,1800*TM7
Greenness	Greenness = -0.2848(TM1)-0.2435(TM2)-0.5436(TM3)+0.7243(TM4)+0.0840(TM5)-0.1800(TM7)
Transformed Vegetation Index	TVI=100 * [((NIR - red) / (NIR + red))½)+0,5]
Infrared Percentage Vegetation Index	$IPVI = \frac{1}{2}(NDVI+1)$
Soil Adjusted Vegetation Index L=0 for high vegetation cover L=0 for low vegetation cover	$SAVI = \frac{NIR - red}{NIR + red + L}(1+L)$
Modified Soil Adjusted Vegetation Index 1	MSAVI1 = ((NIR-red) / (NIR + red +L) x (1+ L) L= 1-(2* slope * NDVI * WDWI) WDWI = NIR – slope * red
Modified Soil Adjusted Vegetation Index 2	$MSAVI2 = 1/2*((2*(NIR+1))-(((2*NIR)+1)^2-8(NIR-red))^{1/2})$

Table 1. Indices used in this study

Principal components of these six indices were calculated and the lists of image eigenvalue loadings for this transformation on all vegetation indices are given in
Table 2. According to this table, correlations of PC1 with the indices are very high except IPVI. This means PC1 has a great amount of information of these 5 indices. To include the spectral information of IPVI in the analysis, PC2 is used because IPVI has a high loading value in this component. 92.27 percent of the spectral information was collected on the first two principal components; PC1 and PC2 are selected as feature components of vegetation indices.

Axis	GREEN NESS	GVI	IPVI	ND	TVI
1	0.7340	-0.9673	-0.2756	0.8746	0.8622
2	0.1738	0.1960	-0.9519	0.1359	0.2095
3	0.5125	0.1605	0.1340	0.4491	0.4354
4	-0.4101	0.0071	0.0015	0.1178	0.0580
5	-0.0131	0.0000	0.0016	-0.0321	0.1404

Table 2. Correlation Between Input Rasters and Principal Components

The relationship between vegetation cover and the indices appears to change over the area according to the certain conditions such as soil cover type. To minimize the effect of soil on vegetation reflectance, second set of indices were used. These indices were Soil Adjusted Vegetation Index (SAVI), Modified Soil Adjusted Vegetation Index 1 (MSAVI1), and Modified Soil Adjusted Vegetation Index 2 (MSAVI2) (Table 1). For this study the first PC acquired from the 3 soil indices contains the spectral information adequate for the classification to normalize the effects that emerge due to the different soil types of the areas with low canopy of vegetation.
(Table 3)

Axis	MSAVI	MSAVI2	SAVI	Eigenvalues (%)
1	0.9999	0.0789	0.0046	84.4805
2	-0.0145	0.9969	-0.0629	15.5186
3	-0.0000	0.0000	0.9980	0.0009

Table 3. Correlation Between Soil Adjusted Vegetation Indices and Principal Components

Besides these feature extraction oriented indices, PCA were performed on raw bands in order to find if vegetation related information could be collected in few explanatory bands. In this transformation, examination of principal components eigenvector loadings determine which PC possesses information related directly to the spectral signatures of vegetation. Eigenvector loadings for PC2 in Table 4 indicate that PC2 describes the difference between the visible channels (TM1, 2, and 3) and the infrared (IR) channels (TM5 and 7) and also this component is commonly thought to be related to vegetation. Eigenvector loadings for PC3 (in Table 3) indicate that PC3 is dominated by vegetation. In this component both the loading values of TM3 and TM4 is negative but the difference between these two band were high because in TM3 chlorophyll is absorbed, on the contrary chlorophyll is highly reflected in the near infrared band. Therefore PC2 and PC3 were selected as feature components.

Axis	TM1	TM2	TM3	TM4	TM7	TM5	Eigen values (%)
1	0.9160	0.9667	0.9781	0.6093	0.9615	0.9329	85.731 0
2	-0.3436	-0.2114	-0.1570	0.6718	0.1201	0.3008	9.2461
3	-0.1460	-0.1245	-0.0459	-0.4201	0.2347	0.1756	3.9920
4	-0.1408	0.0019	0.1247	-0.0047	-0.0164	-0.0336	0.6456
5	-0.0100	0.0107	-0.0155	0.0286	0.0756	-0.0846	0.2556
6	0.0394	-0.0720	0.0284	0.0104	0.0070	-0.0099	0.1296

Table 4. Correlation Between Input Rasters and Principal Components

In addition to principal component bands, Decorrelation Stretched (DS) bands were used in this study. Even though these bands still show the properties of the original bands, the color separation of these bands are enhanced with significant band to band correlation. Decreasing the correlation of spectral data corresponds to exaggerating the color saturation without changing the distribution of hues (or relative color composition) (Gillespe et al., 1987).

At the end of these analyses it is assumed that; selecting PC1 and PC2 of vegetation indices, PC1 of soil indices, PC2 and PC3 of raw bands and DC3 and DC4 as feature components will remove the redundant data among multivariate datasets, such as multispectral remote sensing images and increase the accuracy of the classification.

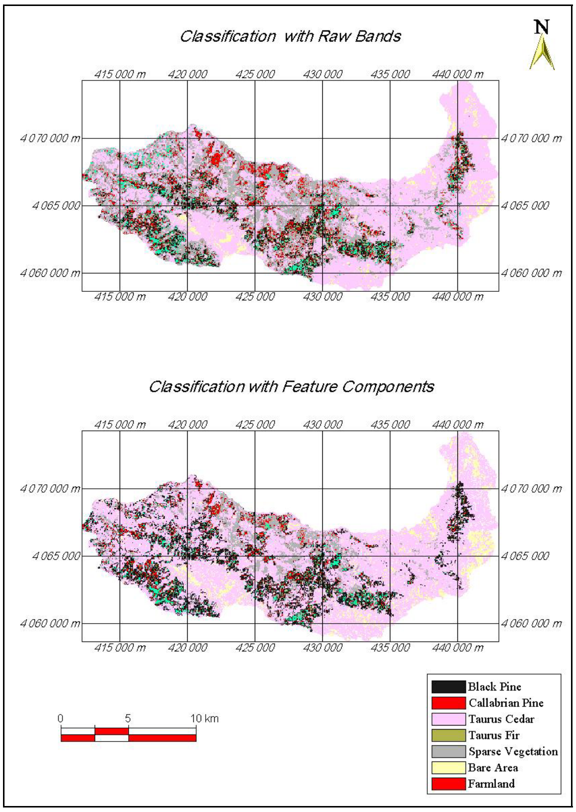


Figure 2. Classification results of both raw bands and feature components.

By using these feature components overall accuracy was increased to 76.92 %. This rise shows that the new formed bands were very successful in the discrimination of vegetation classes with very similar spectral reflectance values.

Error Matrix

Ground Truth Raster... C:/ISPRS/Class4.rvc / GroundTruth1 Help Close

Classification	Ground Truth Data									
Name	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	Total	Accuracy
Callabri	0	2	0	0	0	0	0	1	3	66.67%
Taurus F	2	0	0	0	0	0	0	0	2	100.00%
Black Pi	1	1	0	0	0	2	0	5	9	55.56%
Taurus C	0	0	0	0	0	1	0	0	1	100.00%
Farmland	0	0	0	0	5	0	0	0	5	100.00%
Callabri	0	0	1	0	0	0	0	0	1	100.00%
Bare Are	0	0	0	0	0	0	1	0	1	100.00%
Sparse V	0	0	0	3	0	0	1	0	4	75.00%
Total	3	3	1	3	5	3	2	6	26	
Accuracy	66.67%	66.67%	100.00%	100.00%	100.00%	33.33%	50.00%	83.33%		

Overall Accuracy = 76.92% Khat Statistic = 72.34%

Figure 4. Error matrix of classification which performed on feature components.

3. Discussion and Conclusion

The results obtained from this study show that Principal Components of raw bands and vegetation indices can extract valuable and concentrated vegetation information by creating a new variable set with eliminated interband correlation and reduced dimensionality of the data. In this method Principal Components which were highly loaded with the spectral information of desired band or index considered as a feature component, and used in the classification process. By using this method the accuracy of classification could be increased up to 15%.

This is a simple and fast technique which could easily be implemented on a landscape scaled classification studies.

4. Acknowledgements

We would like to thank WWF-Turkey for providing the necessary data for this study within the project "Priority Forest Areas of Mediterranean Turkey".

5. References

Brook, R. K. and Kenkel, N. C., 2002. A multivariate approach to vegetation mapping of Manitoba's Hudson Bay Lowlands, *International Journal of Remote Sensing*, vol. 23, no. 21, pp. 4761- 4776.

Bryant, J., 1988. On displaying multispectral imagery. *Photogrammetric Engineering and Remote Sensing*, vol. 54, pp.1739-1743.

Drake, N. A., Mackin, S., Settle, J. J., 1999. Mapping Vegetation, Soils, and Geology in Semiarid Shrublands Using spectral Matching and Mixture Modeling of SWIR AVIRIS Imagery. *Remote Sensing of Environment*, vol. 68, pp. 12-25.

Fahsi, A., Tsegaye, T., Tadesse, W., Coleman, 2000. T. Incorporation of digital elevation models with Landsat-TM data to improve land cover classification accuracy. *Forest Ecology and Management*, vol. 128, pp. 57-64.

Gillespie, A.R., Kahle, A.B. and Walker, R.E., 1987, Color enhancement of highly correlated images. Decorrelation and HSI contrast stretches. *Remote Sensing of Environment*, vol. 20, pp. 209-235.

Ricotta, C., Avena,G.C., Volpe, F, 1999. The influence of principal component analysis on the spatial structure of a multispectral dataset, *International Journal of Remote Sensing*, vol. 20, no. 17, pp. 3367- 3376.

Sabine, C., 1999, Remote sensing strategies for mineral exploration. *Remote Sensing for the Earth Sciences – Manuel of Remote Sensing*, 3rd edn, editted by A. Rencz (New York: American Society of Photogrametry and Remote Sensing/ John Wiley and Sons), pp.375-447.