Hyperspectral imagery, vegetation indices and discriminant analysis for monitoring forests. Case of Saadia forest in North-Western, Algeria.

[*1]Adda Ababou, [2]Mohammed Chouieb, [1]Djamel Saidi, [3]Abdelkader Bouthiba, [3]Mahfoud M'hamedi Bouzina and [5]Khalladi Mederbal

[*1] Biology Department, Faculty of Sciences, University Hassiba Ben Bouali, Chlef, Algeria.

[2]Agronomy Department, Faculty of Science and engineering, University Abd El Hamid Ibn Badis, Mostaganem, Algeria.

[3]Biology Department, Faculty of Sciences, University Hassiba Ben Bouali, Chlef, Algeria.

[4]Agronomy Department, Institute of Agronomical Sciences, University Hassiba Ben Bouali, Chlef, Algeria.

[5]Biology Department, Faculty of Science and earth Sciences, University Mustapha Stambouli, Mascara, Algeria.

Abstract

This work carried out in Saadia forest in northwestern Algeria and based on 11 vegetation indices analysed through discriminant analysis, is among a growing number of studies focusing on evaluating forests by using remotely sensed vegetation indices. The main results were the identification of three classes of vegetation, coniferous, deciduous and mixed vegetation, with an overall accuracy of classification of 91%. The discriminant analysis showed that the first discriminant function explained a discrimination of 93.89% and a Pillai's trace of 1.23 highly significant. The most important vegetation indices discriminating between vegetation were leaf pigments (ARI) indicating the weakness of both coniferous and mixed vegetation in the study area, then the canopy water content (NDII) suggesting that mixed vegetation was also subject to a high water stress, whereas, deciduous were mostly affected by moisture and water stress.

Key words: Algeria, Discriminant analysis, Hyperion, Saadia, Vegetation indices.

1. Introduction

With increasing pressure on forest resources and sustainability, accurate studies and information about their state and functioning are highly required. When monitoring vast forest areas traditional methods of evaluation are expensive and time consuming, hence, a reliable and fast method to circumvent these problems is extremely needed, one of the most reliable methods is remote sensing. Indeed remote sensing is a very important tool and robust way of characterizing and monitoring forest [1].

In this context, hyperspectral imagery represents one of the most important technological trends in remote sensing [2]; [3]. Since, every object has a single reflectance and absorption structure (spectrum) in different wavelengths [4], hyperspectral remote sensing is concerned by the measurement, analysis and identification of these spectra [5].

*Corresponding author: Address: Biology Department, Faculty of Sciences, University Hassiba Ben Bouali, Chlef, Algeria. E-mail address: ab_adda@yahoo.fr, Phone: +213552285695 Fax: +21327726304

Besides, using reflectance spectra to identify forest types and vegetal species, spectral vegetation indices derived from hyperspectral imagery are among the most important tools in monitoring forest biophysical, functioning and disturbance states [6].

These vegetation indices (VIs) are combinations of surface reflectance at two or more wavelengths designed to highlight a particular property of vegetation.

As reported by [7] VIs can be divided into five main categories according to the use of each index: (a) broadband indices; (b) narrowband indices; (c) leaf pigment indices; (d) stress indices, and (e) water stress indices.

To segregate the different forest type according to the most relevant vegetation index affecting each forest type, statistical methods are a very important way, the most important statistical methods in this case is discriminant analysis, in contrast to principal components analysis, discriminant analysis is explicitly a multiple-group procedure, and assumes that the groups are known before analysis on the basis of extrinsic criteria and that all individuals are members of one and only one of the known groups [8].

Discriminant analysis investigates the most parsimonious way to distinguish between groups, classifies any case into the group it most closely resembles [9]; [10] and successively identifies the linear combination of attributes [11] known as canonical discriminant functions (DF) which contribute maximally to group separation.

In this context the aim of this study was the use of hyperspectral imagery and vegetation indices combined with discriminant analysis to monitor, characterize and map the forest of Saadia in Northwestern Algeria.

2. Materials and Method

2.1. Study area

Located in northwestern Algeria, the study area (Figure 1) covers approximately 170 km^2, extending from 1° 9′ 33″ to 1°17′ 55″ East, and from 35° 38′ 55″ to 35° 51′ 17″ North and 70 km inland from the Mediterranean sea. With a variable altitude ranging from 250 to 1090 m above the sea level, it’s a typical Mediterranean area in terms of landscape structure composition and climate, characterized by hot and dry summers, with a dry period of 4 months (Mai to September) and relatively rainy winters with an average annual rainfall of 300 mm. In terms of vegetation, the landscape is covered with natural sclerophyllous and sparse vegetation alternating with bare soils.

Geologically, according to [12] and [13] the area characterized by facies of all ages eminently favourable to erosion, is composed mainly of cretaceous limestone, marls, sandstones, calcareous marls, and calcareous sandstone.

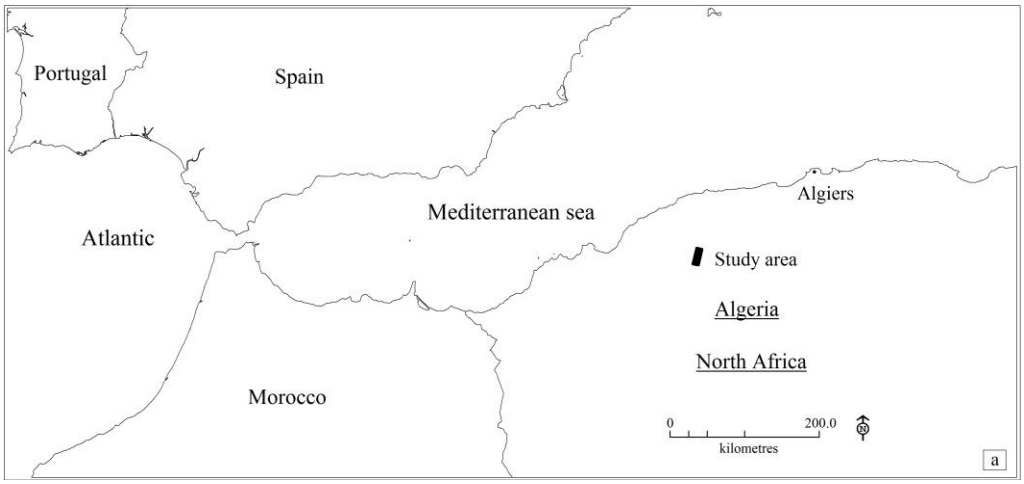

Figure 1. Location of the study area.

2.2. Data pre-processing and methodology

As mentioned by [14] Hyperion data require several phases of pre-processing to enable hyperspectral analysis. The mains pre-processing applied to our Hyperion image were, the removal of non-calibrated bands and bands with poor signal to noise ratio, the remaining bands were rescaled to radiance and converted to reflectance through the atmospheric correction, followed by a cross track illumination, de-smiling, de-striking and spectral bands smoothing, then, a minimum noise fraction analysis (MNF) [15] was applied in order to separate noise from data, followed by pixel purity index (PPI) procedure to cluster the purest pixels into image-derived from the MNF, through the n-D visualiser were determined the most spectrally pure endmembers derived from PPI, the spectral angle mapper (SAM) was then used to classify vegetation.

Finally, both forest classification and vegetation indices were subject to discriminant analysis.

2.3. Vegetation indices

Initially a colinearity test was performed between 25 vegetation indices (VIs), VIs which showed high variance inflation factors and those with no significant influence ($p \geq 005$) were removed, finally, only 10 VIs belonging to greenness VIs, light use Efficiency VIs, leaf pigment VIs and canopy water content, with significant ($p < 0.05$) to highly significant ($p < 0.01$) influence and low variance inflation factors were retained. Beside these 10 VIs, the fire fuel risque was also estimated.

3. Results

3.1. Classification

The Hyperion image was classified using SAM classifier and 25 training areas defined on the basis on several time field visits, as a result, 8 classes were identified (Figure 2) including vegetation (coniferous, deciduous and mixed vegetation) and bare soils, the overall accuracy of classification (Table 1) was equal to 91% and the Kappa coefficient equal to 0.89, which implies an excellent agreement between the SAM classification and the ground truth. According to Hyperion classification, the area covered by forest was 88.5 km^2, among them 17 km^2 of deciduous, 19.5 km^2 of coniferous and 52 km^2 of mixed vegetation.

Table 1. Classification accuracy and Kappa coefficient obtained in Saadia forest.

Land occupation	Producer Accuracy (%)	User Accuracy (%)
Coniferous	95.02	91.17
Deciduous	86.10	91.80
Mixed vegetation	91.69	89.93
Reddish brown fine sandy loam	92.29	83.18
Dark yellowish brown micacious loam	85.18	90.88
Brown fine sandy loam	88.68	92.52
Brown silty loam	85.42	88.36
Overall. Accuracy	90.76%	
Kappa Coefficient	0.894	

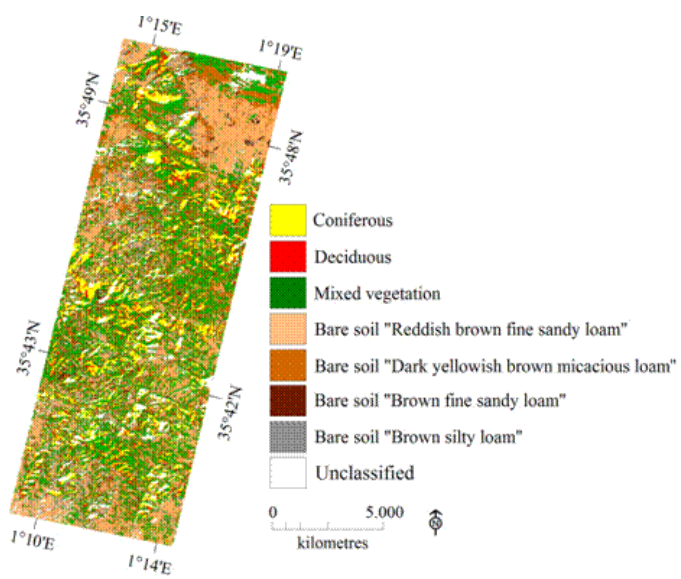

Figure 2. Classification of Hyperion hyperspectral imagery (WRS path 197 and row 35)

3.2. Factorial discriminant analysis

The discriminant analysis showed that coniferous, deciduous and mixed vegetation were well separated according to the 11 vegetation indices, indeed, in order to assesses whether the predictors vary enough to distinguish the different groups, the Lambda Wilks' [16] value was equal to 0.074 with a p-value less than 0.0001 (Table 2), meaning that vegetation indices highly distinguished the three groups of vegetation.

Table 2. Wilks' lambda and Pillai's trace values

	Wilks' Lambda	Pillai's trace
Lambda	0.074	
Trace		1.23
Observed value	119.6	71.22
Critical value	1.553	1.55
p-value	< 0.0001	< 0.0002

The distances between group centroid were all highly significant ($p < 0.0001$) according to F test, especially coniferous and mixed vegetation which were clearly segregated with a distance equal to 353.8, the distance between coniferous and deciduous was equal to 109.4, whereas, deciduous and mixed vegetation showed the smallest distance (60.8) indicating a weak overlap. These high differences between the three groups were also confirmed by the highly significant ($p < 0.0001$) Pillai's trace value equal to 1.23 (Table 2). The first DF with a canonical correlation of 0.94 explained the highest variance percentage between-group (93.89%) (Table 3).

Table 3. Discriminant analysis eigenvalues, cumulative discrimination and canonical correlations.

	Canonical variable	
	D1	D2
Eigenvalue	7.912	0.515
Discrimination (%)	93.894	6.106
Cumulative %	93.894	100.000
Canonical correlations:	0.942	0.583

This DF opposed on the positive side coniferous, which were more characterized by a High NDII (Normalized Difference Infrared index), NDVI (Normalized difference vegetation index), NDWI (Normalized difference water index) and ARI1 (Anthocyanin reflectance index 1), to mixed vegetation on the negative side, with a high MSI (Moisture stress index), SIPI (Structure insensitive pigment index), RGRI (Red green ratio index), ARI2 (Anthocyanin reflectance index 2) and fire fuel.

The second DF with a canonical correlation of 0.58 explained only 6.11% of the total inertia and was negatively correlated to deciduous more characterized ARVI (Atmospherically resistant vegetation index), mNDVI (Modified NDVI) and negatively related to mainly to MSI (Figure 3).

According to the standardized coefficients, leaf pigments (ARI) and canopy water content (NDII) appeared to be the parameters that contributes most to the first canonical variable with a respective standardized coefficients (in absolute value) of 0.74 and 0.675, which accounts for most of the discrimination between groups (Table 4). The second canonical variable was highly related to canopy water content (MSI and NDII) and Greenness (NDVI) as showed by their high standardized coefficients absolute values of respectively 1.481, 0.714 and 0.711.

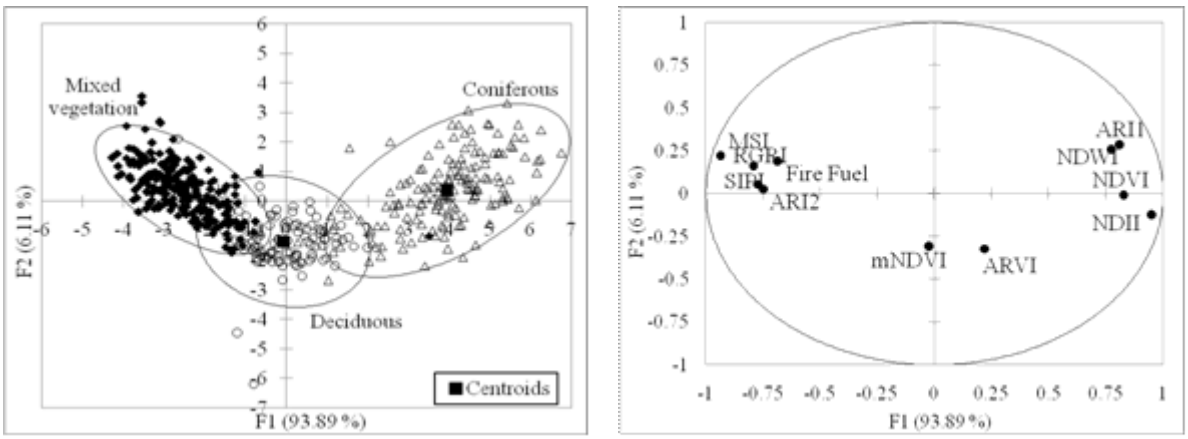

Figure 3. Discriminant analysis of the three vegetation classes and 11vegetation indices

Table 4. Standardized canonical discriminant coefficients of the most influent indices, NDVI, ARI1, ARI2, MSI, and NDII in Saadia forest.

Standardized canonical discriminant coefficients	D1	D2
NDVI	- 0.100	0.711
ARI1	0.740	0.504
ARI2	- 0.742	- 0.321
MSI	0.048	1.481
NDII	0.675	0.714

Furthermore, the confusion matrix used to test the accuracy of the discriminant analysis by displaying the proportion of correct and incorrect predictions produced by the DF equation [17], showed that 96.15% of coniferous, 77.06% of deciduous and 96.67% of mixed vegetation were correctly classified by the discriminant function with an overall accuracy of 92.28% (Table 5).

Table 5. Confusion matrix for the cross-validation results.

	Conifer	Deciduous	Mixed vegetation	Total	Correct
Conifers	150	6	0	156	96.15%
Deciduous	4	84	21	109	77.06%
Mixed vegetation	1	7	232	240	96.67%
Total	155	97	253	505	92.28%

4. Discussion

When monitoring vast forest areas traditional methods of evaluation are expensive and time consuming, the potential of the new algorithms and indices calculations used in remote sensing represents a reliable and fast method to circumvent these problems, among the most reliable methods in forest studies, the spectral angle mapper (SAM) algorithm and vegetation indices, especially, when supported by reliable statistical studies.

As a case study, we tested the above methods in a Mediterranean forest, suffering from anthropogenic, natural pressure and subject to repeated fires, causing the decrease in forest area and biodiversity. Spectral angle mapper classification yielded an overall accuracy equal to 91%, with, both user's accuracy (Probability that a classified pixel was correctly assigned) and a producer's accuracy (How much each category was correctly classified) greater than 80% for each class.

The Kappa coefficient was equal to 0.89, which according to the subjective scale of assessment of [18] implies an excellent agreement between the SAM classification and the ground truth. When focusing on forest biophysical, functioning and disturbance states, as well as external factors affecting canopy reflectance, canonical discriminant analysis through vegetation indices, accurately segregated between three vegetation types in the study area (coniferous, deciduous and mixed vegetation), highly significantly distant according to Mahalanobis distance.

The Pillai's trace value most robust criterion in discriminant analysis [19], suggested that the most important discriminant function was the first one, explaining the highest variance percentage between-group (93.89%), the three groups of vegetation were highly distinguished by VIs as showed by the highly significant Lambda Wilks' value, the accuracy of discrimination was 96.15% for coniferous, 77.06% for deciduous, 96.67% for mixed vegetation and 92.28% of total accuracy.

Each of the three vegetation classes was highly characterized by specific vegetation indices, However, according to the standardized discriminant coefficients, the larger the standardized coefficient, the greater is the contribution of the respective variable to the discrimination between groups [20], the most important vegetation index discriminating between vegetation classes on the first discriminant function was mainly, leaf pigments (Anthocyanin Reflectance Index), indicating the weakness of both coniferous and mixed vegetation in the study area. Indeed,

according to [21] leaf pigments are stress-related pigments, present in higher concentrations in weakened vegetation.

The second important vegetation index was canopy water content (Normalized Difference Infrared Index), this index designed to provide a measure of the amount of water contained in the foliage canopy [22], suggest that mixed vegetation was subject to a high water stress, whereas, deciduous were mostly affected by moisture and water stress on the second discriminant function.

Conclusion

From an operational perspective, the use of hyperspectral imagery and vegetation indices can be of great importance particularly in areas suffering from natural degradation and anthropogenic pressure such as our study area, since it can be associated with the mapping and monitoring of the forest biophysical, functioning and disturbance states frequently pronounced in such areas.

References

[1] Skoupy O, Zejdova L, Hanus J. The use of hyperspectral remote sensing for mapping the age composition of forest stands. J For Sci 2011;57:287-297.
[2] Pearlman J, Barry P, Segal C, Shepanski J, Beiso D, Carman S. Hyperion, a space-based imaging spectrometer. IEEE Trans Geosci Remote Sens 2003;41:1160-1173.
[3] Datt B, McVicar TR, Niel van TG, Jupp DLB, Pearlman JS (2003). Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. IEEE Trans Geosci Remote Sens 2003;43:1246-1259.
[4] Xie Y, Sha Z, Yu M. Remote sensing imagery in vegetation mapping: a review. J Plant Ecol 2008;1:9-23.
[5] Gupta RP. Remote Sensing Geology. 2nd ed. Berlin and Heidelberg: Springer-Verlag; 2003.
[6] Gutierrez M, Norton R, Thorp KR, Wang G. Association of spectral reflectance indices with plant growth and Lint yield in upland Cotton. Crop Sci 2012;52:849-857.
[7] Stagakis S, Markos N, Sykioti O, Kyparissis A. Monitoring canopy biophysical and biochemical parameters in ecosystem scale using satellite hyperspectral imagery: an application on a Phlomis fruticosa Mediterranean ecosystem using multiangular CHRIS/PROBA observations. Remote Sens Environ 2010;114:977-994.
[8] Strauss RE. Discriminating groups of organisms. In: Elewa AMT, editor. Morphometrics for Nonmorphometricians, lecture notes in earth sciences, Berlin: Springer-Verlag; 2010, p. 73-91.
[9] Jongman RHG, Braak ter CJF, Tongeren van OFR. Data Analysis in Community and Landscape Ecology. Cambridge: Cambridge University Press; 1995.
[10] McDowall RM, Hewitt J. Attempts to distinguish morphotypes of the Canterbury–Otago non-migratory *Galaxias* species complex. 1st ed. Wellington: Doc Science Internal Series; 2004.
[11] Mircea G, Pirtea M, Neamtu M, Bazavan S. Risk software application using a credit scoring model. Int J Appl Math Inform 2012;6:1-8.
[12] Mattauer M. Étude géologique de l'Ouarsenis oriental. Alger: Service de la carte géologique de l'Algérie; 1958.

[13] Capolini J, Pieyns S, Ramana R. Bassin versant du Cheliff. Alger: Direction de l'hydraulique; 1969.
[14] Barry P. EO-1 / Hyperion science data user's guide, level 1-B, TRW Space, Defense and Information Systems; 2001.
[15] Lee JB, Woodyatt AS, Berman M. Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. IEEE Trans Geosci Remote Sens 1990;28:295-304.
[16] Moutinho LAM, Hutcheson GD. The SAGE dictionary of quantitative management research, London: Sage Publications; 2011.
[17] Rubin A. Statistics for evidence-based practice and evaluation. 2nd ed. Belmont CA: Brooks Cole; 2010.
[18] Landis JR, Koch GC. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-179.
[19] Cramer D, Howitt DL. The Sage dictionary of statistics. London: Sage Publications; 2004.
[20] Hill T, Lewicki P. Statistics: Methods and applications: a Comprehensive reference for science, industry, and data mining. 1st ed. Tulsa, OK: Statsoft Inc; 2005.
[21] Verrelst J, Koetz B, Kneubuhler M, Schaepman M. Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data. Remote Sens Environ 2008;112:2341-2353.
[22] Colombo R, Meroni M, Marchesi A, Busetto L, Rossini M, Giardino C, Panigada C. Estimation of leaf and canopy water content in poplar plantations by means of hyperspectral indices and inverse modelling. Remote Sens Environ 2008;112:1820-1834.