Abstract. Discrimination and classification are integral processes for interpreting remotely sensed data. Many spectral vegetation indices have been proposed for discriminating between vegetation, soil, and other ground cover categories. Classical remote sensing show that reflectance in the red (R) and near infrared (NIR) bands of the electromagnetic spectrum have been successful in differentiating between vegetation and other ground cover classes and they are commonly used for this purpose. Here we demonstrate how Fisher’s classical statistics can be applied to develop discriminant functions for commonly used vegetation indices simply using the R and NIR bands. We derive a new vegetation index, the Log-Ratio Vegetation Index (LRVI) and demonstrate its utility in discriminating between cork oak trees and surrounding background in woodlands in Portugal. The LRVI performed better than seven previously developed vegetation indices, likely because of its linear properties in the reflectance density spectral space. The robustness and simplicity of LRVI suggests that it deserves further exploration and should be included for comparison with other vegetation indices and functions in discrimination, classification, and modelling studies. We suggest that the demonstrated approach is widely applicable to development of indices composed of other bands than R and NIR for systems or processes that correlate better with reflectance in other regions of the electromagnetic spectrum.

Key words: Cork oak woodlands; Quercus suber; discriminant functions, vegetation indices, remote sensing

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Um Índice de Vegetação, o Log-Ratio Vegetation Index (LRVI), para a Discriminação e a Classificação de Dados de Deteção Remota

Sumário. Discriminação e classificação são processos integrantes da interpretação de dados de deteção remota. Muitos índices espectrais de vegetação têm sido propostos para a discriminação entre vegetação, solo e outras categorias de ocupação do solo. A deteção remota clássica mostra que as refletâncias nas bandas do vermelho (R) e infravermelho (NIR) do espectro eletromagnético têm sido bem-sucedidas na diferenciação entre vegetação e outras classes de ocupação do solo e são frequentemente usadas com esta finalidade. Utilizando apenas as bandas R e NIR, demonstramos como a estatística clássica de Fisher pode ser aplicada para desenvolver funções discriminantes para índices de vegetação comumente usados. Derivamos um novo índice de vegetação, o Log-Ratio Vegetation Index (LRVI), e demonstramos a sua utilidade na discriminação entre as copas de sobreiro e o sob ombro envolvente em montados de sobreiro em Portugal. O LRVI teve um desempenho melhor que sete índices de vegetação desenvolvidos anteriormente, provavelmente devido às suas propriedades lineares no espaço espectral das densidades de reflectância. A robustez e simplicidade do LRVI sugerem que merece ser mais estudado e deve ser incluído para comparação com outros índices de vegetação e funções em estudos de discriminação, classificação e modelação. Sugerimos que a abordagem demonstrada seja amplamente aplicável ao desenvolvimento de índices compostos por outras bandas além de R e NIR para sistemas ou processos que se correlacionem melhor com a reflectância noutras regiões do espectro eletromagnético.

Palavras-chave: Montados de sobreiro; Quercus suber; funções discriminantes; índices de vegetação; deteção remota

Un Indice de Végétation, Log-Ratio Vegetation Index (LRVI), pour la Discrimination et la Classification de Données de Télédétection.

Résumé. La discrimination et la classification font partie intégrante des processus d’interprétation des données de télédétection. De nombreux indices spectraux de végétation ont été proposés pour différencier la végétation, le sol et d’autres catégories d’occupation du sol. La télédétection classique montre que la réflectance dans les bandes rouge (R) et proche infrarouge (NIR) du spectre électromagnétique a réussi à différencier la végétation d’autres types de classes d’occupation du sol et elles sont couramment utilisées à cette fin. Ici, nous montrons comment la statistique classique de Fisher peut être appliquée pour développer des fonctions discriminantes pour les indices de végétation couramment utilisés en utilisant simplement les bandes R et NIR. Nous dérivons un nouvel indice de végétation, le Log-Ratio Vegetation Index (LRVI) et démontrons son utilité dans la discrimination entre les couronnes de chêne-liège et le sous-bois environnant dans des peuplements de chêne-liège du Portugal. Le LRVI a donné de meilleurs résultats que sept indices de végétation développés précédemment, probablement en raison de ses propriétés linéaires dans l’espace spectral des densités de...
A Log-Ratio Vegetation Index (LRVI) 47

réflectance. La robustesse et la simplicité du LRVI suggèrent qu’il mérite une exploration plus approfondie et devrait être inclus pour comparaison avec d’autres indices et fonctions de végétation dans les études de discrimination, de classification et de modélisation. Nous suggérons que l’approche démontrée est largement applicable au développement d’indices composés d’autres bandes que R et NIR pour des systèmes ou des processus qui correspondent mieux à la réflectance dans d’autres régions du spectre électromagnétique.

Mots-clés: Peuplements de chêne-liège; *Quercus suber*; fonctions discriminantes; indices de végétation; télédétection
Introduction

Discrimination and classification of remote sensing data

Discrimination and classification are related processes often complementary but different in their objectives. Discrimination means making a distinction between objects while classification means systematically arranging similar objects into groups or classes (HAWKINS et al., 1991). These two processes often complement each other in the analysis and interpretation of remote sensing data in relation to vegetation and soil. Vegetation indices (VIs) are mathematical combinations of spectral bands, mostly in the visible and near infrared regions of the electromagnetic spectrum (WU, 2014), that take advantage of the unique vegetation spectral signature in this region, namely strong absorption in the visible domain and very high near infrared reflectance (HUETE, 2012). They have been widely used for quantitative and qualitative evaluations of vegetation cover, vigor, and growth dynamics, among other applications (XUE and SU, 2017). Although VIs are not intrinsic physical quantities, they are optical measures of vegetation canopy "greenness", a composite property of leaf chlorophyll, leaf area, canopy cover, and canopy architecture and therefore they are frequently used to establish functions that relate their value with continuous biophysical and biochemical variables, including canopy chlorophyll content, leaf area index (LAI), green vegetation fraction, gross primary productivity (GPP), and fraction of photosynthetically active radiation absorbed by the vegetation (FAPAR) (JIANG et al., 2008). Another different use of vegetation indices has been in classification processes for vegetation or land cover mapping (see as examples BENEDETTI et al., 1994; DECHKA et al., 2002; WANG et al., 2015). Image classification, in a broad sense, can be defined as the process of extracting differentiated classes or themes (XIE et al., 2006). When these vegetation indices are used to discriminate between two classes, for example forest against non-forest (LUCAS and BUNTING, 2006), woody vegetation (trees and shrubs) against non-woody vegetation (KUERMERLE et al., 2006) or tree cover against surrounding background in a very high spatial resolution image (see STRAND et al., 2008 and also this work) they can be seen as discriminant functions and their discrimination and classification value assessed by the resulting separation of the two groups.

Two main approaches exist in developing vegetation indices, namely the empirical-statistical approach, where different functions using different bands are statistically evaluated for their relationships with known soil and vegetation
properties and undesired external influences (see as examples TUCKER, 1979; HUETE, 1988; BARET et al., 1993; HABOUDANE et al., 2004; WU, 2014), and the physically based approach based on radiative transfer theory (see as examples JIN and EKLUNDH, 2014; GOEL and QIN, 1994).

Early work on this subject such as those of JORDAN (1969) or TUCKER (1979) indicated that the combination of data from the near infrared and the red bands were successful in predicting important vegetation characteristics, for example green leaf area or biomass. Many other studies since have confirmed these early findings and as these bands are available for the major part of remote sensing sensors the two-band vegetation indices based only on near infrared and red bands are most often used (WU, 2014; JIANG et al., 2006; BARET and GUYOT, 1991). As a result, this paper will only use reflectance values in the near infrared (NIR) and red (R) bands. Furthermore, hyperspectral or specific narrowband VIs in these wavelength regions are also outside the scope of this paper. Simplicity of formulation and wider use of the indices are among the criteria for the choice of the vegetation indices analysed and presented in this work.

The statistical approach to developing the discriminant function

Statistical approaches following the pioneer work of FISHER (1936) can be applied to develop a discriminant function that maximizes the separation between classes of vegetation or other ground cover classes. In statistics, discrimination means finding a transform function that converts an original set of variables into a single discriminant variable that produces the maximum difference between two previously defined groups representing a method of collapsing a multivariate problem down into a univariate problem (DAVIS, 1986). In the first paper on discriminant functions FISHER (1936) applied these principles to four measurements on flowers of plants of two species of *Iris*. Fisher estimated the linear function (or compound measurement) which best discriminated between populations of the two species by maximizing the ratio of the difference between the specific means to the standard deviations within the species.

Since the original work by FISHER (1936), many other authors have proposed linear combinations of variables as discriminant functions, and linear discriminant analysis is a popular option in most statistical packages. However, this analysis has the limitation of implying that the best discriminant function is a linear combination of the original variables which is not necessarily always the
case. The methods for calculating the linear discriminant function are presented in most statistics books applied to a variety of fields, for example geology (DAVIS, 1986) or ecology (LUDWIG and REYNOLDS, 1988).

In the case of using reflectance data of two bands (near infrared and red) the calculations start by computing the means (m) of the two groups (1 and 2) in the original variables NIR and R (corresponding to the two bands) and their differences (d) as:

\[
d_{\text{NIR}} = m_{1 \text{NIR}} - m_{2 \text{NIR}}
\]
\[
d_{R} = m_{1 R} - m_{2 R}
\]

The second step is to compute the variances (s_{\text{NIR}}^2 and s_{R}^2) and covariance (s_{\text{NIR} R}) for each of the two groups:

s_{1\text{NIR}}^2, s_{1R}^2, s_{1\text{NIR} R}

s_{2\text{NIR}}^2, s_{2R}^2, s_{2\text{NIR} R}

If we accept that the variances and covariances of the two groups can be combined, we can compute a pooled estimate of the variances for NIR and R (s_{p\text{NIR}}^2 and s_{pR}^2) and a pooled covariance (s_{p\text{NIR} R}) as the average of the variances and covariance obtained for each group weighted by the corresponding degrees of freedom. The assumption that the variances and covariance of the groups can be considered equal and combined is often not tested or not possible to accept by statistical tests. However, departure from this assumption can have important consequences for the results. The assumption that the two groups have equal variances and covariances implies that the point along the discriminant function line which is exactly halfway between the centres of the two groups can be used as the threshold for classification (DAVIS, 1986).

We can now compute the coefficients of the linear discriminant function (here termed as Discriminant Function Vegetation Index - DFVI) centred at the midpoint between the averages of the two groups along the discriminant function:

\[
\text{DFVI} = c_{\text{NIR}} \text{NIR} + c_{R} \text{R} - Z_{0}
\]

where:

\[
c_{\text{NIR}} = (s_{pR}^2 d_{\text{NIR}} - s_{p\text{NIR} R} d_{R}) / (s_{p\text{NIR}}^2 s_{pR}^2 - s_{p\text{NIR} R}^2)
\]
\[
c_{R} = (s_{p\text{NIR}}^2 d_{R} - s_{p\text{NIR} R} d_{\text{NIR}}) / (s_{p\text{NIR}}^2 s_{pR}^2 - s_{p\text{NIR} R}^2)
\]
\[
Z_{0} = (c_{\text{NIR}} m_{1 \text{NIR}} + c_{R} m_{1 R} + c_{\text{NIR}} m_{2 \text{NIR}} + c_{R} m_{2 R}) / 2
\]

For DFVI = 0, this equation can be solved to give the line of separation between the two groups in the NIR-R space:

\[
\text{NIR} = Z_{0} / c_{\text{NIR}} - c_{R} / c_{\text{NIR}} \text{R}
\]
Various types of transformation functions have been proposed by combining data from two or more spectral bands from remote sensing data into indices developed to extract the plant signal only. These vegetation indices can be estimated in many ways, for example by calculating ratios, differences, ratios of differences and sums, and by linear combinations of spectral data.

Vegetation indices can be seen as discriminant functions that are not only derived statistically but by combining the NIR and R bands in different ways. One of the first indices to be proposed (JORDAN, 1969) was based on the differences between near infrared and red data, the Difference Vegetation Index (DVI), and it can be considered a special case of a linear discriminant function where:

\[ \text{DVI} = \text{NIR} - \text{R} \]  

Another vegetation index that is equivalent to a linear discriminant function is the Perpendicular Vegetation Index (PVI) proposed by RICHARDSON and WIEGAND (1977):

\[ \text{PVI} = \frac{(\text{NIR} - a - b \, \text{R})}{(1 + b^2)^{1/2}} \]  

representing, in the NIR-R space, the distance to a line of equation \( a + b \, \text{R} \), the soil line (the line with zero vegetation), giving a measure of vegetation density (\( a \) and \( b \) are the slope and intercept of the soil line, respectively).

Another common approach is using ratios instead of differences. This was originally proposed by PEARSON and MILLER (1972) as a Ratio Vegetation Index (RVI) computed as:

\[ \text{RVI} = \frac{\text{NIR}}{\text{R}} \]  

The first and very popular combination of ratios and differences was proposed by ROUSE et al. (1974) with their Normalized Difference Vegetation Index (NDVI):

\[ \text{NDVI} = \frac{\text{NIR} - \text{R}}{\text{NIR} + \text{R}} \]  

Since then many other equations using differences and ratios with NIR and R have been proposed such as the Soil Adjusted Vegetation Index (SAVI) proposed by HUETE (1988) and developed from NDVI to take into account soil background influences:

\[ \text{SAVI} = 1.5 \frac{(\text{NIR} - \text{R})}{(\text{NIR} + \text{R} + 0.5)} \]  

Many other vegetation indices of the SAVI family have been since proposed referred to as being Optimized (OSAVI) (RONDEAUX et al., 1996), Modified (MSAVI) (QI et al., 1994), Transformed (TSAVI) (BARET et al., 1989) or Generalized (GESAVI). This latter index is computed according to GILABERT et al. (2002) as:


GESAVI = (NIR - a - b R) / (R + 0.35)  

(13)
showing similarities with PVI as there is also a reference line of equation a + b R.

For these two indices the values of a and b can be related to soil characteristics as they follow the 'soil line' concept, which considers that all soils can be represented by a 'global' soil line in NIR-R space (a reference soil line) and different authors have been searching for a general equation with variable success. In fact, HUETE et al. (1984) demonstrated that the "global" soil line concept does not apply. This concept is only valid in the visible-near infrared domain for a given soil type with varying conditions in only one of the soil variables (e.g. surface roughness, surface moisture, organic content, mineral composition, etc.) (BARET et al., 1993). Therefore, in many studies, in order to achieve better results, a different soil line equation for each situation is estimated statistically (GILABERT et al., 2002).

Apart from these described vegetation indices we propose to test a different index, similar to the Ratio Vegetation Index (RVI), but using reflectance density instead of reflectance. Reflectance density is also a measure of the extent to which a surface reflects radiation and it is computed as the logarithm to base 10 of the reciprocal of the reflectance similarly to what is described for transmittance in classical remote sensing books (LILLESAND and KIEFER, 1979). Our hypothesis is that the relationship between densities can be more adequately assumed to be linear compared to the relationship between NIR and R reflectance values. We then have:

\[
\begin{align*}
DNIR & = - \log_{10} (NIR) \\
DR & = - \log_{10} (R)
\end{align*}
\]

(14)  (15)

where DNIR and DR are the reflectance densities corresponding to reflectances of the NIR and R bands.

We propose a Log-Ratio Vegetation Index (LRVI) based on a ratio of densities as:

\[
LRVI = DNIR / DR = - \log_{10} (NIR) / - \log_{10} (R)
\]

(16)

for NIR € ] 0, 1 [ and R € ] 0, 1 [.

We demonstrate the utility of the proposed LRVI for distinguishing between trees and surrounding background in cork oak (Quercus suber) woodlands in Portugal (Figure 1), using a very high spatial resolution image (2.4 m).
Figure 1 - Photographs of cork oak (*Quercus suber*) woodlands with a herbaceous understory, east of Lisbon, Portugal. In photo (a) the photographer is facing the woodlands while photo (b) gives an overhead perspective. Photo credit Nuno Onofre
Methods

The evaluation of the discriminant value of a function

Since the work of Fisher, the evaluation of the discriminant value of any function is always assessed by the ratio of the difference between the mean to the standard deviation within each group (Fisher, 1936). This can be done simply by calculating a Student's t value with no assumption of the equality of variances (Steel and Torrie, 1980) as:

\[ t = \frac{(m_1 - m_2)}{(s_{12}^2 / n_1 + s_{22}^2 / n_2)^{1/2}} \]  \hspace{1cm} (17)

where

- \( m_1 \) and \( m_2 \) are the averages of the two groups,
- \( s_{12}^2 \) and \( s_{22}^2 \) are the variances, and
- \( n_1 \) and \( n_2 \) are the sizes of the two groups.

This t value can then be used as the measure of the discriminant value of any variable (original or combined) to distinguish between two groups.

When the variances of the two groups are not statistically different, a pooled estimate of the variance (\( s_p^2 \)) is often used instead of \( s_{12}^2 \) and \( s_{22}^2 \) and computed as the average of the two variances weighted by the corresponding degrees of freedom:

\[ s_p^2 = \frac{[(n_1-1) s_{12}^2 + (n_2-1) s_{22}^2]}{n_1 + n_2 - 2} \] \hspace{1cm} (18)

In this case it is also common to use the F statistic instead of the t value as:

\[ F = \frac{(m_1 - m_2)^2}{(s_p^2 / n_1 + s_p^2 / n_2)} \] \hspace{1cm} (19)

This F value will be used in this paper for the comparison of the discriminating value of the different functions.

The evaluation of the classification value of a function

Classification starts when the discriminant function developed is used to allocate new objects to one of the original groups (Davis, 1986). Typically, these two different steps (developing the discriminant function and using it for classification) are involved in the same analysis and therefore their different objectives and scope are often overlooked. Sometimes these two steps are clearly distinct. For instance, in the pioneer paper of Fisher, when the author noticed some overlap in the distribution of two species of Iris in the discriminating linear function, he indicated "that a certain diagnosis of these two species could not be based solely on these four measurements of a single flowers taken on a
A Log-Ratio Vegetation Index (LRVI)

plant growing wild" (FISHER, 1936) avoiding entering into the field of classification.

In any case, the classification process associated with a discriminant function requires a specific approach. First, it is necessary to set up criteria to evaluate the results of the classification, and those criteria can then be used to determine the optimum threshold (or cut-off point) in the value of the function to classify any object into a class.

The main objective in classification is to minimize two types of error, often termed errors of commission and omission or type I and type II. These errors can be quantified in our case as the proportion \( p_{12} \) of objects (pixels from remote sensing data) belonging to class 1 that are misclassified as the class 2 and the proportion \( p_{21} \) of pixels belonging to class 2 that are misclassified as class 1. When the two types of error are considered to be equivalent, one reasonable objective is to minimize the average of \( p_{12} \) and \( p_{21} \). We can display the two-way cross-classification table and the corresponding symbols (Quadro 1).

Using the symbols of Table 1 we can compute the average error:

\[
\text{Average error (AE)} = \frac{p_{12} + p_{21}}{2} = \frac{b}{a+b} + \frac{c}{c+d} \]

\[(20)\]

Quadro 1 - Example of a cross-classification table for two classes. The table describes correct classifications of Class 1 and 2 respectively as a and d. Incorrect classification of class 2 as class 1 is represented by b and the reverse is represented by c.

| True Class | Classification Result |
|------------|-----------------------|
|            | Class 1 | Class 2 |
| Class 1    | a       | b       |
| Class 2    | c       | d       |

Distinguishing tree cover from background in cork oak formations in Portugal

Cork oak (Quercus suber) occupies more than 700 thousand hectares in Portugal, about 30% of the species worldwide producing around 150 million kg of cork annually with a total value of around 400 million euros (DGRF, 2007). Typical open cork oak stands (Figure 1) are of extreme importance for many wildlife species and ecosystem services.

The research work of CADIMA (2008) was dedicated to finding methods to discriminate cover of cork oak canopies from the surrounding background.
using two very high resolution images from QuickBird obtained in the summers of 2003 and 2006 on an 80 km² area (8°05'05" - 8°00'47"W; 39°04'05" - 38°57'19"N) to the east of Lisbon, northeast of Cabeção village, Portugal. The area of the images is mainly occupied by forests (~80%) from which around 73% are cork oak trees in open and spatially irregular stands that vary from around 38 to 260 trees per ha (CADIMA, 2001). The understory shrub and herbaceous vegetation of the cork oak stands is frequently used as pasture for sheep and cattle. Main shrub communities in the area are gorse scrubs, either *Thymo capitellati-Stauracanthetum genistoides* or *Thymo villosi-Ulicetum airensis*. In the herbaceous layer we can find the following main plant communities, among others: *Anthyllido hamosae-Malcomion lacerae* and *Chamaemelo mixti-Vulprietum alopecuroris* (CAPELO and MESQUITA, 2001).

The area is relatively flat with a moderate ondulated topography. Elevation above sea level varies between 77.5 m and 191.0 m with a mean elevation of 132.5 m. The region is characterized by a Mediterranean climate (Köppen-Geiger Climate Classification subtype: Csa (KÖPPEN-GEIGER, 1928)) with cold and rainy winters and hot and dry summers. Mean annual precipitation is around 650 mm. Minimum temperature of the coldest month (January) around -4.7°C and the maximum temperature of the hottest month (July-August) is around 43°C.

Geologically the area is covered predominantly by sedimentary rocks of sandstone and clay of the lower miocene and pleistocene but also by sandstone and conglomerates from the plio-pleistocene in some areas (ALEXANDRE et al., 2002). In cork oak areas of the images the predominant soil groups of the World Reference Base for Soil Resources (F.A.O., 1998) are Cambissols, but also Arenosols and Regosols (ALEXANDRE and GOMES, 2000).

The images were orthorectified and reflectance data obtained after atmospheric corrections with the ATCOR2 module of PCI Geomatics, *Geomatica* v.10 software. The multispectral images (with four bands and 2.4 m resolution) were fused with the corresponding panchromatic images (one band with 0.6 m resolution). These pan-sharpened images were used to manually delineate the tree crowns within 30 circular sample plots already studied in the field. Plots that were under cloud cover were discarded. The pixels of the atmospherically corrected multispectral images (2.4 m resolution) that were inside the boundaries of crowns were then classified as ‘crowns’ and those outside classified as ‘background’ (including bare ground, dried grasses and small shrubs not resolved by the pixel spatial resolution). This process resulted for the image of 2003 in a total of 4648 pixels classified as tree cover and 9674 pixels
classified as background. The same procedure was applied independently for the image of 2006.

The process described allowed for the calculation of the previously described vegetation indices, as well as all the parameters necessary (optimal thresholds, F-statistics, soil line parameters, etc.) to discriminate between tree crowns and background. In order to compare the different functions in their discriminant and classification value, we used the thresholds (and the slope and intercept coefficients of the reference lines) that resulted in the maximum F statistic and the minimum average error (AE).

The best values of F and AE were then used to compare the results of the different discriminant functions for the derived indices DVI, RVI, PVI, NDVI, SAVI, and GESAVI, for the statistically derived linear discriminant function (DFVI) and for the proposed Log-Ratio Vegetation Index (LRVI) for the images of the year 2003. LRVI was also computed independently for the year 2006.

**Results and discussion**

**Discrimination and classification**

The visual inspection of the frequency distribution of the two groups (crown and background) along the discriminant functions is possible by the analysis of Figure 2 and the corresponding quantitative results shown in Table 2.

We conclude that the reflectance data analysed individually for both near infrared and red bands are the least accurate, especially for NIR where the two groups largely overlap (Figure 2 and Table 2). The frequency distributions of the two groups in R space already show some separation of the two groups that exhibit frequency curves with very different shapes however with much overlap. These results are consistent with the expected spectral response of the two groups in these spectral bands, namely lower reflectance in the red band for pixels with photosynthetic active components ("crown pixels") and higher reflectance for the "background pixels", since in the summer, these pixels are mainly occupied by bare soils and senescent vegetation in different proportions (CADIMA, 2008). In fact, the spectra of these two last components (soil and senescent vegetation) is very similar in the visible-near infrared (VNIR) region of the electromagnetic spectrum for broadband sensors (NAGLER et al., 2000; OKIN, 2007). In the near infrared it was expected that both groups (crows and
background) would have higher and more similar reflectance than in the red wavelength region.

**Figure 2** - Frequency distributions of pixels of the 2003 image corresponding to tree crowns (green) and to background (grey) for individual band reflectance data (NIR and R), for various vegetation indices and functions in the literature (DVI, RVI, DFVI, NDVI, SAVI, GESAVI, PVI) and for the proposed Log-Ratio Vegetation Index (LRVI). Note that the axis for LRVI is in reverse order to facilitate comparison.
Combination of the NIR and R bands by utilizing differences (DVI) versus ratios (RVI) and by statistical linear combination (DFVI) were able to separate the two groups better but the frequency distribution of the two groups is still very different and similar (inversely) to what was observed in R (Figure 2). These indices produce intermediate results for the value of F and for the average error (Table 2). The normalisation produced by the NDVI results in more similar frequency distributions and better discrimination of the two groups (higher F value) but still with a degree of overlap that does not allow for the classification to be more successful than that of RVI. In fact, PERRY and LAUTENSCHLAGER (1984) demonstrated that, from the mathematical point of view, these two indices are functionally equivalent and contain the same information, since they can be transformed one into the other. The authors therefore concluded that, for making a set of decisions, the decisions made on the basis of one index could have been equally well made on the basis of the other. However, the normalization in NDVI made it more sensitive to sparse vegetation densities than RVI, but less sensitive to high vegetation densities (JACKSON and HUETE, 1991; LIANG, 2004). This can also be observed in our data (Figure 2) when comparing the change of the form of the histogram for background (sparse vegetation) and crown (higher vegetation densities) from RVI to NDVI, namely a wider spread of values for the background histogram and a concentration of

Table 2 - Discriminant and classification results for original reflectance data for the two bands (NIR and R), for different vegetation indices in the literature (DVI, RVI, NDVI, SAVI, GESAVI, PVI), for the statistically derived linear discriminant function (DFVI) and for the proposed Log-Ratio Vegetation Index (LRVI) for the images of the year 2003. LRVI is also computed independently for the year 2006.

| Vegetation Index | F-value | Threshold | Average Error (%) |
|------------------|---------|-----------|-------------------|
| NIR              | 424.9   | 0.271     | 43.3              |
| R                | 6217.3  | 0.100     | 21.0              |
| DVI              | 19506.0 | 0.138     | 12.1              |
| DFVI             | 20222.6 | 0.000     | 10.1              |
| RVI              | 25646.6 | 0.482     | 9.7               |
| NDVI             | 27054.2 | 0.414     | 9.7               |
| SAVI             | 30000.3 | 0.239     | 8.9               |
| GESAVI           | 30561.4 | 0.224     | 8.2               |
| PVI              | 28809.6 | 0.000     | 8.2               |
| LRVI (2003)      | 31897.7 | 0.623     | 8.2               |
| LRVI (2006)      | 37161.8 | 0.596     | 7.6               |
values for the crown histogram, resulting in more similar frequency distributions (Figure 2).

The better performance of the ratio based indices (in this case RVI and NDVI, but also SAVI and PVI) over indices that are linear combinations of bands (DVI and DFVI) have been attributed by several authors to the attenuating effect that the ratio of bands have on the effect that variable illumination have on surface reflectance (or shadowing) due to topography, but also to soil roughness or vegetation components (BARET and GUYOT, 1991; LAWRENCE and RIPPLE, 1998; RICHARDS and JIA, 1999; AKKARTAL et al., 2004; SILLEOS et al., 2006). In our case, the effect of variable illumination due to topography cannot be considered of major concern since the terrain is relatively flat, but tree shadows are particular important, since we are working with a high spatial resolution image, where the average height of trees (around 9 m) is larger than the pixel size, and trees cast shadows on neighboring background pixels, but also on neighboring crown pixels which exert important influences on the reflectance in the visible and near-infrared of the two groups. Since the spatial resolution of the image is not fine enough relatively to the average dimension of the tree crowns, intra-crown shadowing effects are not as influential as they would be in higher spatial resolution imagery (a few centimeters), since they are averaged in the pixel reflectance value. The attenuating effect that the ratio of bands have on shadowing, could, at least in part, explain the better performance of the ratio based indices over the non-ratio based DVI and DFVI in our work.

Vegetation indices that are soil adjusted and that correspond to orthogonal or angular distances to reference soil lines (PVI, SAVI, GESAVI) produce very good results as shown by the higher F values and by the lower average error values (Table 2). SAVI and GESAVI show a relatively better discrimination of the groups than PVI, but the classification accuracy is not higher.

Soil adjusted vegetation indices are supposed to give better results than NDVI and RVI especially in partially vegetated environments of intermediate canopy covers with high reflective soil backgrounds (HUETE and JACKSON, 1987; HUETE, 1988; ELVIDGE and CHEN, 1995), typical for semi-arid environments, as in our study, where the gaps between vegetation elements (horizontally) or the gaps between low layers of green components in the vertical sense (intermediate LAI values) permit the surrounding and background soil, or other non-photosynthetic elements like trunk, stems, branches or senescent vegetation, to have a strong influence on the reflectance received by the sensor. One drawback with those indices and in the case of PVI and GESAVI is that the coefficients of the reference lines have to be determined statistically,
while in the case of SAVI the formulation of the index introduces a coefficient L in the NDVI equation, that should vary with vegetation density, from L=0 for very high vegetation cover to L=1 for very low vegetation cover. However, since the vegetation density is seldom known and can be spatially variable, it is difficult to optimize this index (LIANG, 2004) and L=0.5 for intermediate vegetation amounts is used most often.

Finally, the proposed Log-Ratio Vegetation Index (LRVI) performed very well. The frequency distributions of the two groups have similar shapes (Figure 2), and the index produced the higher F value and an average error value equal to the best (GESAVI and PVI) (Table 2). The same analysis using data from 2006 resulted in similar conclusions. It is also important to note that the threshold value for LRVI in 2003 and 2006 is very similar (around 0.61) indicating that this index might be robust enough to be applied to multi-temporal data. Figure 3 illustrates the classification for a subset of the 2006 image (643.2 m x 427.2 m).

The assumption of linearity in the NIR-R space

In this study we hypothesized that linearity of index isolines was probably more adequate in the space of reflectance densities than that of original reflectance data in the NIR and R space. A visual inspection to the pattern of distribution of pixels of crown and background in the original space and in the log-transformed space seems to support this hypothesis (Figure 4a and b).

The vegetation indices analyzed in this work (with the exception of the proposed LRVI), are all represented by linear index isolines in the NIR-R reflectance space. Index isolines are the lines formed by points of constant values of the index and each vegetation index has its own index isolines, that are artificial lines obtained independently of canopy optical properties (YOSHIOKA et al., 2000). These lines have fixed behavior in NIR-R space for each VI and an illustration of these isolines for each index can be found in QI et al. (1994), JIANG et al. (2007) and GILABERT et al. (2002) for GESAVI. In our work we are concerned with the problem of finding the line for each of the analyzed VIs that best discriminates and simultaneously better classifies (above and below the line) the overall cloud of points of each group in NIR-R space (Figure 4a).
Figure 3 - Subset of the 2006 QuickBird image (643.2 m x 427.2 m) showing cork oak stands: a. Pansharpened image (0.6 m spatial resolution), false color composition RGB=NRG; b. Multispectral image (2.4 m spatial resolution), false color composition RGB=NRG; c. Crown pixels classified by an LRVI < 0.61 in green over the image b
Figure 4 - Pixels corresponding to tree crowns (green) and to background (black) for the 2003 image. a., in the NIR-R reflectance space and b., in the logarithmic space, showing the optimal crown-background threshold line (LRVI = 0.623) and the approximated soil line (LRVI = 0.8). In graph a., we also show the optimal Fisher's discriminant function isoline (DFVI = 0). Note that the axis in the logarithmic space are inverted to facilitate comparisons.
Apart from external factors such as atmospheric condition (MYNENI and ASRAR, 1994), solar and viewing geometries (GOEL and QIN, 1994), spatial and spectral resolution of the sensor (TEILLET, STAENZ and WILLIAMS, 1997), the distribution of these points in the NIR-R space, in other words, the reflectance of the pixels in these spectral bands, are influenced not only by the optical properties of green vegetation elements (reflectance and transmittance), canopy structure (height, shape, size orientation and spatial arrangement of vegetation elements), but also by the optical properties of surface cover and surrounding non-photosynthetic elements like soil, stems, trunk, branches and senescent vegetation. Due to multiple scattering between at least two of these elements the spectral mixing process is generally nonlinear (HUETE et al., 1985). The vertical structure of the canopy and the spatial distribution of the plants play an important role in nonlinearity (JI et al., 2017). Furthermore, the amplitude of multiple photon scattering is wavelength-dependent, being more important in the near-infrared than in the visible part of the spectrum, due to high leaf transmittance in the near-infrared and high absorption in the visible (CHEN and VIERLING, 2006; SOMMERS et al., 2009).

These multiple physical processes often result in a variety of displacements of the pixel spectral position in NIR-R reflectance space, and most of the time, these displacements are neither collinear nor orthogonal to each other (VERSTRAETE and PINTY, 1966). These authors showed that vegetation indexes can be made sensitive or insensitive at will to various processes by manipulating their mathematical definition. Graphically, in the NIR-R space, "this corresponds to positioning, orienting, and bending the isolines of the index to achieve the desired degree of sensitivity", however they also showed that it is not possible to define a vegetation index sensitive to only one process or to design a universal vegetation index which would be optimal for all applications.

VIs relationships with vegetation biophysical quantities (such as Leaf Area Index (LAI), fraction of green cover (FGC), biomass, etc.) are in general not linear (GAMON et al., 1995; HUETE et al. 2002; HABOUDANE et al. 2004; GLENN et al., 2008) and these nonlinearities are modeled à posteriori by establishing nonlinear relationships between them. Unlike these studies, in our work we are concerned with a one-step method of image classification, commonly known as 'thresholding'. Therefore, it is desirable that nonlinearities between VIs and ground targets are treated in the first step and a nonlinear VI in the NIR-R space, like LRVI, is preferable.

To simulate nonlinear relationships between VIs and surface biophysical parameters, non-linear indices have been proposed and some examples are NLI
A Log-Ratio Vegetation Index (LRVI)

(Non-Linear Index, GOEL and QIN, 1994), RDVI (Renormalized Difference Vegetation Index, ROUGEAN and BREON 1995) and MSR (Modified simple Ratio, CHEN, 1996). In our work we are not concerned with quantifying vegetation amounts or related biophysical quantities, but in distinguishing functionally and spectrally distinct groups or land cover themes that at sub-pixel scales contain an unknown quantity of vegetation and other ground elements, with some degree of variability. GLENN et al. (2008) argue that vegetation indices should not be used as surrogates for detailed features of canopy architecture (like LAI, leaf angle distribution (LAD) or FGC), but simply as a measurement of canopy light absorption or canopy light reflected. An index, like LRVI, that contrasts the target reflectance densities in the near infrared and red wavelengths is a step in that direction and seems to give good results for the classification of spectrally distinct thematic classes demonstrated by its linear properties in the reflectance density space.

Conclusions

The results of this study are obviously only strictly applicable to the situation studied. However, due to the similar findings for both years evaluated in this research, we believe that the assumption of linearity of the relationship between NIR and R reflectance data in this log-transformed space deserves further exploration. We propose this new Log-Ratio Vegetation Index (LRVI) as the ratio of log-transformed reflectance data of NIR and R (logarithms at any base) as a very simple index that does not require the development or the statistical estimation of reference lines. This index performed very well and seems robust for the situation analysed, allowing the use of a common threshold of 0.61 to distinguish between crowns and surrounding background in summer images of cork oak formations. The promising results observed and the simplicity of calculations suggests that this index should be included for comparison with other vegetation indices and functions in discrimination and classification studies. We further suggest that the statistical approach demonstrated in the development of LRVI can be applied to other band combinations and used in discrimination between ecosystem components.
Acknowledgements

The authors would like to thank Sandra Mesquita for the delineation of the tree crowns on the pansharpened images. The QuickBird images used in this work were funded by INIAV's research project PIDDAC 237/02.

References

AKKARTAL, A., TURUDU, O., ERBEK, F.S., 2004. Analysis of changes in vegetation biomass using multisensor satellite data Proceedings of the ISPRS Congress-Youth Forum Archives, XXXV Part B8: 181-185.
ALEXANDRE, C., GOMES, A.A., 2000. Relatório da Componente de Solos do Projeto PAMAF 4069 - Tipificação dos montados de sobre e azinho nos concelhos de Avis, Mora e Arraiolos e dos sistemas de exploração associados. Identificação dos principais problemas e medidas para a sua gestão integrada e conservação. UE, EFN, Évora, Lisboa, 11 pp + Anexos.
ALEXANDRE, C., GOMES, A.A., ROSADO, M.J., MESQUITA, S., CADIMA, I.S.P., REGO, F.C., 2002. Caracterização preliminar dos solos ocupados com montados de sobre e azinho nos concelhos de Avis, Mora e Arraiolos (Preliminary characterization of soils occupied with cork oak and holm oak stands in Avis, Mora and Arraiolos region), Revista de Ciências Agrárias, Edição Especial da SPCS, Vol. XXV, n°s 3 e 4: 271-279.
BARET, F., GUYOT, G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. Remote Sensing of Environment 35: 161-173.
BARET, F., GUYOT, G., MAJOR, D., 1989. TSAVI: A Vegetation Index Which Minimizes Soil Brightness Effects on LAI and APAR Estimation, Proceedings of the 1989 International Geoscience and Remote Sensing Symposium (IGARSS'89) and the Twelfth Canadian Symposium on Remote Sensing, Vancouver, Canada, pp. 1355-1358.
BARET, F., JAQUEMOUD, S., HANOCQ, J.F., 1993. About the soil line concept in remote sensing, Advances in Spatial Research 13(5): 281-284.
BENEDETTI, R., ROSSINI, P., TADDEI, R., 1994. Vegetation classification in the Middle Mediterranean area by satellite data. International Journal of Remote Sensing 15(3): 583-596.
CADIMA, I.S.P., 2001. Relatório Final da Componente 'Povoamento Florestal: Dendrometria e Sanidade'. In Relatório Final do projeto PAMAF 4069 - Tipificação dos montados de sobre e azinho nos concelhos de Avis, Mora e Arraiolos e dos sistemas de exploração associados. Identificação dos principais problemas e medidas para a sua gestão integrada e conservação, Vol. III, EFN, Lisboa pp. 78-181.
A Log-Ratio Vegetation Index (LRVI)

CAdima, I.S.P., 2008. Cartografia dos montados a partir de imagens multispectrais de muito alta resolução espacial: Exploração de uma metodologia. Trabalho de Tese apresentado como dissertação original para efeitos de acesso à categoria de Investigador Auxiliar, INIAP, EFN, Oeiras, 225 pp.

Capelo, J., Mesquita, S., 2001. Relatório Final da Componente ‘Vegetação Paisagem Vegetal e Clima’. In Relatório Final do projeto PAMAF 4069 - Tipificação dos montados de sobro e azinhol nos concelhos de Avis, Mora e Arraiolos e dos sistemas de exploração associados. Identificação dos principais problemas e medidas para a sua gestão integrada e conservação, Vol. III. EFN, Lisboa, pp. 5-77.

Chen, J., 1996. Evaluation of Vegetation Indices and modified simple ratio for boreal applications. Canadian Journal of Remote Sensing 22: 229-242.

Chen, X., Vierling, L., 2006. Spectral mixture analysis of hyperspectral data acquired using a tethered balloon. Remote Sensing of Environment 103: 338-350.

Davis, J.C., 1986. Statistics and Data Analysis in Geology (second edition), John Wiley and Sons, New York, 646 pp.

DGRF, 2007. National Forest Strategy. Portugal. Government Order No 114/2006. DGRF, Lisboa, 109 pp.

Elvidge, C.D., Chen, Z., 1995. Comparison of Broad-Band and Narrow-band Red and Near-Infrared Vegetation Indices. Remote Sensing of Environment 54: 38-48.

FAO., 1998. World Reference Base for Soil Resources. Food and Agriculture Organization of the United Nations. World Soil Resources Report No 84, Roma.

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics 7: 179-188. Cambridge University Press, Cambridge.

Gamon, J.A., Field, C.B., Goulden, M.L., Griffin, K.L., Hartley, A.E., Joel, G., Penuelas, J., Valentini, R., 1995. Relationships between NDVI, Canopy structure and photosynthesis in three Californian vegetation types. Ecological Applications 5(1): 28-41.

Gilabert, M.A., González-Piqueras, J., García-Haro, F.J., Meliá, J., 2002. A generalized soil-adjusted vegetation index. Remote Sensing of the Environment 82: 303-310.

Glenn, E.P., Huete, A.R., Nagler, P.L., Nelson, S.G., 2008. Relationship Between Remotely-sensed Vegetation Indices, Canopy Attributes and Plant Physiological Processes: What Vegetation Indices Can and Cannot Tell Us About the Landscape. Sensors 8: 2136-2160.

Goel, N.S., Qin, W., 1994. Influences of canopy architecture on relationships between various vegetation indices and LAI and FPAR: A computer simulation. Remote Sensing Reviews 10(4): 309-347.

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. Remote Sensing of Environment 90: 337-352.
HAWKINS, J.M., WESTON, J., SWANNELL, J.C. (compilers), 1991. *The Oxford Study Dictionary*, Oxford University Press. Oxford.

HUETE, A.R., POST, D.F., JACKSON, R.D., 1984. Soil Spectral Effects on 4-Space Vegetation Discrimination. *Remote Sensing of Environment* **15**: 155-165.

HUETE, A.R., JACKSON, R.D., POST, D.F., 1985. Spectral Response of a Plant Canopy with Different Soil backgrounds. *Remote Sensing of Environment* **17**: 37-53.

HUETE, A.R., JACKSON, R.D., 1987. Suitability of Spectral Indices for Evaluating Vegetation Characteristics on Arid Rangelands. *Remote Sensing of Environment* **23**: 213-232.

HUETE, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of the Environment* **25**: 295-309.

HUETE, A.R., 1988. Soil and atmosphere Influences on the Spectra of Partial Canopies. *Remote Sensing of Environment* **25**: 89-105.

HUETE, A.R., DIDAN, A.K., MIURA, T., RODRIGUEZ, E.P., GAO, X., FERREIRA, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* **83**: 195-213.

HUETE, A.R., 2012. Vegetation Indices, Remote Sensing and Forest Monitoring. *Geography Compass* **6**(9): 512-532.

JACKSON, R.D., HUETE, A.R., 1991. Interpreting vegetation indices. *Preventive Veterinary Medicine* **11**: 185-200.

JIN, C., JIA, Y., GAO, Z., WEI, H., LI, X., 2017. Nonlinear spectral mixture effects for photosynthetic/non-photosynthetic vegetation cover estimates of typical desert vegetation in western China. *PLoS ONE* **12**(12): e0189292. https://doi.org/10.1371/journal.pone.0189292.

JIANG, Z., HUETE, A.R., CHEN, J., CHEN, Y., LI, J., YAN, G., ZHANG, X., 2006. Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sensing of Environment* **101**: 366-378.

JIANG, Z., HUETE, A.R., LI, J., QI, J., 2007. Interpretation of the modified soil-adjusted vegetation index in red-NIR reflectance space. *Journal of Applied Remote Sensing* **1**(1), 013503. https://doi.org/10.1117/1.2709702.

JIANG, Z., HUETE, A.R., DIDAN, K., MIURA, T., 2008. Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment* **112**: 3833-3845.

JIN, H., EKLUNDH, L., 2014. A physically based vegetation index for improved monitoring of plant phenology. *Remote Sensing of Environment* **152**: 512-525.

JORDAN, C.F., 1969. Derivation of leaf area index from quality of light on the forest floor. *Ecology* **50**: 663-666.

KÖPPEN, W., GEIGER, R., 1928. *Klimate der Erde*. Gotha: Verlag Justus Perthes. Wall-map 150 cm x 200 cm.

KUEMMERLE, T., RÖDER, A., HILL, J., 2006. Separating grassland and shrub vegetation by multidate pixel-adaptive spectral mixture analysis. *International Journal of Remote Sensing* **27**(15): 3251-3271.
A Log-Ratio Vegetation Index (LRVI)
STEEL, R.G.D., TORRIE, J.H., 1980. *Principles and Procedures of Statistics. A Biometrical Approach* (second edition),” McGraw-Hill Book Company, New York, 672 pp.

STRAND, E.K., VIERLING, L.A., SMITH, A.M.S., BUNTING, S.C., 2008. Net changes in aboveground woody carbon stock in western juniper woodlands, 1946-1998. *Journal of Geophysical Research Biogeoosciences* 113, G01013. doi:10.1029/2007JG000544.

ROUGEAN, J.L., BREON, F. M., 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment* 51: 375-384.

TEILLET, P.M., STAENZ, K., WILLIAMS, D.J., 1997. Effects of Spectral, Spatial, and Radiometric Characteristics on Remote Sensing Vegetation Indices of Forested Regions. *Remote Sensing of Environment* 61: 139-149.

TUCKER, C.J., 1979. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of the Environment* 8: 127-150.

VERSTRAETE, M.M., PINTY, B., 1996. Designing optimal spectral indices for remote sensing applications. *IEEE Transactions on Geoscience and Remote Sensing* 34(5): 1254-1264.

WANG, J., ZHAO, Y., LI, C., YU, L., LIU, D., GONG, P., 2015. Mapping global land cover in 2001 and 2010 with spatial-temporal consistency at 250 m resolution. *ISPRS Journal of Photogrammetry and Remote Sensing* 103: 38-47.

WU, W., 2014. The Generalized Difference Vegetation Index (GDVI) for Dryland Characterization. *Remote Sensing* 6: 211-1233.

YOSHIOKA, H., MIURA, T., HUETE, A.R., GANAPOL, B.D., 2000. Analysis of Vegetation Isolines in Red-NIR Reflectance Space. *Remote Sensing of Environment* 74: 313-326.

XIE, Y., SHA, Z., YU, M., 2008. Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology* 1: 9-23.

XUE, J., SU, B., 2017. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, vol. 2017, Article ID 1353691, 17 pages. https://doi.org/10.1155/2017/1353691.