SPECTRAL BAND DISCRIMINATION FOR SPECIES OBSERVED FROM HYPERSPECTRAL REMOTE SENSING

N. Dudeni, P. Debba, M. Cho, R. Mathieu

The Council for Scientific and Industrial Research (CSIR)
P.O. Box 395, Pretoria, 0001, South Africa

ABSTRACT

In vegetation spectroscopy, compositional information of leaves contained at band level or across the electromagnetic spectrum (EMS) and parts thereof, plays a huge rule in the analysis of spectra and their relations to the reflectance patterns across the spectrum. Spectral matching is often achieved by means of matching algorithms such as the Spectral Angle Mapper (SAM), Spectral information divergence (SID) and mixed measures of SAM and SID using either the tangent or the sine trigonometric functions, SID(TAN) or SID(SIN). The performance of these measures in distinguishing between objects of interest, such as species, is often compared using the relative spectral discriminatory probability (RSDPB). In this study, these measures are used to assess whether various sets of bands including the full spectrum, the visible (VIS), the near infrared (NIR), the shortwave infra-red (SWIR) region, as well as sets of bands identified by the stepwise discriminant analysis (SDA), can be used to discriminate the different species. This is done to identify the important regions of the EMS to distinguish seven common savannah tree species observed in the Kruger National Park, South Africa's largest game reserve. The magnitude of variation of the species in any part of the spectrum can be linked to the importance of that spectral region in distinguishing the species. In addition, classification accuracy of these sets of bands was assessed and the SDA bands often gave better classification accuracy compared to using all bands, bands in the NIR, and SWIR parts of the EMS.

***Index Terms*—** hyperspectral data, species, discrimination, SAM, SID, mixed measures, classification.

1. INTRODUCTION

The compositional information of leaves in savannah trees and vegetation spectroscopy in general, obtained by means of hyperspectral remote sensing technologies, can provide good basis for spectral analysis of leaves as well as information about leaves contained across the electromagnetic spectrum (EMS) and parts thereof. The advancement in high spectral resolution data obtained through hyperspectral remote sensing makes it possible to distinguish spectrally similar species, but with many additional problems, for example, large number of bands and greater within species variability than between species variability.

A number of studies (those conducted by [1], [2], [3], and [4]) were able to show that specific parts of the hyperspectral EMS can be used to discriminate species, whereas studies such as the one conducted by [5] contradicted their findings as the different parts of EMS did not show any significant difference in species discrimination. This study assesses the ability of certain parts of the spectrum, as well as the important bands identified by the stepwise discriminant analysis (SDA), to contribute to the identification of each of the seven major savannah tree species found in the Kruger National Park, the largest game reserve in South Africa.

Common spectral similarity measures are used to compare the manner in which the band configurations such as the visible (VIS), near infra-red (NIR), shortwave infra-red (SWIR), full spectrum, as well as the best sets of bands selected from SDA (65, 30, 20, and 10 band sets respectively). The performance of these measures in distinguishing between the various sets of bands are compared by means of the relative spectral discriminatory probability (RSDPB), using each of the species in order to establish the association between the sets of bands and the species. We essentially assess the significance of information provided by hyperspectral sets of bands, in discriminating each of the species. We further studied the mean and variances of the seven species and related them to the most discriminatory parts of the EMS.

This study was concluded by looking at several common classification techniques, such as, the Spectral Angle Mapper (SAM), Spectral information divergence (SID) and mixed measures of SAM and SID which incorporate the sine and tangent trigonometric functions, SID(TAN) and SID(SIN).

2. DATA DESCRIPTION AND METHODS

2.1. Data

Hyperspectral measurements of leaves of the seven savannah trees were acquired from the Kruger National Park, in an attempt to assess tree species diversity in the park. The hyperspectral data consisted of 2151 spectral bands and seven plant tree species, which were measured using the Analytical Spectral Device (ASD) spectrometer. Prior to the analysis, these spectral bands were reduced to 1552 as the water absorption spectral bands on the leaf area were removed. The seven tree species included the Lonchocarpus capassa (LC), Combretum apiculatum (CA), Combretum heroense (CH), Combretum zeyherrea (ZC), Gymnospora buxifolia (GB), Gymnospora senegalensis (GS), and Terminalia sericia (TS). Each tree species had 10 measurements recorded with the exception of the GB, which had only seven.

2.2. Methods

2.2.1. *Stepwise discriminant analysis (SDA)*

SDA was performed initially to eliminate redundant information recorded by the ASD spectrometer and to find bands that discrimi-

N. Dudeni and P. Debba are with CSIR Built Environment. M. Cho and R. Mathieu are with the CSIR Natural Resources and the Environment. Corresponding author: N. Dudeni email: ndudeni@csir.co.za tel: +27 12 841 2958 fax: +27 12 841 3037

nate between the seven plant tree species across the spectrum. SDA builds a model of how we can best predict the groups to which each case belongs. At each step, each band is evaluated to establish its contribution to the discrimination between groups and the band with the most discriminatory power is then included in the model. This process is channeled by the F statistic, which evaluates the variables that are entered into or removed from the discriminant model, while the discriminatory power of the model is measured by Wilks' lamba. The F value for a variable indicates its statistical significance to the prediction of group membership while the Wilks' lambda tests for differences between the group means. The Wilks' lamba ranges from zero to one with values around zero indicating differences in group means, whereas values close to one indicate similarities between group means.

This procedure was performed in order to select the spectral bands with more discriminatory power between the seven plant tree species while finding the discriminant function that best predicts the classification of observations to relative species. A total of 65 spectral bands were selected by SDA. All the 65 SDA bands were used in the analysis. In addition, we considered selecting, through SDA, the top 30, 20 and 10 spectral bands respectively. This was done because as the number of bands selected grew larger, the model became unstable due to singularities resulting from high correlation between bands. Table 1 shows the spectral bands that were selected by SDA as well as the regions of the EMS considered for species discrimination.

Table 1. Spectral band configurations used in the analysis

	Wavelength (nm)	Number of bands
All bands	401-1350, 1600-1800, 2100-2500	1552
VIS bands	401-701	301
NIR bands	702-1300	601
SWIR bands	1301-1350, 1600 -1800, 2100-2500	650
SDA.65 bands	402, 404, 407, 422, 436, 442, 451, 618,644, 646, 657, 659, 665, 679, 700, 742, 955, 968, 2100,2109, 2110, 2114, 2131, 2144, 2147, 2170, 2178, 2205, 2207, 2210, 2214, 2231, 2235, 2236, 2239, 2256,2263, 2272, 2273, 2296, 2297, 2301, 2321, 2325, 2341, 2345, 2348, 2349, 2350, 2351, 2365, 2367, 2378, 2383, 2393, 2409, 2416, 2421, 2426, 2427, 2431, 2463,2465, 2468, 2495	65
SDA.30 bands	402, 407, 436, 442, 451, 618, 657, 659, 665, 679, 700, 729, 2114, 2131, 2147, 2178, 2186, 2235, 2236, 2263, 2296, 2321, 2325 2348, 2351, 2365, 2378, 2383, 2431, 249	30
SDA.20 bands	402, 436, 442, 451, 659, 679, 729, 2114, 2131, 2147, 2186, 2235, 2236, 2263, 2296, 2321, 2325, 2351, 2365, 2431	20
SDA.10 bands	402, 436, 2147, 2186, 2235, 2263, 2296, 2321, 2351 2431	10

2.2.2. *Spectral Angle Mapper (SAM)*

SAM measures the similarity by computing an angle between two spectra. These spectra are vectors in space with dimensionality equal to the number of bands [6]. The angle between spectra is then used as a measure of discrimination [7] between species or any objects of interest. SAM is defined as:

$$\mathrm{SAM}(s_i, s_j) = \cos^{-1}\left(\frac{\sum_{l=1}^{L} s_{il}s_{jl}}{\left[\sum_{l=1}^{L} s_{il}\right]^{\frac{1}{2}}\left[\sum_{l=1}^{L} s_{jl}\right]^{\frac{1}{2}}}\right), \quad (1)$$

where s_i and s_j are the two spectral signatures, and L is the total number of bands that were considered. The smaller the angle between spectra, the more similar the two spectra are. SAM is a useful similarity measure in species identification [7].

2.2.3. *Spectral information divergence (SID)*

SID measures the discrepancy probability between spectral vectors by computing the distances between probability distributions produced by spectral signatures [8]. The information divergence is measured between the probability distributions generated by spectra. The SID approach models the spectrum of hyperspectral data as a probability distribution, and is useful in capturing variations among spectral bands [7]. SID is defined as:

$$\mathrm{SID}(s_i, s_j) = \mathrm{D}(\mathrm{s_i}||\mathrm{s_j}) + \mathrm{D}(\mathrm{s_j}||\mathrm{s_i}), \quad (2)$$

where s_i and s_j are the two spectral signatures, $\mathrm{D}(\mathrm{s_i}||\mathrm{s_j})$ is the average discrepancy in self-information of s_j relative to that of s_i also known as the Kullback-Leibler information measure, and $\mathrm{D}(\mathrm{s_j}||\mathrm{s_i})$ is the average discrepancy in self-information of s_i with respect to the self-information of s_j. Smaller values of SID indicate greater similarity between the two spectra.

2.2.4. *SID-SAM mixed measure*

SID-SAM mixed measure is said to increase discriminability between two similar spectra by making them even more similar and by making the two dissimilar spectra even more distinct. Du et. al. [7] proposed two versions of SID-SAM mixed measures, one based on the tangent of the function between SAM and SID, while the other based on the sine function. These two are defined as:

$$\begin{aligned} \mathrm{SID(TAN)} &= \mathrm{SID}(s_i, s_j) \times \tan(\mathrm{SAM}(s_i, s_j)) \quad (3) \\ \mathrm{SID(SIN)} &= \mathrm{SID}(s_i, s_j) \times \sin(\mathrm{SAM}(s_i, s_j)) \quad (4) \end{aligned}$$

where s_i and s_j are the two spectral signatures. The smaller values of SID(TAN) and SID(SIN) indicate greater similarity between the two spectra.

2.2.5. *Relative Spectral discriminatory probability (RSDPB)*

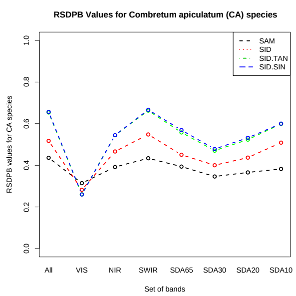
RSDPB computes the likelihood that a spectral signature t will be identified by a selective set of spectral signatures, $\triangle$. In this application, RSDPB is used to determine the likelihood that certain spectral signatures (from the entire spectrum, VIS, NIR, SWIR, and also SDA bands) can be used in discriminating between the seven predominant plant tree species found in the Kruger National Park. Therefore, the discriminatory ability of the bands in the above mentioned regions is determined through the relative spectral discriminatory probability. The relative spectral discriminatory probability is defined as:

$$P_{t,\triangle}(k) = \frac{m(t, s_k)}{\sum_{j=1}^{L} m(t, s_j)} \quad \text{for} \quad k = 1, \ldots, K \quad (5)$$

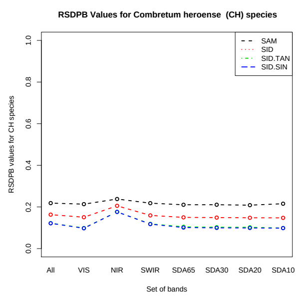
where $\sum_{j=1}^{L} m(t, s_j)$ is the normalization constant determined by similarity measures in the endmember matrix. Further, $m(t, s_k)$ is any of the predefined spectral similarity measures for the target spectral signature relative to other spectra s_k, and K is the total number of species. The higher the RSDPB value, the more likely the spectra discriminate from others in that part of the EMS.

3. RESULTS AND DISCUSSION

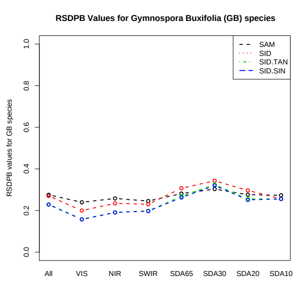
Due to the similar patterns shown by certain species, only the results for a few species are shown in this section. Figure 1(a) shows the plot of the RSDPB values for CA, using each of the discrimination methods at various parts of the EMS. For CA, using all the bands, the SWIR, and 10 SDA bands showed greater discriminatory probability compared to the other regions of the EMS. It can also be observed that the mixed measures, namely, SID(SIN) and SID(TAN) showed more discriminatory ability over the SID and SAM measures. GS showed a discrimination pattern similar to that of CA, expect that the 20 SDA bands had greater discriminatory ability (particulary for the mixed measures) compared to the other SDA bands.



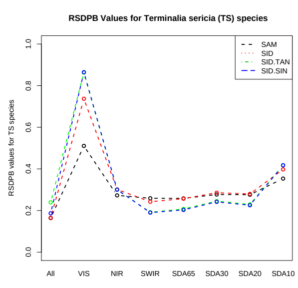
(a) RSDPB values for CA



(b) RSDPB values for CH



(c) RSDPB values for GB



(d) RSDPB values for TS

Fig. 1. RSDPB plots for species

Figure 1(b) shows the plot of the RSDPB values for CH using each of the classification methods at various parts of the EMS. The SAM method showed relatively greater discrimination than the other methods, where the largest discrimination was observed at the NIR region. All the measures, however, failed to discriminate most band sets, particulary the SDA bands. CZ showed the discrimination pattern similar to that of CH.

The GB discrimination plot shown in Figure 1 (c), indicates that using VIS, NIR, and SWIR parts of EMS showed lower discrimination compared to using the SDA bands. The 30 SDA bands showed greater discrimination than other bands.

Figure 1(d) shows the plot of the RSDPB values for TS using each of the classification methods at various parts of the EMS. The discriminatory pattern for TS appeared to be different from others, but similar to that of LC. It is also quite noticeable from this figure that the VIS region appeared to be largely discriminated by all the measures, particularly by the mixed measures.

We also linked the magnitude of spectral variation in species with the spectral regions mentioned above to examine whether the level of variation across the EMS or at specific regions thereof, gives useful information for the identification of species. We assume that the mean and variance reflectance patterns of species, at different parts of the EMS, contribute to the identification of species at such parts (if not at the entire spectrum).

Figures 2 and 3 are the spectral plots for the mean and the variances for each of the seven species. For CA, the mean and variance at SWIR region are much higher compared to the other species. According to the average wavelength for all species shown in Figure 2, the GS, just like the CA, seemed to be more different than other species at the SWIR region. GS, however, showed lower averages and smaller variations across the EMS (especially at the SWIR part), hence the other regions, except for the VIS region, also showed larger contributions to the identification of GS.

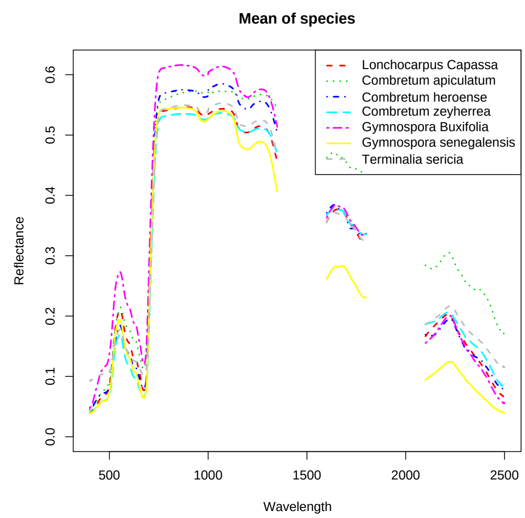


Fig. 2. Mean of species across the EMS

Figure 3 indicates that the variability within the TS species were relatively different from other species in the VIS region, hence the importance of this region in the discrimination of TS. GB did not show a very different pattern of variation at any specific part of EMS, as a result, no clear identification of GB appeared at any part of EMS. Hence it was observed that at any part of the EMS where the mean or variance patterns of species were different from other species (whether extremely large or small), that part had greater impact on species separability. Therefore, the ability of any EMS region to discriminate between all species lies in the magnitude of variation of species in that region. This could be the reason that the EMS regions in this study, as well as those used in the study by [5], failed to give any clear distinction between all the species.

Each of the 67 observations were then classified according to SAM, SID, SID(SIN), and SID(TAN). The accuracy of these sets of bands in classifying the species was examined using the kappa coefficient as defined by [9].

Figure 4 shows the classification results for each of the band regions using the different classification methods. As can be observed, SID(SIN) and SID(TAN) often performed better than SAM and SID

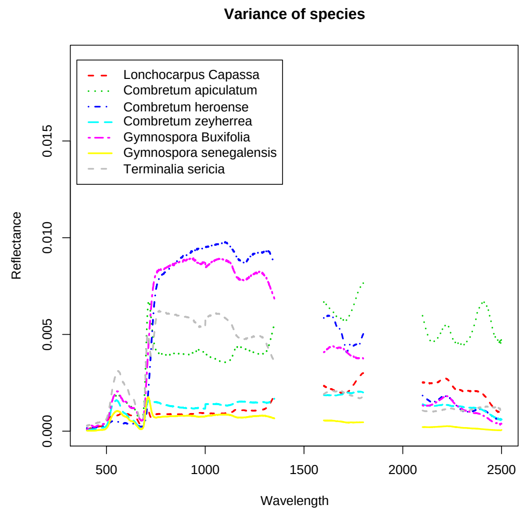


Fig. 3. Variance of species across the EMS

especially in VIS, NIR, SWIR, 20 SDA, and 10 SDA bands configurations. The bands selected from SDA often gave higher classification accuracy compared to other regions, especially for 30 SDA bands where SID classification accuracy was the highest (approximately 74% accuracy). This figure also shows that the mixed measures identified the VIS region as the region that achieved the largest degree of classification of observations to respective species, with a kappa coefficient of about 0.77.

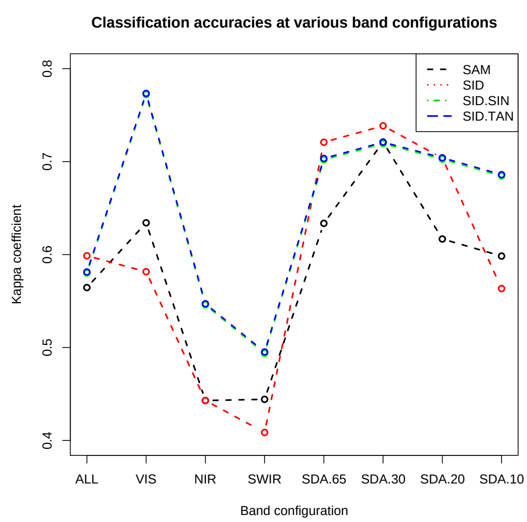


Fig. 4. Classification accuracy for each set of bands

4. SUMMARY AND CONCLUSIONS

This study used various sets of bands that could potentially discriminate between species. These sets of bands included all the bands in the EMS, VIS, NIR, and SWIR regions as well as the sets of 65, 30, 20, and 10 bands selected from SDA. Using the spectral composition of each of the seven species observed from hyperspectral remote sensing, these of bands were identified for potential discrimination by means of spectral similarity measures. This was done to establish which of the band sets or regions could be associated with the species based on discriminatory probabilities of each band set. The results indicated that certain parts of the EMS can be associated with certain species depending on the reflectance contribution exhibited by the species at those parts. The bands selected from the stepwise discriminant analysis were seen to be important for identification of species such as GB. The classification accuracy of each of the band sets was also assessed based on the classification methods such as SAM, SID, and mixed measures thereof. The mixed measures, SID(SIN) and SID(TAN) often gave better classification results for the majority of the band sets. According to these mixed measures, the VIS bands classified the species better than other sets of bands. The SDA bands, however, generally gave better classification accuracy compared to other parts of EMS. This gives is an indication that the stepwise discriminant analysis selects the bands that are statistically important and are able to provided better classification of observations to relative species even though these bands might not (in reality) be important in discrimination of various species because of the high within variability of the species compared to the between variability of the species.

5. REFERENCES

[1] M. D. Atkinson, A. P. Jervis, and R. S. Sanghar, "Discrimination between betula pendula, betula pubescens, and their hybrids using near-infrared reflectance spectroscopy," *Canadian Journal of Forest Research*, vol. 27, pp. 1896–1900, 1997.

[2] D.L. Verbyla, *Satellite Remote Sensing of Natural Resources*, New York: CRS Lewis, 1995.

[3] M. A. Cochrane, "Using vegetation reflectance variability for species level classification of hyperspectral data," *International Journal of Remote Sensing*, vol. 21, pp. 2075–2087, 2000.

[4] P. Gong, R. Pu, and B. Yu, "Conifer species recognition: An exploratory analysis of in situ hyperspectral data," *Remote sensing of Environment*, vol. 62, pp. 1827–1850, 1997.

[5] I. Sobhan, *Species discrimination from a hyperspectral perspective*, Ph.D. thesis, C. T. de Wit Graduate School for Production and Resource Conservation (PE&RC) in Wageningen University, the Nertherlands, 2007.

[6] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz, "The spectral image processing system-the interactive visualization and analysis of imaging spectrometer data," *Remote Sensing of Environment*, vol. 44, pp. 157–159, 1993.

[7] Y. Du, C. I. Chang, H. Ren, C. C. Chang, J. O. Jensen, and F. M. D'Amico, "New hyperspectral discrimination measure for spectral characterization," *Optical Engineering*, vol. 43, pp. 1777–1784, 2004.

[8] F. van der Meer, "The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 8, pp. 1– 4, 2006.

[9] J. Cohen, "A coefficient of agreement for nominal scale," *Educational and Psychological Measurement*, vol. 20, pp. 36–46, 1960.