Evaluation of discrimination measures to characterize spectrally similar leaves of African Savannah trees

Dudeni, Nontembeko
Council for Scientific and industrial Research (CSIR), Built Environment
Meering Naude Road, Brummeria
Pretoria (0184), South Africa
E-mail: ndudeni@csir.co.za

Debba, Pravesh
Council for Scientific and industrial Research (CSIR), Built Environment
Meering Naude Road, Brummeria
Pretoria (0184), South Africa
E-mail: pdebba@csir.co.za

Introduction

Hyperspectral sensors acquire hundreds to thousands of channels which provide more spectral information (Du et al., 2004) than the multispectral sensors, for instance. The high spectral resolution of hyperspectral data helps in discovering minor differences in narrow-band reflectance caused by various vegetation types and characteristics thereof, which are not detectable with multispetral data (Bajwa et al., 2004). These minor differences are essential in characterization of vegetation types. Much of the information provided by hyperspectral data is, however, redundant due to spectral and spatial correlations between individual bands (Bajwa et al., 2004). Identification of the most important bands in the characterization of vegetation types as well as removal of redundant information, is an interesting aspect in vegetation spectroscopy (Bajwa et al., 2004).

In this study, hyperspectral measurements of leaves of the seven predominant savannah trees were acquired from the Kruger National Park, South Africa's largest game reserve, in an attempt to assess tree species diversity in the park. The leaf reflectance samples for each of the species were measured using the Analytical Spectral Device (ASD) spectrometer. The seven 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).

This study uses the stepwise discriminant analysis (SDA) to identify the most important hyperspectral bands for characterization of the seven species. The role played by different regions of the electromagnetic spectrum (EMS), such as, the visible (VIS), near-infrared (NIR), and the short-wave infrared (SWIR), in the characterization of savannah trees at leaf level, is assessed. In addition, the most important bands selected by the SDA are also used to characterize these savannah tress. Spectral characterization is based on two groups of spectral similarity measures, namely, the deterministic and the stochastic measures. Deterministic measures are used to determine the geometric characteristics of spectra (Sobhan, 2007) by either measuring the angle, or the distance, or correlation between a set of spectra. In this study, a deterministic similarity measure known as the Spectral Angle Mapper (SAM), which determines spectral similarity by computing an angle between spectra, is used. The stochastic similarity measures evaluate statistical distributions of spectral reflectance values (van der Meer, 2006) of the concerned spectrum. These measures essentially define spectral variations by modelling spectral information as a probability distribution. An example of such as measure used in this study is the Spectral Information Divergence (SID). A new spectral similarity measure referred to as the SID-SAM mixed measure, which combines the deterministic measure (SAM) and a stochastic measure (SID), recently developed by (Du et al., 2004), is also used to characterize spectral properties of savannah trees. The performance of the similarity measures is compared at each band configuration, for each species, using the relative spectral discriminatory probability (RSDPB). Further details on these techniques are discussed in the following section.

Stepwise Discriminant Analysis(SDA)

SDA was primarily used to reduce redundancies in the hyperspectral measurements recorded by the Analytical Spectral Device (ASD) spectrometer and to determine the spectral bands with a greater potential for discriminating between the seven savannah tree species. SDA builds a step-by-step model which evaluates the contribution of each spectral band with respect to the discriminatory power of the model. The discriminatory power of the model is measured by the Wilk's lambda. A spectral band therefore enters the model if it, according to the Wilk's lambda criterion, contributes more to the discrimination of the tree species, while it is removed if it contributes least to the discriminatory power of the model. A discriminant model can generally be expressed as follows:

$$(1) \quad T_i = a + b_{i1} \times s_1 + b_{i2} \times s_2 + \ldots + b_{in} \times s_n,$$

where T_i represents the i^{th} tree species, a is a constant and b_1 to b_n are model coefficients, while s_1 to s_n are the bands in the EMS. The statistical significance of each spectral band in the discrimination between the various tree species is indicated by the F value.

Spectral Angle Mapper (SAM)

SAM measures the similarity between spectra by computing an angle between them, and is defined as:

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

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 (Du et al., 2004).

Spectral information divergence (SID)

Given the spectra s_i and s_j , SID is defined as:

$$(3) \quad \mathrm{SID}(s_i, s_j) = \mathrm{D}(s_i||s_j) + \mathrm{D}(s_j||s_i),$$

where $\mathrm{D}(s_i||s_j)$ 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 $\mathrm{D}(s_j||s_i)$ 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.

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., 2004) 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:

$$
\begin{aligned}
(4) \quad \text{SID(TAN)} &= \text{SID}(s_i, s_j) \times \tan(\text{SAM}(s_i, s_j)), \textit{and} \\
(5) \quad \text{SID(SIN)} &= \text{SID}(s_i, s_j) \times \sin(\text{SAM}(s_i, s_j)),
\end{aligned}
$$

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.

Relative Spectral discriminatory probability (RSDPB)

RSDPB computes the likelihood that a spectral signature $\mathbf{t}$ 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 each of the seven predominant tree species in the Kruger National Park. The relative spectral discriminatory probability is defined as:

$$(6) \quad P_{\mathbf{t},\triangle}(k) = \frac{m(\mathbf{t}, \mathbf{s}_k)}{\sum_{j=1}^{L} m(\mathbf{t}, \mathbf{s}_j)} \quad \text{for} \quad k = 1, \ldots, K,$$

where $\sum_{j=1}^{L} m(\mathbf{t}, \mathbf{s}_j)$ is the normalization constant determined by similarity measures in the known spectral matrix. Further, $m(\mathbf{t}, \mathbf{s}_k)$ is any of the predefined spectral similarity measures for the target spectral signature relative to other spectra $\mathbf{s}_k$, 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.

Results

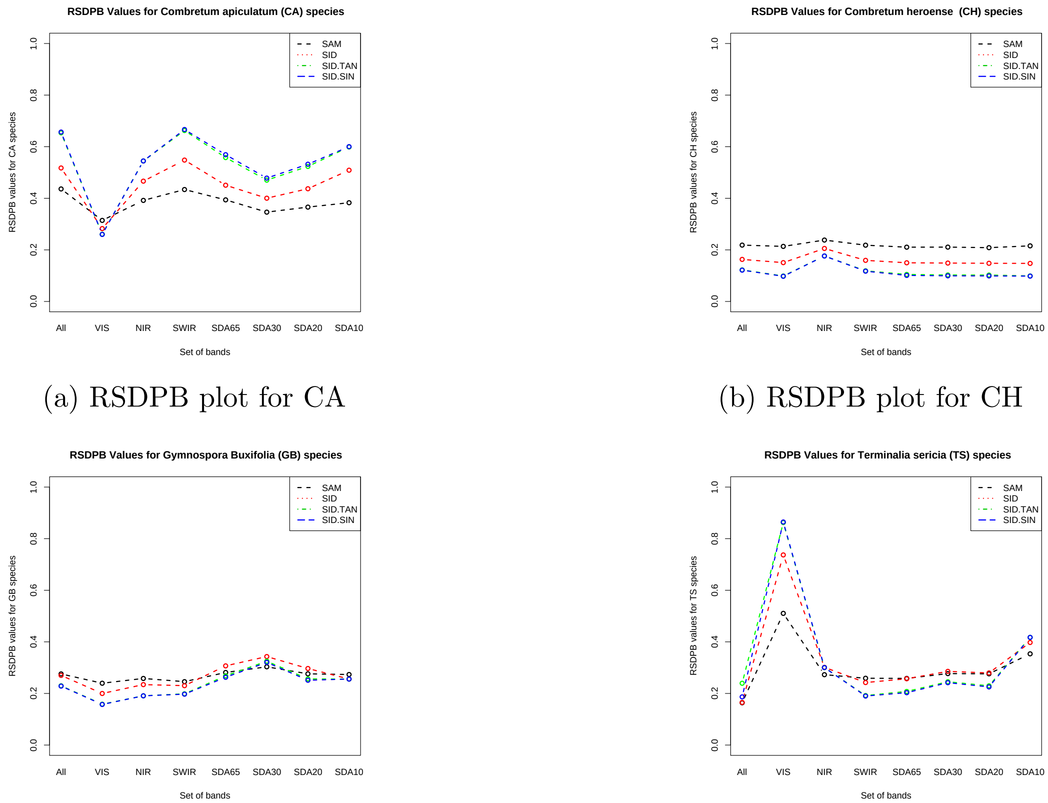
The results for the two species (LC and GS) are not shown as they exhibited similar characterization patterns with one or more of species shown in this section. Figure 1(a) compares the performance of the measures in characterization of CA and assesses it at various parts of the EMS. For CA, using all 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, except that the 20 SDA bands had greater discriminatory ability (particulary for the mixed measures) compared to the other SDA bands.

Figure 1(b) shows that for CH, SAM showed relatively greater characterization ability at all parts of the EMS, where the largest characterization was observed at the NIR region. The characterization pattern shows, however, that all the measures failed to clearly characterize CH at different EMS part, particulary the SDA bands. CZ also showed the characterization pattern similar to that of CH.

For characterization of GB, Figure 1 (c) generally indicates that no major role was played parts of EMS, while there is an indication that the 30 SDA bands are relatively important in the identification of GB.

Figure 1(d) indicates that the characterization of TS is largely associated with the VIS bands. 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. LC species also exhibited the same pattern, but, the SDA bands also seemed to have a relatively bigger role in its characterization.

The magnitude of spectral variation in species was linked with the spectral regions mentioned above to examine whether the level of variation across the EMS or at specific regions thereof, gives



(a) RSDPB plot for CA

(b) RSDPB plot for CH

(c) RSDPB plot for GB

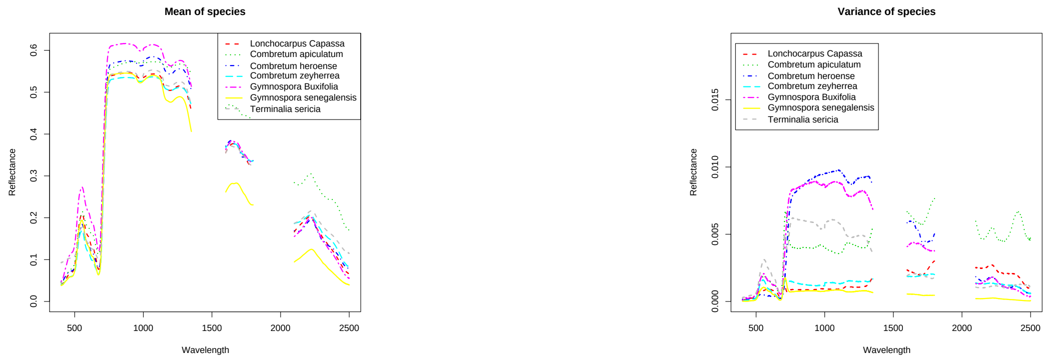
(d) RSDPB plot for TS

Figure 1: Spectral characterization of species (RSDPB plots)

useful information for the identification of species. It was assumed 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).

Figure 2 represents 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 mean 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, hence this region played a larger role in their characterization. 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.

Figure 2 (b) 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 characterize all species also depends in the magnitude of variation of species in that region. This could be the reason none of the EMS regions in this study, as well as those used in the study by (Sobhan, 2007), failed to characterize all the species.



(a) Mean of species (b)Variance of species

Figure 2: Mean and Variance of species

Summary and Conclusions

This study assessed the role played by each of the regions of the EMS as well as the SDA bands in discriminating each of the seven savannah tress. The regions of the EMS considered included the VIS, NIR, and SWIR regions as well as the sets of 65, 30, 20, and 10 bands selected from SDA. In addition, the performance of the deterministic and stochastic spactral similarity measures in the characterization of species was compared using the relative discriminatory probability. From the results, the SWIR region was observed to be important in the characterization of CA and GB species, while the VIS part of EMS was seen to be largely significant in the characterization of the TS and LC. The results also indicated that certain parts of the EMS play a significant role in the characterization of species, depending on the variability contribution of species at those parts. None of the EMS parts, however, had the ability to characterize all the species. As expected, the SID-SAM mixed measures of spectral similarity generally performed better than the other measures with respect to the species characterization given their leaf reflectance properties.

References

Bajwa, S., Bajcsy, P., Groves, P., & Tian, L. F. (2004). Hyperspectral image data mining for band selection in agricultural apllications. *American Society of Agricultural Engineers*, 47(3), 895–896.

Du, Y., Chang, C. I., Ren, H., Chang, C. C., Jensen, J. O., & D'Amico, F. M. (2004). New hyperspectral discrimination measure for spectral characterization. *Optical Engineering*, 43, 1777–1784.

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

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