WITHIN- AND BETWEEN-CLASS VARIABILITY OF SPECTRALLY SIMILAR TREE SPECIES

P. Debba[1], *M.A. Cho*[2], *R. Mathieu*[2]

The Council for Scientific and Industrial Research (CSIR)
[1]CSIR Built Environment
[2]CSIR Natural Resources of the Environment
P. O. Box 395, 0001, South Africa
email: pdebba@csir.co.za

ABSTRACT

In this paper, a comparison is made through evaluating the within- and between-class species variability for the original, the first derivative and second derivative spectra. For each, the experiment was conducted (i) over the entire electromagnetic spectrum (EMS), (ii) the visible (VIS) region, (iii) the near infrared (NIR) region, (iv) the short wave infrared (SWIR) region, (v) using band selection, for example, best 10, 20, 30 and 65 bands selected, through linear stepwise discriminant analysis (vi) using sequential selection of bands, for example, every 5th, 9th, 15th, 19th or 25th band selected and (vii) spectral degradation of the spectral bands by averaging the reflectance values for every 5th, 9th, 15th, 19th or 25th band. We concluded that for this data set, there are important bands from the original spectra, the first and second derivative spectra and from various regions of the EMS (VIS, NIR, SWIR) that is important for species separability. Furthermore, there did not seem to be any decrease in species separability, for this data set, by degrading the spectral bands through averaging the reflectance. This implies that hyperspectral (high spectral) measurements did not prove useful in species separability compared to lower spectral resolution data.

Index Terms— Species separability, within-class variability, between-class variability, spectrally similar, classification, spectral unmixing, band selection, spectral resolution

1. INTRODUCTION

Spectral signatures for samples of the same species of vegetation could have high within-class species variability [1] and when comparing spectrally similar species this issue of high within-class species variability enhances the problem of obtaining high accuracy in image classification and/or spectral unmixing [2]. Hence techniques that are presently used for image classification and/or spectral unmixing often cannot be directly used for vegetation studies to distinguish spectrally similar species [3]. Most researchers in the past have ignored this or have used the mean spectrum for the class for which image classification and/or spectral unmixing techniques are based on the mean spectra. This results in (i) a loss of valuable information as a result from individual samples and (ii) undistinguishable mean spectra for spectrally similar vegetation species.

The two main types of variability, which is necessary for any image classification and/or spectral unmixing technique are (i) the variability within a species class, and (ii) the similarity between the species classes [4]. When the variability within a species class is small compared to the variability between the species classes, this results in relatively good accuracy in the results for image classification and/or spectral unmixing. However, when the species spectra is similar, the within-species variability can be large compared to the between-species class variability, which is more prominent in vegetation studies, and thus producing poor results for image classification and/or spectral unmixing techniques.

This research studies the variability within a species class and the variability between the species classes of seven spectrally similar tree species and presents ways in which the within-species class variability can be reduced compared to the between-species class variability.

2. DATA DESCRIPTION

The Analytical Spectral Device (ASD) spectrometer (FieldSpec3 Pro FR) was used to record hyperspectral measurements of leaf samples taken from several different savannah trees in the Kruger National Park in South Africa, in an attempt to assess tree species diversity in the park. The hyperspectral data consist of 2151 spectral bands at a spectral resolution of 1 nm for seven common plant tree species in the area. The seven tree species include Lonchocarpus capassa, Combretum apiculatum, Combretum heroense, Combretum zeyherrea, Gymnospora buxifolia, Gymnospora senegalensis, and Terminalia sericia. Each tree species has 10 measurements recorded (see Figure 1 for Combretum apriculatum and Figure 2 for Terminalia sericia) with the exception of Gymnospora Buxifolia, which has only seven. The total data set therefore had 67 observations for the species measurements. The mean and the variance of the spectral reflectance for each of the seven species is shown in Figures 3 and 4. From Figure 3, the similarity between the seven species can clearly be seen, whereas Figures 1, 2, and 4 indicates the high within-species variability.

3. METHOD AND RESULTS

Let y_i^k denotes the d-dimensional feature vector (d represents the number of bands) selected from the i^{th} sample of the k^{th} class, c_k, with n_k samples in the k^{th} class. Furthermore, let μ_k $(k = 1, \ldots, c)$ be the mean vector of k^{th} class and μ be the total mean vector in this d-dimensional feature space. The within-class, S_w, and between-class, S_b, variances can be calculated in this feature space as follows:

$$S_w = \frac{1}{c} \sum_{k=1}^{c} \left[\frac{1}{n_k} \sum_{i=1}^{n_k} \left(y_i^k - \mu_k \right)^T \left(y_i^k - \mu_k \right) \right] \quad (1)$$

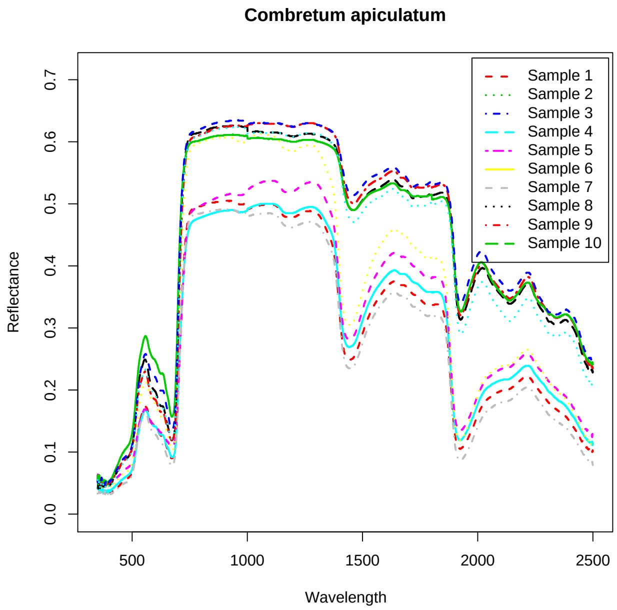

Fig. 1. Reflectance spectra of the 10 samples for Combretum apriculatum.

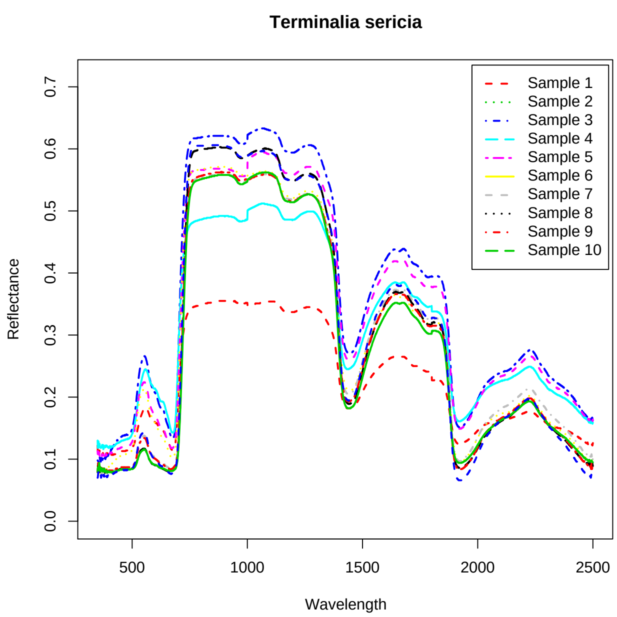

Fig. 2. Reflectance spectra of the 10 samples for Terminalia sericia.

and

$$S_b = \frac{1}{c}\sum_{k=1}^{c} (\mu_k - \mu)^T (\mu_k - \mu) \,. \tag{2}$$

The ratio of the between-class variability to the within-class variability, commonly known as Fisher's criterion ratio, is a measure for class separation, with high values indicating greater class separation.

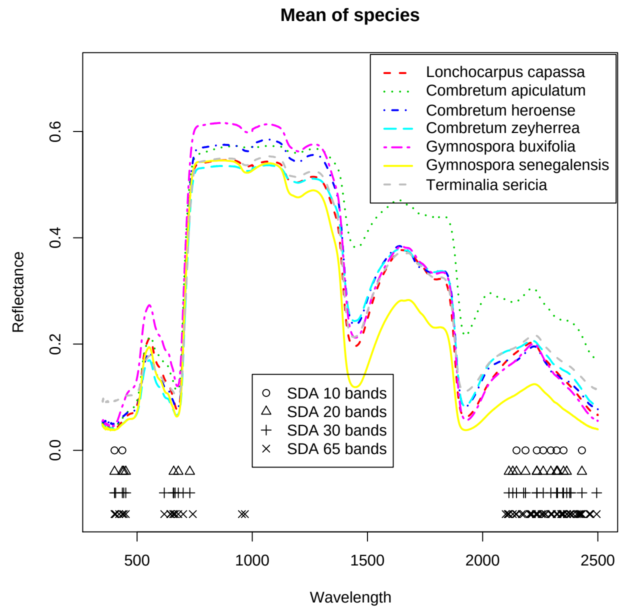

Fig. 3. Mean spectral reflectance for all seven species. Also, band selection using stepwise discriminant analysis. For SDA the results for the best 10, 20, 30 and 65 selected bands are shown.

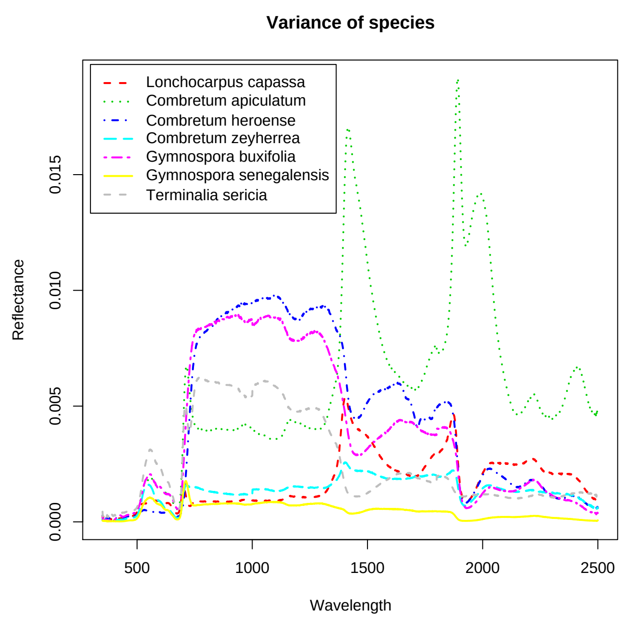

Fig. 4. Variance of the spectral reflectance for all seven species.

In the past several, researchers have considered using the first and second derivatives of the spectra to improve the image classification and/or spectral unmixing results [5, 6, 7] without actually reflecting the within- and between-class species variability. A comparison is made through evaluating the within-class species variability and the between-class species variability for the original, the first derivative and second derivative spectra. For each, the experiment was conducted (i) over the entire electromagnetic spectrum (EMS) (0.350–2.500 μm), (ii) the visible (VIS) (0.400–0.740 μm) region, (iii) the near infrared (NIR) (0.741–1.300 μm) region, (iv) the short wave infrared (SWIR) (1.301–2.500 μm) region, (v) using band selection, for example, best 10, 20, 30 and 65 bands selected, through linear stepwise discriminant analysis (SDA), for which the selected bands are shown in Figure 3 (vi) using sequential selection of bands, for example, every 5th, 9th, 15th, 19th or 25th band selected and (vii) spectral degradation of the spectral bands by averaging the reflectance values for every 5th, 9th, 15th, 19th or 25th band. Each of the above was considered so as to reduce the high correlations between species, since it is already well known that high correlations produces an unstable inverse matrix and results in a decrease in the accuracy for image classification and/or spectral unmixing [8, 5, 9]. The results for the above can be seen in Tables 1 and 2.

Bands	Within-class variability	Between-class variability	Ratio
All			
Original	5.574	5.030	0.902
1st derivative	9.007×10^{-3}	4.000×10^{-3}	0.444
2nd derivative	1.522×10^{-2}	3.582×10^{-2}	0.235
VIS			
Original	0.316	0.291	0.920
1st derivative	2.220×10^{-4}	1.160×10^{-4}	0.523
2nd derivative	1.787×10^{-4}	2.797×10^{-5}	0.157
NIR			
Original	2.090	0.481	0.230
1st derivative	1.163×10^{-4}	4.254×10^{-4}	0.366
2nd derivative	2.557×10^{-4}	7.420×10^{-5}	0.290
SWIR			
Original	3.162	4.241	1.341
1st derivative	3.594×10^{-4}	1.568×10^{-4}	0.436
2nd derivative	7.013×10^{-4}	8.371×10^{-5}	0.119
SDA10			
Original	0.013	0.021	1.600
1st derivative	6.621×10^{-8}	1.445×10^{-7}	2.183
2nd derivative	4.763×10^{-11}	1.273×10^{-10}	2.672
SDA20			
Original	0.026	0.038	1.463
1st derivative	1.339×10^{-6}	8.253×10^{-7}	0.616
2nd derivative	2.061×10^{-7}	3.661×10^{-8}	0.178
SDA30			
Original	0.037	0.055	1.473
1st derivative	4.520×10^{-6}	3.194×10^{-7}	0.707
2nd derivative	2.061×10^{-7}	6.138×10^{-8}	0.247
SDA65			
Original	0.095	0.135	1.428
1st derivative	7.298×10^{-6}	4.482×10^{-6}	0.614
2nd derivative	3.271×10^{-6}	3.130×10^{-7}	0.096

Table 1. Within- and between-class variability for the original, 1st and 2nd derivative for various regions of the EMS or selection of bands.

Bands	Within-class variability	Between-class variability	Ratio
Every 5th spectrum			
Original	1.114	1.006	0.902
1st derivative	6.060×10^{-5}	5.062×10^{-5}	0.835
2nd derivative	1.135×10^{-6}	5.649×10^{-7}	0.498
Every 9th spectrum			
Original	0.619	0.559	0.903
1st derivative	3.002×10^{-5}	2.636×10^{-5}	0.878
2nd derivative	1.755×10^{-7}	1.327×10^{-7}	0.756
Every 15th spectrum			
Original	0.371	0.335	0.902
1st derivative	1.658×10^{-5}	1.480×10^{-5}	0.893
2nd derivative	6.056×10^{-8}	5.305×10^{-8}	0.876
Every 19th spectrum			
Original	0.293	0.264	0.902
1st derivative	1.232×10^{-5}	1.113×10^{-5}	0.904
2nd derivative	3.592×10^{-8}	3.265×10^{-8}	0.909
Every 25th spectrum			
Original	0.223	0.201	0.902
1st derivative	8.837×10^{-6}	7.753×10^{-6}	0.877
2nd derivative	2.038×10^{-8}	1.702×10^{-8}	0.835
Every 5th averaged			
Original	1.114	1.005	0.902
1st derivative	5.514×10^{-5}	4.819×10^{-5}	0.874
2nd derivative	5.582×10^{-7}	3.339×10^{-7}	0.598
Every 9th averaged			
Original	0.619	0.559	0.902
1st derivative	2.799×10^{-5}	2.511×10^{-5}	0.897
2nd derivative	1.260×10^{-7}	1.089×10^{-7}	0.864
Every 15th averaged			
Original	0.371	0.334	0.901
1st derivative	1.504×10^{-5}	1.356×10^{-5}	0.902
2nd derivative	4.713×10^{-8}	4.219×10^{-8}	0.895
Every 19th averaged			
Original	0.293	0.264	0.902
1st derivative	1.090×10^{-5}	9.880×10^{-6}	0.906
2nd derivative	2.737×10^{-8}	2.478×10^{-8}	0.905
Every 25th averaged			
Original	0.222	0.201	0.902
1st derivative	7.383×10^{-6}	6.591×10^{-6}	0.893
2nd derivative	1.399×10^{-8}	1.193×10^{-8}	0.853

Table 2. Within- and between-class variability for the original, 1st and 2nd derivative for selected bands and comparison to spectrally degraded bands.

4. DISCUSSION AND CONCLUSIONS

From the results in Tables 1 and 2, it can be seen that the within-class species variability is often higher than the between-class species variability, thus making Fisher's criterion ratio less than one. This implies that the accuracy for image classification and/or spectral unmixing will be low. From these tables, the derivative spectra for this data set will often be even more problematic in terms of species separability compared to the original spectra and furthermore, the within- and between- class species variability are smaller in magnitude for the derivatives as compared to the original spectra. This could affect the reliability of the estimates, in terms of rounding, when using derivative spectra because of the extremely small values.

From Table 1, the VIS and SWIR regions of the EMS have higher species separability compared to using all spectral bands, at least for the original spectra. For NIR region, the first derivative has the highest species separability, while the original spectra has the lowest species separability. When the selected bands from linear stepwise discriminant analysis were used, the between-class species variability is higher than the within-class species variability for the original spectra, thus making Fisher's criterion ratio greater than one. Except for when the best 10 bands were selected, the derivative spectra have lower species separability compared to the original spectra. It is interesting that for the 10 best bands selected from linear stepwise discriminant analysis, the second derivative had higher species separability, while the original spectra had the lowest species separability. Eight of the 10 best bands were selected from the SWIR region, while the remaining two are from the VIS region (see Figure 3). This contradicts the results in [10, 11, 12] where the authors identified the VIS (0.400–0.700 μm) region to be most appropriate to discriminate between soil and vegetation, but is in agreement with [13] where the authors identified SWIR2 (2.050–2.500 μm) region to be most appropriate to discriminate between soil and vegetation.

From Table 2, there does not seem to be any difference in terms of species separability for the original spectra when all bands was used, or using selected bands, namely, every 5th, 9th, 15th, 19th or 25th band selected or when the bands were spectral degraded by averaging the reflectance values for every 5th, 9th, 15th, 19th or 25th band. However, when the derivative spectra were used, the species separability increased up to when every 19th band was selected. The species separability then deceased when every 25th band was selected using derivative spectra.

In general, we can conclude that for this data set, there are important bands from the original spectra, the first and second derivative spectra and from various regions of the EMS (VIS, NIR, SWIR) that is important for species separability. Hence, we recommend further research and improvement to selecting spectral bands from a broader combined set of the original, first and second derivative spectra. Having selected the most discriminating set of spectral bands, these can be used to further increase the between-class species variability compared to the within-class species variability, thereby increasing the accuracy for image classification and/or spectral unmixing. Furthermore, for this data set, there does not seem to be any decrease in species separability by degrading the spectral bands through averaging the reflectance. This implies that hyperspectral (high spectral) measurements did not prove useful in species separability compared to lower spectral resolution data.

5. REFERENCES

[1] L. Gomez-Chova, J. Calpe, G. Camps-Valls, J.D. Martin, E. Soria, J. Vila, L. Alonso-Chorda, and J. Moreno, "Semi-supervised classification method for hyperspectral remote sensing images," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium, 2003. IGARSS '03*, 2003.

[2] Erin L. Hestir, Shruti Khanna, Margaret E. Andrew, Maria J. Santos, Joshua H. Viers, Jonathan A. Greenberg, Sepalika S. Rajapakse, and Susan L. Ustin, "Identification of invasive vegetation using hyperspectral remote sensing in the California Delta ecosystem," *Remote Sensing of Environment*, vol. 112, pp. 4034–4047, 2008.

[3] Gregory S. Okina, Dar A. Roberts, Bruce Murraya, and William J. Okin, "Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments," *Remote Sensing of Environment*, vol. 77, pp. 212–225, 2001.

[4] Jinkai Zhang, Benoit Rivard, Arturo Sánchez-Azofeifa, and Karen Castro-Esau, "Intra- and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: Implications for species identification using HYDICE imagery," *Remote Sensing of Environment*, vol. 105(2), pp. 129–141, 2006.

[5] P. Debba, E. J. M. Carranza, F. D. van der Meer, and A. Stein, "Abundance estimation of spectrally similar materials by using derivatives in simulated annealing," *IEEE Geoscience and Remote Sensing*, vol. 44, no. 12, pp. 3649–3658, 2006.

[6] J. Zhang, B. Rivard, and A. Sanchez-Azofeifa, "Derivative spectral unmixing of hyperspectral data applied to mixtures of lichen and rock," *IEEE Transactions of Geoscience and Remote Sensing*, vol. 42(9), pp. 1934–1940, 2004.

[7] R.L. Pu, P. Gong, G.S. Biging, and M.R. Larrieu, "Extraction of red edge optical parameters from Hyperion data for estimation of forest leaf area index," *IEEE Transactions on Geosciences and Remote Sensing*, vol. 41(4), pp. 916–921, 2003.

[8] J. Settle, "On the effect of variable endmember spectra in the linear mixture model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, pp. 389–396, 2006.

[9] F. D. Van der Meer and S. M. De Jong, "Improving the results of spectral unmixing of Landsat Thematic Mapper imagery by enhancing the orthogonality of end-members," *International Journal of Remote Sensing*, vol. 21, no. 15, pp. 2781–2797, 2000.

[10] X. Miao, P. Gong, S. Swope, R. Pu, R. Carruthers, J. S. Anderson, G. L. Heaton, and C. R. Tracy, "Estimation of yellow starthistle abundance through CASI-2 hyperspactral imagery using linear spectral mixture models," *Remote Sensing of Environment*, vol. 101, pp. 329–341, 2006.

[11] J.A.N. van Aardt and R.H. Wynne, "Spectral separability among six southers tree species," *Photogrammetric Engineering and Remote Sensing*, vol. 67, pp. 1367–1375, 2001.

[12] R. Dehaan, J. Louis, A. Wilson, A. Hall, and R. Rumbachs, "Discrimination of blackberry (Rubus fruticosus sp. agg.) using hyperspectral imagery in Kosciuszko National Park, NSW, Australia," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, pp. 13–24, 2007.

[13] Gregory P. Asner and David B. Lobell, "A biogeophysical approach for automated SWIR unmixing of soils and vegetation," *Remote Sensing of Environment*, vol. 74, pp. 99–112, 2000.