MULTIPLE ENDMEMBER SPECTRAL-ANGLE-MAPPER (SAM) ANALYSIS IMPROVES DISCRIMINATION OF SAVANNA TREE SPECIES

Moses Azong Cho[1]*, Renaud Mathieu*[1] *and Pravesh Debba*[2]

[1]Council for scientific and industrial Research (CSIR), Natural Resources and Environment (NRE)
[2]CSIR, Built Environment, P.O. BOX 395, Pretoria, 0001, South Africa

mcho@csir.co.za, PDebba@csir.co.za, rmathieu@csir.co.za

ABSTRACT

Differences in within-species phenology and structure driven by factors including topography, edaphic properties, and climatic variables across the landscape present important challenges to species differentiation with remote sensing. The objective of this paper was to evaluate the classification performance of a multiple-endmember spectral angle mapper (SAM) classification approach in discriminating seven common African savanna tree species and to compare the results with the traditional SAM classifier based on a single endmember per species or class. The leaf spectral reflectances of seven common tree species in the Kruger National Park, South Africa, *Combretum apiculatum, Combretum hereroense, Combretum zeyheri, Gymnosporia buxifolia, Gymnosporia senegalensis, Lonchocarpus capassa and Terminalia sericea* were used in this study. Discriminating species using all training spectra for each species as reference endmembers (i.e. the multiple endmember approach or more conventionally termed K-nearest neighbour classifier) yielded a higher classification accuracy of 60% compared to the conventional SAM classifier based on the mean of the training spectra for each species (overall accuracy = 44%). Further analysis using endmembers selected after cluster analysis of all the spectra for each species yielded the highest classification accuracy for the species (overall accuracy = 74%). This study underscores two important phenomena; (i) within-species spectral variability affects the discrimination of savanna tree species with the SAM classifier and (ii) the effect of within-species spectral variability can be minimised by adopting a multiple endmember approach with the SAM classifier. This study further highlights the importance of the quality of the reference endmember or spectral library.

Index Terms— *Savanna tree species; Spectral variability; Multiple endmember approach; Spectral angle mapper*

1. INTRODUCTION

The ability to map vegetation to the species level is of wide interest in Ecology. Species-level maps of vegetation have important applications in resource inventories, biodiversity assessment, and fire hazard assessment. Species mapping with remote sensing is based on the assumption that each species has a unique spectral signature. Spectral signatures of vegetation vary according to biochemical content, physical structure of plant tissue and canopy architecture.

Several mapping methods are applied in remote sensing to quantify species or vegetation community distribution at the local to regional scale. The most commonly used methods include maximum likelihood, spectral mixture analysis (SMA)[1] and spectral angle mapper (SAM)[2]. The application of some of these methods including SAM and SMA has become popular with the advent of hyperspectral remote sensing. SAM determines the degree of similarity between two spectra by treating the spectra as vectors in a space with dimensionality equal to the number of bands[2]. Each vector has a certain length and direction. The length of the vector represents brightness of the target while the direction represents the spectral feature of the target. Variations in illumination mainly affect changes in vector length, while spectral variability between different spectra affects the angle between their corresponding vectors[2]. SAM is appropriate for species-level monitoring at the regional scale as it is not sensitive to differences in illumination or albedo[3, 4]. SMA on the other hand is a subpixel classifier that determines the relative abundance of materials that are depicted in multispectral or hyperspectral imagery based on the materials' spectral characteristics. The reflectance at each pixel of the image is assumed to be a linear combination of the reflectance of each material (or endmember) present within the pixel.

Spectral Angle Mapper (SAM) is a physically-based spectral classification that uses an n-D angle to match pixels to reference spectra. The conventional SAM classifier in ENVI software compares the angle between the endmember spectrum vector and each spectral vector in n-D space. Smaller angles represent closer matches to the reference spectrum. Spectra further away than the specified maximum angle threshold in radians are not classified. ENVI provides the option of no threshold or a single threshold for all classes or a different threshold for each species or class can be applied. However, the important point to note here is that a single reference spectrum is used for each class. This means that spectral variability within each class, denoted as the intra-class variability, is not preserved. Some studies have gone through a stepwise approach by first classifying possible classes for each species before merging them into one class[3, 5]. We argue in this paper that by not sufficiently considering the variability around the means of the reference spectra, the ability of the classifier to classify species with high intra-species variability is weakened. So, our research hypothesis centres on the fact that a multiple-endmember approach involving endmembers that are representative of the within-class spectral variability would provide a higher classification accuracy compared to the conventional SAM classifier involving a single spectrum as the reference for each class. The multiple endmember approach has proven successful with SMA[6] but has not been assessed for SAM. This is an experimental study designed to evaluate the classification performance of a multiple endmember approach (conventionally called nearest neighbour approach) of SAM in discriminating seven common African savanna tree species and compare the results with the traditional SAM classifier based on a single endmember per species or class.

In the basic nearest neighbours classifier, each training spectrum is used as a reference spectrum and the unknown (test) spectrum is assigned to the class of the closest (i.e. spectrally most similar) reference spectrum [7]. But the point we are making in this study is that specific target discrimination based on a spectral library such those of plant species is based on the assumption that each species has a unique spectral signature termed endmember for that species. But this is not often the case because of phenological differences or differences resulting from the same species growing on different substrates. We would therefore rather work on the assumption that each species has a set of unique identities (IDs) or multiple endmembers. The question is how to establish the set of (IDs) for each species? The representativeness of the endmembers of IDs must be the most important criterion for establishing a spectral library for each species. In study we have evaluated classification accuracies for seven tree species involving multiple endmembers selected (i) through random sampling from a set of spectra for each species and (ii) from representative spectra for each species chosen after cluster analysis.

2. METHODS AND MATERIAL

2.1. Leaf spectral measurement

The leaf spectral reflectances of seven common tree species in the Kruger National Park, South Africa, *Combretum apiculatum, Combretum hereroense, Combretum zeyheri, Gymnosporia buxifolia, Gymnosporia senegalensis, Lonchocarpus capassa and Terminalia sericea* were used in this study. To simulate intra-species spectral variability for the purpose of this study, leaves representing varying phenological stages for each species were used in the study. The leaf spectra were measured using the leaf probe of the ASD spectrometer (FieldSpec3 Pro FR, Analytical spectral Device, Inc, USA) in the 350-2500 nm range (Fig.1). The measurements were made on sunlit leaves.

2.2. Data analysis

First, the spectral data for each species were randomly split into the training (1/3) and test (2/3) data. A resampling procedure was adopted in the splitting the data. Ten replicates were created by repeated resampling with replacement. Subsequently, two types of reference endmember spectra were used to classify the species in the test data set: (i) the mean spectrum of the training data set for each species and (ii) all training spectra for each species in a multiple-endmember approach conventionally called nearest neighbour approach. Additionally, the classification accuracy for the species involving training spectra selected after cluster analysis was assessed, i.e. cluster analysis was used to organise the spectra of each species into similar groups. The training spectra was then constituted be selecting one spectrum from each group.

SAM was calculated between spectral pairs for the whole spectral range (350-2500 nm). The overall, users and producers classification accuracies were determined as the mean values for the ten replicates from the confusion matrices between the observed and predicted data. The effect of within species spectral similarity or dissimilarity was also assessed.

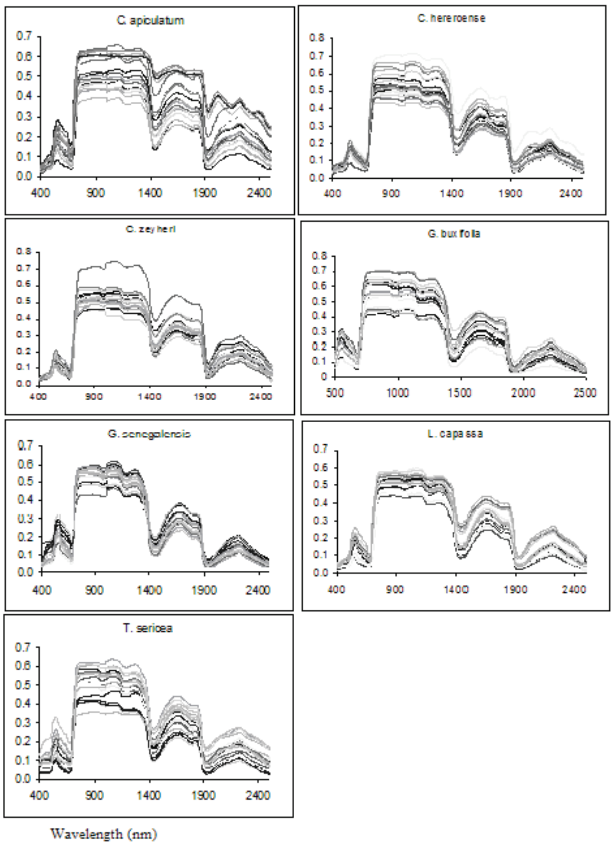


Fig.1. leaf spectral reflectance of 7 savanna tree species: *Combretum apiculatum*, *Combretum hereroense*, *Combretum zeyheri*, *Gymnosporia buxifolia*, *Gymnosporia senegalensis*, *Lonchocarpus capassa*, *Terminalia sericea*

3. RESULTS

The highest within-species spectral dissimilarity was observed for *C. apiculatum* (Fig.1&2), followed by *L. capassa* and *T. sericea*. The leaves of *C. apiculatum* varied from fresh to highly senesced leaves as revealed by their spectral reflectance (Fig.1&2). This accounted for the high within-species spectral dissimilarity. *G. senegalensis* showed the highest within-species spectral similarity, followed by *C. hereroense*. High between species similarity was observed between the following species (Fig.3.):

- C. apiculatum and C. zeyheri. In fact, the combretums were closer to each other when compared with the rest of the species.
- G. buxifolia and G. senegalensis

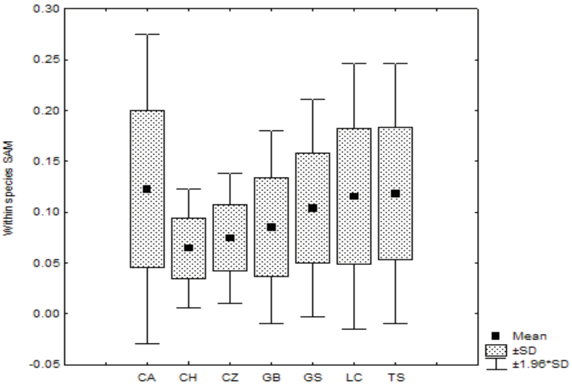


Fig.2. Within-species spectral similarity (spectral angle mapper - SAM). CA= *Combretum apiculatum*, CH=*Combretum hereroense*, CZ=*Combretum zeyheri*, GB=*Gymnosporia buxifolia*, GS=*Gymnosporia senegalensis*, LC = *Lonchocarpus capassa*, TS=*Terminalia sericea*

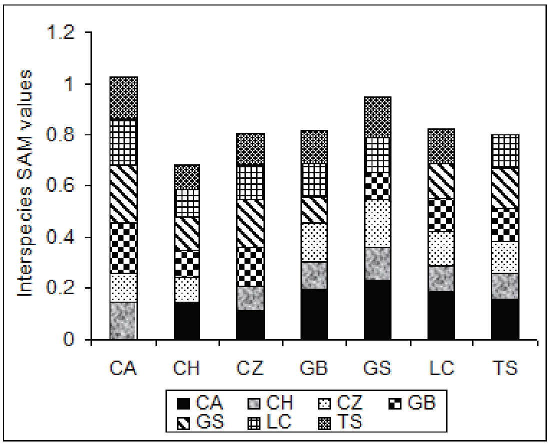
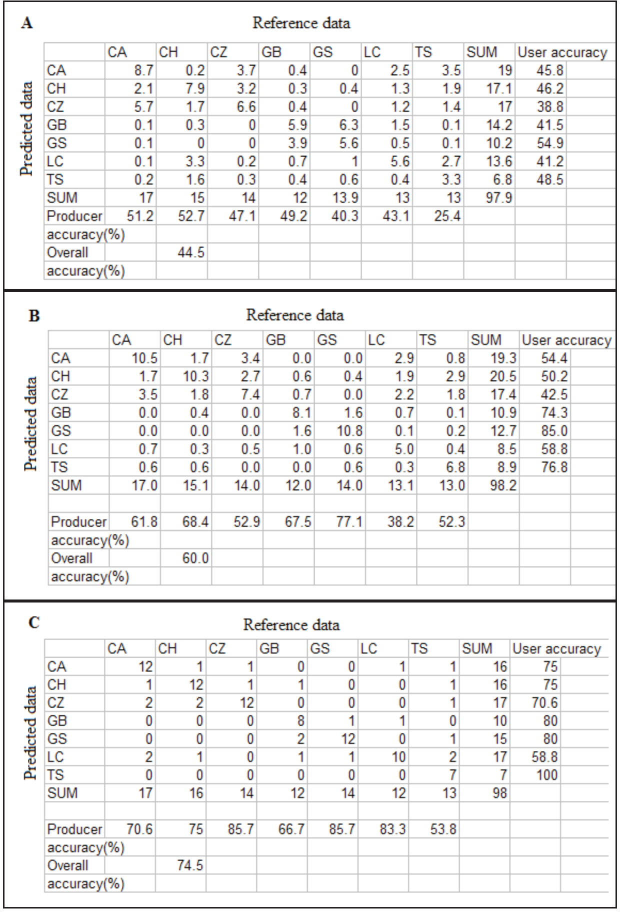


Fig.3. Interspecies similarity using spectral angle mapper (SAM). The lower the SAM the greater the similarity between two species e.g. *Combretum apiculatum* (CA) shows the highest similarity with *Combretum zeyheri* (CZ).

Discriminating the 7 tree species using the multiple endmember approach yielded a higher classification accuracy (overall accuracy = 60%) compared with the classification involving the mean spectrum of each species as training spectrum (overall accuracy = 44%) (Table 1 A and B). However, the species were most accurately classified using the multiple endmember classifier involving endmembers selected after performing cluster analysis (overall accuracy = 74%).

Table 1. Classification accuracies for 7 savanna tree species using (A) the mean spectrum of the training set for each species as reference endmember, (B) all training spectra selected by random sampling for each species as reference endmembers and (C) all training spectra selected after cluster analysis as endmembers: CA= *Combretum apiculatum*, CH=*Combretum hereroense*, CZ=*Combretum zeyheri*, GB=*Gymnosporia buxifolia*, GS=*Gymnosporia senegalensis*, LC = *Lonchocarpus capassa*, TS=*Terminalia sericea*



A

Reference data

Predicted data

	CA	CH	CZ	GB	GS	LC	TS	SUM	User accuracy
CA	8.7	0.2	3.7	0.4	0	2.5	3.5	19	45.8
CH	2.1	7.9	3.2	0.3	0.4	1.3	1.9	17.1	46.2
CZ	5.7	1.7	6.6	0.4	0	1.2	1.4	17	38.8
GB	0.1	0.3	0	5.9	6.3	1.5	0.1	14.2	41.5
GS	0.1	0	0	3.9	5.6	0.5	0.1	10.2	54.9
LC	0.1	3.3	0.2	0.7	1	5.6	2.7	13.6	41.2
TS	0.2	1.6	0.3	0.4	0.6	0.4	3.3	6.8	48.5
SUM	17	15	14	12	13.9	13	13	97.9	
Producer accuracy(%)	51.2	52.7	47.1	49.2	40.3	43.1	25.4		
Overall accuracy(%)		44.5							

B

Reference data

Predicted data

	CA	CH	CZ	GB	GS	LC	TS	SUM	User accuracy
CA	10.5	1.7	3.4	0.0	0.0	2.9	0.8	19.3	54.4
CH	1.7	10.3	2.7	0.6	0.4	1.9	2.9	20.5	50.2
CZ	3.5	1.8	7.4	0.7	0.0	2.2	1.8	17.4	42.5
GB	0.0	0.4	0.0	8.1	1.6	0.7	0.1	10.9	74.3
GS	0.0	0.0	0.0	1.6	10.8	0.1	0.2	12.7	85.0
LC	0.7	0.3	0.5	1.0	0.6	5.0	0.4	8.5	58.8
TS	0.6	0.6	0.0	0.0	0.6	0.3	6.8	8.9	76.8
SUM	17.0	15.1	14.0	12.0	14.0	13.1	13.0	98.2	
Producer accuracy(%)	61.8	68.4	52.9	67.5	77.1	38.2	52.3		
Overall accuracy(%)		60.0							

C

Reference data

Predicted data

	CA	CH	CZ	GB	GS	LC	TS	SUM	User accuracy
CA	12	1	1	0	0	1	1	16	75
CH	1	12	1	1	0	0	1	16	75
CZ	2	2	12	0	0	0	1	17	70.6
GB	0	0	0	8	1	1	0	10	80
GS	0	0	0	2	12	0	1	15	80
LC	2	1	0	1	1	10	2	17	58.8
TS	0	0	0	0	0	0	7	7	100
SUM	17	16	14	12	14	12	13	98	
Producer accuracy(%)	70.6	75	85.7	66.7	85.7	83.3	53.8		
Overall accuracy(%)		74.5							

By using the multiple endmember SAM approach, the effect of within-species spectral variability on the classification accuracy was minimised. This can be observed in the higher negative correlation between the intra-species SAM and producer's accuracy for the classification involving the mean spectra of the training sets are reference endmembers (r = 0.52) compared to the same relationship for the classification involving the multiple endmember approach (r = 0.33) (Fig.4.). The negative correlation between the intra-species SAM and producer's accuracy is lowest (r = 0.29) for the classification involving multiple endmembers selected after cluster analysis. Another factor which affect species discrimination is the interspecies similarity. For example, the high similarity between C. apiculatum and C. zeyheri, affected the classification accuracies for these species as a good number of spectra of each species were wrongly classified as the other (Table 1).

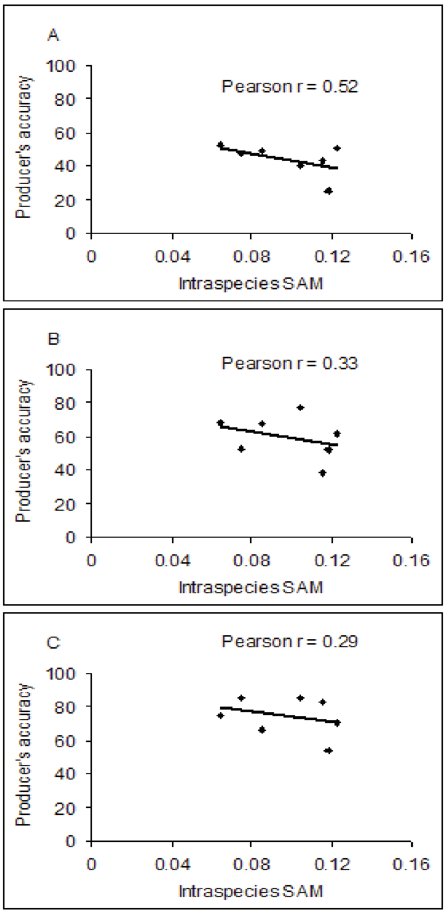


Fig.4. Effects of within-species similarity as determined from the spectral angle measure or mapper (SAM) on the classification accuracy for: **A.** using the mean spectrum of the training spectra for each species as reference (endmember) spectrum, **B.** using all training spectra for each species as reference spectra and C. using all training spectra selected after cluster analysis as reference spectra

4. DISCUSSION AND CONCLUSIONS

This study highlights three important phenomena; (i) within-species spectral variability affects the discrimination of savanna tree species at the leaf level with the SAM classifier, (ii) the effect of within-species spectral variability driven by differences in leaf phenology on the discrimination of savanna tree species can be minimised by adopting a multiple endmember approach with the SAM classifier or similarity measure and (iii) the classification accuracy is affected by the quality of the training endmembers. The results of this study suggest that each species multiple IDs or endmembers should be truly representative of the multiple groups existing within the population. Therefore, field sampling methods for collecting endmember spectra should ensure that the within-species variability is adequately captured. This is particularly relevant for the Kruger National Park given the high within-species variability across the landscape because of differences in rainfall and soil quality within relatively short distances. If a large training set is collected, it important to construct a small portion of them such that a high classification performance of the 1-nearest neighbour rule is achieved, hence the usual criticism of the k-nearest neighbour classifier in terms of space requirement to store the complete set of training data and a high computational cost for the evaluation of new targets.

5. ACKNOWLEDGMENTS

Our gratitude goes to the Council for Scientific and industrial Research (CSIR), South Africa for providing the funding for this study. Many thanks to A. Ramoelo, L. Naidoo, B. Majeke, R. Main. J.v Aardt , and C. Munyati for their collaboration.

6. REFERENCES

[1] J. B. Adams, M. O. Smith, and P. E. Johnson, "Spectral mixture modeling: a new analysis of rock and soil types at the Viking Lander 1 site.," *Journal of Geophysical Research*, vol. 98 B8, pp. 8098-8112, 1986.

[2] 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 (SIPS)--interactive visualization and analysis of imaging spectrometer data," *Remote Sensing of Environment*, vol. 44, pp. 145-163, 1993.

[3] E. L. Hestir, S. Khanna, M. E. Andrew, M. J. Santos, J. H. Viers, J. A. Greenberg, S. S. Rajapakse, and S. 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.

[4] Y. Du, C. I. Chang, H. Ren, C. C. Chang, and J. O. Jensen, "New hyperspectral discrimination measure for spectral characterisation," *Opt. Eng.*, vol. 43, pp. 1777-1786, 2004.

[5] J. T. Mundt, N. F. Glenn, K. T. Weber, T. S. Prather, L. W. Lass, and J. Pettingill, "Discrimination of hoary cress and determination of its detection limits via hyperspectral image processing and accuracy assessment techniques," *Remote Sensing of Environment*, vol. 96, pp. 509-517, 2005.

[6] D. A. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer, and R. O. Green, "Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. ," *Remote Sensing of Environment* vol. 65, pp. 267-279, 1998.

[7] M. Analoui and M. F. Amiri, "Feature reduction of nearest neighbor classifiers using genetic algorithm," presented at Proceedings of world academy of science, engineering and tehchnology, 2006.