

SPECTRAL VARIABILITY WITHIN SPECIES AND ITS EFFECTS ON SAVANNA TREE SPECIES DISCRIMINATION

Moses A. Cho¹, Pravesh Debba², Renaud Mathieu¹, Jan van Aardt³, Greg Asner⁴, Laven Naidoo¹, Russell Main¹, Abel Ramoelo¹, Bongani Majeke¹.

¹Council for Scientific and Industrial Research (CSIR), Natural Resources and the Environment, Ecosystems- Earth Observation, P.O. Box 395, Pretoria, 0001, South Africa.

²CSIR, Built Environment, Box 395, Pretoria, South Africa

³ RIT: Center for Imaging Science, Laboratory for Imaging Algorithms and Systems
54 Lomb Memorial Drive, Building 17-3173, Rochester NY 14623, USA

⁴Department of Global Ecology, Stanford University, 260 Panama St., Stanford CA 94305, USA

ABSTRACT

Differences in within-species phenology and structure driven by factors including topography, edaphic properties, and climatic variables present important challenges for species differentiation with remote sensing in the Kruger National Park, South Africa. The objective of this study was to examine probable factors including intraspecific spectral variability and the spectral sample size that could affect remote sensing of Savanna tree species across a land-use gradient in the Kruger National park. Eighteen species were examined: *Acacia gerrardii*, *Acacia nigrescens*, *Combretum apiculatum*, *Combretum collinum*, *Combretum hereroense*, *Combretum imberbe*, *Combretum zeyheri*, *Dichrostachys cinerea*, *Euclea* sp (*E. divinurum* and *E. natalensis*), *Gymnosporia* sp (*G. buxifolia* and *G. senegalensis*), *Lonchocarpus capassa*, *Peltoforum africanum*, *Piliostigma thonningii*, *Pterocarpus rotundifolia*, *Sclerocarya birrea*, *Strychnos* sp (*S. madagascariensis*, *S. usambarensis*), *Terminalia sericea* and *Ziziphus mucronata*. Discriminating species using the K-nearest neighbour (K = 1) classifier with spectral angle mapper (SAM) yielded a higher classification accuracy (48% overall accuracy) compared to 16% for the classification involving the mean spectra for each species as the training spectral set. Within-species spectral variability and the training sample size were identified as important factors affecting classification accuracy of the tree species. We recommend a non-parametric classifier such as K-nearest neighbour classifier for classifying and mapping tree species in a highly complex environment such as the savanna system of the Kruger National Park.

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 tissues 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 (imaging spectroscopy). 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]. This is often a problem associated with airborne imagery. 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 as applied in one of the most common image processing softwares, ENVI, 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. A single reference spectrum assumed to be the unique identity of each target is used for the classification. If region of interests (ROI) are selected to represent the spectral set for each class, the mean spectrum is computed for all the spectra in the ROI and used as the reference endmember. 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].

Within-species (intraspecies) structural and chemical variability create important challenges in mapping savanna tree species. For example, the savanna landscape across the Kruger National Park (KNP), South Africa and, neighbouring private game reserves and communal lands present varying vegetation structural characteristics albeit with similar species composition. The high intraspecies variability in the KNP region is caused by differences in edaphic conditions (e.g. presence of gabbro and granitic substrates within relatively short distances), rainfall, herbivory, bushfire and human activities. These factors do influence the vegetation structure (e.g. green leaf area index) and biochemical composition (e.g. differing chlorophyll and nutrient concentrations resulting from phenological differences).

The objective of this study was to examine probable factors including intraspecies spectral variability and the spectral sample size that could affect remote sensing of Savanna tree species across a land-use gradient in the Kruger National park. We have explored two classification approaches with spectral angle mapper: (i) using a spectral library composed of one spectrum (endmember) per species and (ii) a multiple endmember approach conventionally called K-nearest neighbour classifier.

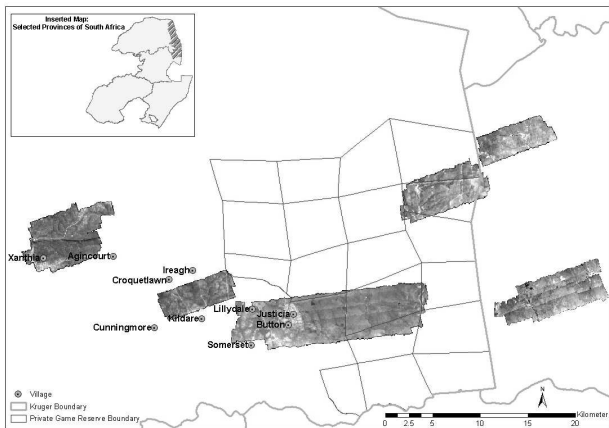


Figure 1. Study area showing CAO image scenes

2. METHODS AND MATERIAL

2.1. Study area

The study area is located in the “lowveld” savanna biome in the North-East of South Africa (Figure 1). Eight sites were chosen for the study including two sites in the KNP, two sites in private game reserves and four sites in communal lands. The species data used in this study consist of tree species generally more than 2 m tall identified and geo-registered using a Leica differential global positioning system (GPS). Eighteen dominant species are examined in the study. These include *Acacia gerrardii*, *Acacia*

nigrescens, *Combretum apiculatum*, *Combretum collinum*, *Combretum hereroense*, *Combretum inberbe*, *Combretum zeyheri*, *Dichrostachys cinerea*, *Euclea sp* (*E. divinurum* and *E. natalensis*), *Gymnosporia sp* (*G. buxifolia* and *G. senegalensis*), *Lonchocarpus capassa*, *Peltoforum africanum*, *Piliostigma thonningii*, *Pterocarpus rotundifolia*, *Sclerocarya birrea*, *Strychnos sp* (*S. madagascariensis*, *S. usambarensis*), *Terminalia sericea* and *Ziziphus mucronata*.

2.2. Airborne imagery

Airborne hyperspectral data were acquired in May 2008 with the Carnegie Airborne Observatory (CAO) system for eight land use sites across the study area (Figure 1). The CAO hyperspectral system acquired the spectral information in the 384-1054 nm range at a spatial resolution of 1m. The data were atmospherically and geometrically corrected by the CAO. The species point map was overlaid on the imagery and the spectral profiles collected via region of interest tool in the ENVI software.

Table 1. Classification accuracy for 18 savanna species using (A) mean spectrum for each species training set as reference endmember and (B) all training spectra for each species as reference endmembers. A. *gerrardii*(AG), A. *nigrescens*(AN), C. *apiculatum*(CA), C. *collinum*(CC), C. *hereroense*(CH), C. *inberbe*(CI), C. *zeyheri*(CZ), D. *cinerea*(DC), *Euclea sp*(ED), *Gymnosporia sp*(GY), L. *capassa*(LC), P. *africanum*(PA), P. *thonningii*(PT), P. *rotundifolia*(PR), S. *birrea*(SB), *Strychnos sp*(SY), T. *sericea*(TS) and Z. *mucronata*(ZM).

A		AG	AN	CA	CC	CH	CI	CZ	DC	ED	GY	LC	PA	PT	PR	SB	SY	TS	ZM	sum	User accuracy(%)
AG	7	5	2	2	0	4	0	0	3	1	2	2	1	0	0	1	4	4	38	18.4	
AN	0	1	1	0	1	1	0	1	1	0	2	0	0	0	0	0	0	0	1	9	11.1
CA	1	1	0	0	1	1	3	2	2	1	0	1	0	0	0	6	0	0	1	20	0.0
CC	11	3	0	12	0	3	0	4	3	2	6	2	5	0	8	1	2	3	65	18.5	
CH	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	4	0	0	0	6	16.7
CI	2	3	1	0	1	3	0	1	1	0	1	0	0	0	0	6	0	0	0	19	15.8
CZ	0	0	1	0	0	3	15	0	0	0	0	1	0	1	12	1	4	0	38	39.5	
DC	0	1	2	1	1	1	0	0	4	0	0	0	0	3	1	0	0	0	14	0.0	
ED	4	1	0	2	0	1	1	3	6	0	3	0	0	0	8	5	1	2	37	16.2	
GY	1	9	1	3	2	8	5	5	3	13	4	3	0	0	20	0	22	2	101	12.9	
LC	1	0	0	3	2	2	5	2	0	0	1	0	0	0	1	1	9	0	27	3.7	
PA	1	3	1	1	0	0	0	1	2	0	0	0	0	1	0	0	0	0	10	0.0	
PT	0	0	0	18	0	0	0	0	13	2	2	2	11	0	12	1	3	0	64	17.2	
PR	1	1	3	0	0	0	2	0	4	0	1	0	14	0	0	2	0	0	28	50.0	
SB	0	2	0	1	1	0	1	0	2	3	2	0	0	0	1	0	2	0	1	16	6.3
SY	3	3	1	2	1	0	0	2	0	0	1	0	1	0	1	0	2	1	3	20	5.0
TS	0	0	0	0	0	1	0	0	1	0	0	0	0	0	2	0	0	0	4	0.0	
ZM	1	4	0	1	1	0	1	2	1	0	0	0	0	3	1	1	1	1	17	5.9	
sum	33	37	14	46	12	28	30	24	43	27	24	12	18	19	86	12	53	15	533		
Producer accuracy(%)	21.2	2.7	0.0	26.1	8.3	10.7	50.0	0.0	14.0	48.1	4.2	0.0	61.1	73.7	1.2	8.3	0.0	6.7			
Overall accuracy(%)	16																				

B		AG	AN	CA	CC	CH	CI	CZ	DC	ED	GY	LC	PA	PT	PR	SB	SY	TS	ZM	sum	User accuracy(%)
AG	12	4	1	2	1	2	0	1	0	1	0	0	0	4	2	1	1	0	32	37.5	
AN	3	18	1	0	1	2	0	0	2	2	0	2	0	0	3	0	1	0	35	51.4	
CA	1	0	4	0	1	0	2	0	1	0	0	0	0	0	0	0	0	0	9	44.4	
CC	1	0	0	26	1	2	0	3	5	1	2	3	3	0	3	0	0	0	2	52	50.0
CH	0	1	0	1	2	0	1	1	0	0	0	0	0	0	1	0	2	0	9	22.2	
CI	1	1	1	3	2	10	0	0	2	0	1	1	0	0	2	0	1	1	26	38.5	
CZ	0	0	2	0	0	0	20	1	0	0	0	2	0	0	10	0	2	0	37	54.1	
DC	5	1	0	1	0	3	0	6	3	0	0	0	0	0	1	3	3	0	26	23.1	
ED	3	2	0	2	1	2	0	4	23	0	2	0	0	0	3	0	0	0	1	43	53.5
GY	0	2	0	0	0	0	0	1	14	0	0	0	0	0	1	0	0	0	20	70.0	
LC	0	1	0	1	0	2	0	1	2	0	8	0	3	0	6	1	0	0	25	32.0	
PA	1	0	0	0	0	1	1	0	0	1	0	3	0	0	0	0	0	0	7	42.9	
PT	4	0	0	3	0	0	0	0	1	2	0	8	0	2	0	0	1	21	38.1		
PR	0	1	0	0	0	0	0	0	1	0	0	0	12	0	0	0	0	0	2	16	75.0
SB	0	3	2	4	1	2	2	4	2	5	0	2	1	45	1	4	1	4	1	81	55.6
SY	1	2	0	2	2	1	0	0	0	1	2	0	2	3	0	0	0	0	17	17.6	
TS	0	1	3	1	0	1	4	1	2	4	2	0	2	2	3	39	1	66	59.1		
ZM	1	0	0	0	0	0	0	1	0	0	1	0	0	0	3	0	0	5	11	45.5	
sum	33	37	14	46	12	28	30	24	43	27	24	12	18	19	86	12	53	15	533		
Producer accuracy(%)	36.4	48.6	28.6	56.5	16.7	35.7	66.7	25.0	53.5	51.9	33.3	25.0	44.4	63.2	52.3	25.0	73.6	33.3			
Overall accuracy(%)	48																				
Overall accuracy(%)	48																				

2.3. Data analysis

The spectral data for each species were randomly split into the training and test data in a 1:4 ratio. Subsequently, two types of reference endmember spectra were used to evaluate the effect of within-species spectral variability on the discrimination accuracy of the tree species; (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 (K-nearest neighbour classifier). 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. SAM was calculated between spectral pairs for the whole spectral range (384-1054 nm). A target spectrum was classified as species x based on the minimum SAM criterion between a reference spectrum of species x and the target spectrum. The overall, users and producers classification accuracies were subsequently, determined from the confusion matrices between the observed and predicted data.

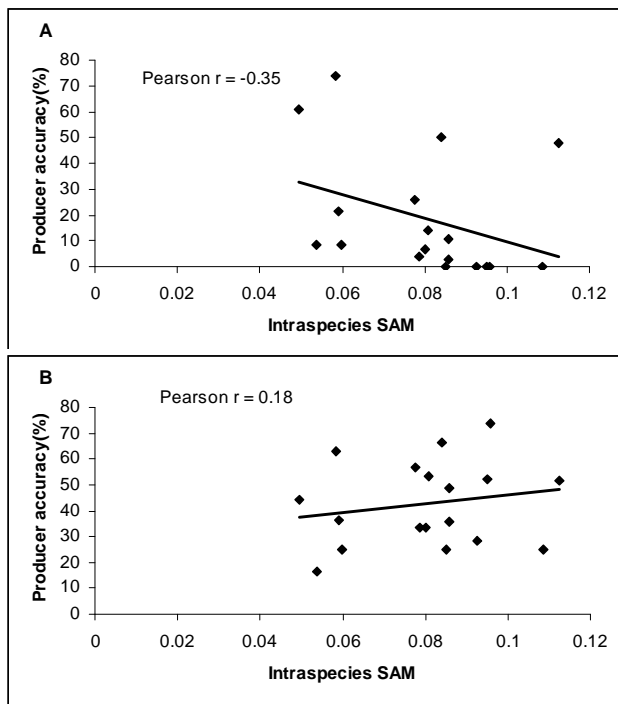


Figure 2. The effect of intraspecies dissimilarity based on intraspectral spectral angle measure (SAM) on the producer's accuracy for A. classification involving the mean spectra as reference endmembers and B. all training spectra per species as reference endmembers

3. RESULTS

The K-nearest neighbour ($k = 1$) classification provided a higher overall classification accuracy (48%) when compared to the classification involving the mean spectra of the training samples as

reference endmembers (overall accuracy = 16%) (Table 1). Seven of the eighteen species showed producer's accuracies of $\geq 50\%$ for the K-nearest neighbour classifier compared to three species for the classification involving the mean spectra as reference endmembers. Also, a higher number of species (8) showed user's accuracies of $\geq 50\%$ for the K-nearest neighbour classifier compared to the conventional SAM classifier (3 species).

The intraspecies SAM showed a negative but non-significant correlation ($r = -0.35$, $p = 0.15$) with the producer's accuracy for the classification involving the mean of the training spectra as reference endmembers (Figure 2). A significant negative correlation ($r = -0.58$, $p = 0.01$) was however observed in the above relationship when the analysis was conducted without including one outlier. A highly non-significant relationship ($r = 0.18$, $p = 0.47$) was observed between the producer's accuracy and intraspecies SAM for the classification involving K-nearest neighbour classifier (i.e. the multiple endmember approach).

The sample size of the training data showed a higher influence on the classification accuracy when compared to the intraspecies spectral variability or dissimilarity (Figure 3). A significant positive correlation ($r = 0.58$, $p = 0.01$) was observed between the producer accuracy and the training data size for the classification involving the K-nearest neighbour classifier. No such influence was observed for the classification based on the mean spectra of the training sets as reference endmembers.

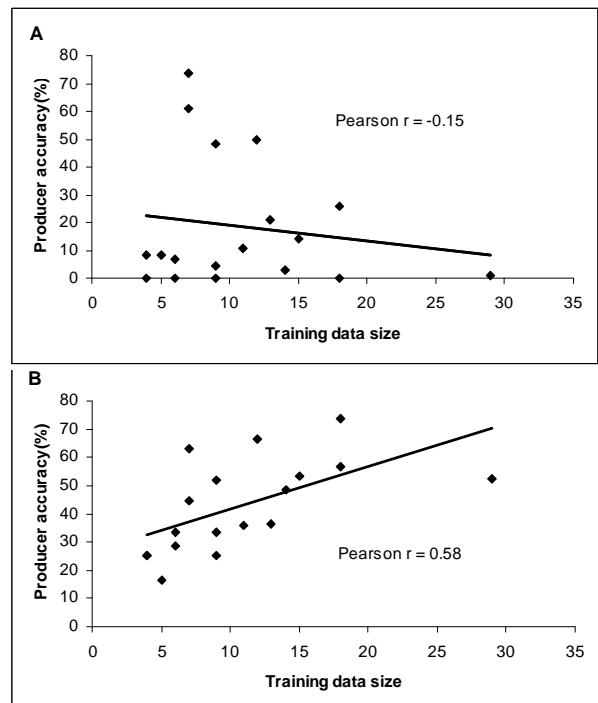


Figure 3. The effect of training data size on the producer's classification accuracy for A. classification involving the mean spectra as reference endmembers and B. all training spectra per species as training endmembers

4. DISCUSSION AND CONCLUSIONS

This study highlights two important factors that affect discrimination of savanna tree species, namely, intraspecies spectral variability and the training data sample size. The latter being an important factor for the K-nearest neighbour classifier. K= 1 nearest neighbours classifier was adopted in this study, which minimises computational cost, but the classification accuracy may improve with higher K values, particularly for classifications involving large training samples. The complexity of the Kruger National Park system requires the use of large training samples for each species. The increasing accuracy of classification with increasing training data size could imply that the classification of the species could be limited to the dominant species. The criteria to determine the dominant species have to be established.

We recommend the utilisation of the multiple endmembers SAM approach as opposed to the traditional SAM classifier involving single spectrum endmember per species for mapping of Kruger National Park species and emphasis that the training endmembers should be truly representative of the different distributions in the population. Higher K values for the nearest neighbour classifier with SAM may be required but the value of K must be as low as possible to minimise the computing space requirement to store the complete set of training data and a high computational cost for the evaluation of new targets. Lastly, the spectral data for this study was limited to the visible-near infrared. The utilisation of the full spectral range from the visible to the shortwave infrared needs to be assessed.

5. ACKNOWLEDGMENTS

Our gratitude goes to the Council for Scientific and industrial Research (CSIR), South Africa and to the Andrew W. Mellon Foundation for providing the funding for this study and to the South African National Parks Board (SANParks), the scientific services of the Kruger National Park in particular. Note of thanks also goes to Barend Erasmus and Jolene Fisher of Wits University and Konrad Wessels of the Meraka Institute.

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.