IEEE Transactions on Geoscience and Remote Sensing

Improving discrimination of savanna tree species through a multiple endmember spectral-angle-mapper (SAM) approach: canopy level analysis

Journal:	*Transactions on Geoscience and Remote Sensing*
Manuscript ID:	Draft
Manuscript Type:	Hyperspectral Image and Signal Processing Special Issue
Date Submitted by the Author:	
Complete List of Authors:	Cho, Moses; The Council for Scientific and Industrial Research (CSIR), Natural Resources and Environment, Ecosystem-Earth Observation unit Debba, Pravesh; Council for Scientific and Industrial Research, Built Environment Mathieu, Renaud; The Council for Scientific and Industrial Research (CSIR), Natural Resources and Environment, Ecosystem-Earth Observation unit Naidoo, Laven; The Council for Scientific and Industrial Research (CSIR), Natural Resources and Environment, Ecosystem-Earth Observation unit van Aardt, Jan; Rochester Institute of Technology, Center for Imaging Science Greg, Asner; Carnegie Institution, Dept of Global Ecology
Keywords:	Remote sensing, Vegetation mapping

Improving discrimination of savanna tree species through a multiple endmember spectral-angle-mapper (SAM) approach: canopy level analysis

Moses Azong Cho[1], Pravesh Debba[2] Renaud Mathieu[1], Laven Naidoo[1], Jan van Aardt[3] and Gregory Asner[4]*

[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
[3]RIT: Center for Imaging Science, Laboratory for Imaging Algorithms and Systems, 54 Lomb Memorial Drive, Rochester NY 14623, USA
[4]Carnegie Institution for Science, 260 Panama Street, Stanford, CA 94305, USA

mcho@csir.co.za, PDebba@csir.co.za, rmathieu@csir.co.za, LNaidoo@csir.co.za, vanaardt@cis.rit.edu, gpa@stanford.edu

Abstract - Differences in within-species phenology and structure are controlled by genetic variation, as well as topography, edaphic properties, and climatic variables across the landscape and present important challenges to species differentiation with remote sensing. The objectives of this paper were to (i) evaluate the classification performance of a multiple-endmember spectral angle mapper (SAM) classification approach (conventionally known as the nearest neighbour) in discriminating ten common African savanna tree species and (ii) compare the results with the traditional SAM classifier based on a single endmember per species. The canopy spectral reflectance of the tree species (*Acacia nigrescens*, *Combretum apiculatum, Combretum Imberbe, Dichrostachys cinerea, Euclea natalensis, Gymnosporia buxifolia, Lonchocarpus capassa, Pterocarpus rotundifolius, Sclerocarya birrea and Terminalia sericea)* were extracted from airborne hyperspectral imagery that was acquired using the Carnegie Airborne Observatory (CAO) system in the Kruger National Park, South Africa, in May 2008. This study highlights four important phenomena: (i) intra-species spectral variability affected the discrimination of savanna tree species with the SAM classifier, particularly the producer's accuracy, (ii) the effect of intra-species spectral variability was minimised by adopting a multiple endmember approach, (iii) the classification accuracy of the multiple endmember classifier was affected by the quality of the training endmembers, and (iv) targeted band selection improved be the classification of savanna tree species. We furthermore proposed bootstrapping as a method to obtain the best training subset for the classification.

Index Terms - *savanna tree species; spectral variability; multiple endmember approach; spectral angle mapper, hyperspectral remote sensing, band selection*

I. INTRODUCTION

The ability to map vegetation at the species level is of broad 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 is characterised by a set of unique biophysical features and biochemical composition that control the variability in its spectral signature. The advent of high spatial and spectral resolution imaging spectrometers, i.e., sensors that provide contiguous spectral data in narrow bands, has offered new opportunities for mapping vegetation at species-level, while also renewing demands for algorithm or methodological protocol development.

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 discriminant analysis, spectral mixture analysis (SMA) [1] and spectral angle mapper (SAM) [2]. The application of some of these methods, especially 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 insensitive to scaling, e.g., to differences in illumination or albedo [3, 4] and is therefore more appropriate for species-level monitoring at the regional scale compared to the Euclidean distance similarity measure. SMA, on the other hand, is a sub-pixel classifier that determines the relative abundance of materials that are depicted in multispectral or hyperspectral imagery based on the materials' spectral characteristics [5]. The reflectance for each pixel of the image is assumed to be a linear combination of the reflectance of each material (or endmember) present within the pixel.

Discriminant analysis (DA) is a commonly used supervised classification method with conventional multispectral data [6], [7]. DA searches for the linear combination of variables (spectral features) that best discriminates among classes. Both first order variations (e.g., mean values) and second order variations (e.g., covariance matrices) are considered in DA [7]. However, there is a limitation with the application of the linear classifier on hyperspectral data. When applied to hyperspectral data, a large number of training samples are required because of the high dimensionality of hyperspectral data. Although the problem of high data dimensionality with hyperspectral data is more severe with non-parametric classifiers such as the nearest neighbour classifier than with parametric classifiers [7], the former has the advantage that it makes no distribution assumptions regarding independent variables [8]. This makes the non-parametric classifier an attractive alternative in the context of high intra-species spectral variability.

Spectral Angle Mapper (SAM) is a non-parametric spectral classification approach that uses an n-D angle to match pixels to reference spectra. The conventional SAM classifier used for species discrimination [9] compares the angle between the reference endmember spectrum vector and each target spectral vector in n-D space. A target spectrum is classified in a group or class based on the minimum SAM criterion between a reference

spectrum and the unknown target spectrum (conventionally called nearest neighbour classifier). In the conventional SAM classification, each species or vegetation class is assumed to have a unique spectral identity or signature [10], [11], [12], [13]. Typically, the endmembers are extracted from a species spectral library. The assumption of a unique spectral identity per species means that spectral variability within each species, denoted as the intra-species variability, is not preserved [10]. This assumption is challenged by the fact that reflectance from vegetation is controlled by a number of biochemical and biophysical parameters that vary over space and time [14]. Differences in phenology and plant structure are driven by factors such as topography, edaphic properties, and climatic variables across the landscape. For example, high intra-species spectral variability within the southern region of the Kruger National Park (KNP), South Africa, is governed by differences in leaf phenology across an east-west direction as a result of a rainfall gradient and structural differences, driven by factors such bush fire and herbivory between conserved and subsistence farming, communal lands [14].

We argue in this paper that the ability of the SAM classifier to classify species with high intra-species variability is weakened when one does not consider the variability around the means of the reference endmember spectra e.g. SAM showed that the lowest performance for discriminating rainforest species compared to linear discriminant analysis and maximum likelihood classifiers [10]. Our research hypothesis therefore centres on the fact that a multiple-endmember approach, involving many endmember spectra per class, would provide higher classification accuracies when compared to the conventional SAM classifier involving a single endmember spectrum per class for discriminating tree species. [3] attempted the multiple endmember SAM approach by dividing water hyacinth into two spectral classes according to phenological classes, in their study on identification of invasive vegetation using hyperspectral data. The multiple endmember SAM is similar to the *k*-nearest neighbours classifier. In the *k*-nearest neighbours classifier, an object is classified by a majority vote of its neighbours, with the object being assigned to the class most common amongst its *k*-nearest neighbours (*k* is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of its nearest neighbour. However, in the *k*-nearest neighbours classifier, often applied using the Euclidean distance measure [15], [8],16], a larger portion of the data is used for training and the remainder for validation of the classification [8], [15]. The *k*-nearest neighbour classifier has been criticised as costly in terms of computer memory space requirement to store the complete set of training data and a high computational time for the evaluation of new targets [17], [16]. The question is whether a small number of representative reference endmembers per species (multiple identities per species) can be constructed for the training of the SAM *k*-nearest neighbours classifier ($k = 1$) in contrast to using one endmember per species and the traditional *k*-nearest neighbour classifier that uses a larger proportion of the spectral data as the training set and $k \geq 1$, hence, the naming; multiple endmember SAM.

The high computational requirement of the nearest neighbour classifier might be even more acute when applied on hyperspectral data which consist of hundreds of bands. Furthermore, it has been demonstrated that as the number of dimensions increases, the sample size needs to increase exponentially in order to have an effective estimate of multivariate densities [18]. Others have suggested that feature extraction or selection algorithms are important to find the lower dimensional space in which the most important

discriminant structure exists [7]. We have assessed the utility of two feature (band) selection protocols in this study for the discrimination of savanna tree species. One is built on plant biological features, such as absorption features of biochemicals, while the other is a mathematical procedure that is based on bands that maximise inter-species SAM described by [19]. The overall aim of the study is to suggest a protocol for the application of SAM in discriminating Savanna tree species with hyperspectral data.

Insert Table 1

II. METHODS AND MATERIAL

A. Tree image spectra

Image-based canopy spectra of ten common tree species (Table 1) in the Kruger National Park (31°37'7.32"E, 24°49'38.13"S and 31°20'32.37"E, 24°50'47.67"S), South Africa, were used in this study. The study area is located in the "lowveld" savanna biome in the northeast South Africa. Eight sites were chosen for the study, including two sites in the Kruger National Park, two sites in private game reserves, and four sites in an adjacent communal-subsistence farming area. The species data consisted of tree species generally more than 2 m tall, identified and geo-registered using a Leica differential global positioning system along several transects in seven out of the eight sites. Airborne hyperspectral data were acquired in May 2008 with the Carnegie Airborne Observatory (CAO) system [20]. The data were atmospherically and geometrically corrected by the CAO research team. The CAO system used in this study consisted of (i) a high-fidelity imaging spectrometer (HFIS), (ii) a discrete return light detection and ranging (LiDAR) scanner, and (iii) a global positioning system-inertial measurement unit (GPS-IMU). The pushbroom HFIS sampled the scenes in the visible-near infrared (VNIR) spectral region between 384.8-1054.3 nm (72 bands) at approximately 9.2 nm spectral resolution (full-with-half-maximum) and a spatial resolution of 1 m.

A tree mask consisting of trees of more than 2m was built using a tree height map produced from the discrete return LiDAR imagery. The tree mask subsequently was used to subset the CAO hyperspectral imagery. The species point map was overlaid on the CAO tree imagery and the spectral profiles of various tree pixels were collected via the region of interest tool in ENVI software. The following number of spectral profiles were collected per species: 54 *Acacia nigrescens*, 26 *Combretum apiculatum,* 88 *Combretum Imberbe,* 34 *Dichrostachys cinerea,* 46 *Euclea natalensis,* 21 *Gymnosporia buxifolia,* 36 *Lonchocarpus capassa,* 35 *Pterocarpus rotundifolius,* 116 *Sclerocarya birrea*, *and* 71 *Terminalia sericea*.

B. Band selection

It has been shown [19] that superior classification results can be obtained from a subset of the hyperspectral bands through band selection. The method for selecting the most useful bands that discriminate the various species is described by [19]. It uses the spectral angle mapper (SAM) as a distance measure and starts by selecting the two bands that have the highest spectral angle on average among all mean species spectral signatures for the various classes. SAM is defined as:

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

where L is the number of bands, $S_i = S_{i1,}\ldots, S_{iL}$ and $S_i = S_{j1,}\ldots, S_{jL}$. In this section, for the SAM classifier as defined in Eq. 1, S_i is the mean reflectance spectrum for species i and S_j is the mean reflectance spectrum for species j.

The procedure is to then add bands sequentially as being the next most important for discriminating the various species until no bands contribute further to the discriminatory power of these classes. This method is described in detain in [19] as Band Add-on (BAO) procedure. Essentially, for spectral signatures S_i and S_i, we can partition $S_i = [S_i^a, S_i^b]$ and $S_j = [S_j^a, S_j^b]$, where L=a+b. Then

$$SAM(s_i, s_j) = \cos^{-1}\left(\frac{\sum_{l=1}^{L} s_{il} s_{jl}}{\left[\sum_{l=1}^{L} s^2_{il}\right]^{1/2}\left[\sum_{l=1}^{L} s^2_{jl}\right]^{1/2}}\right) = \cos^{-1}\left(\frac{\sum s^a_{il} s^a_{jl} + \sum s^b_{il} s^b_{jl}}{\left[\sum (s^a_{il})^2 + (s^b_{jl})\right]^{1/2}\left[\sum (s^a_{il})^2 + (s^b_{il})\right]^{1/2}}\right) \quad (2)$$

$$= SAM(s_i^a, s_j^a)\cos^{-1}\left(\frac{1 + \frac{\sum s^b_{il} s^b_{jl}}{\sum s^a_{il} s^a_{jl}}}{\left[1 + \frac{\sum (s^b_{il})^2}{\sum (s^a_{il})^2}\right]^{1/2}\left[1 + \frac{\sum (s^b_{jl})^2}{\sum (s^a_{jl})^2}\right]^{1/2}}\right) = SAM(s_i^a, s_j^a)\cos^{-1}(\beta)$$

The calculation of β is based on the remaining bands and the iterative procedure is repeated until no bands satisfy $\beta \le 1$.

Insert Fig. 1

C. Data analysis

A bootstrapping procedure was adopted to select the training and test data sets from the spectral data (Fig. 1). One-third and two-thirds of the data were used for the training and test sets, respectively. Twenty replicates for the training and test data per species 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 using the SAM classifier (Eq. 1), (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. In this section, for the SAM classifier as defined in Eq. 1, S_i is the reference spectrum and S_j is the target spectrum.

SAM was calculated between spectral pairs for:

(i) the full VNIR spectral range considering the full spectral range (394-1054 nm),

(ii) wavebands of known spectral features of biochemical and biophysical properties at 432 nm and 460 nm (chlorophyll), 507 nm, 536 nm, 556 nm, and 574 nm (xanthophylls), 640 nm and 696-744 nm (chlorophylls and leaf area index), and 970-989 nm (leaf water) [21], [22], [23], [24], [25], and

(iii) band selection based on a mathematical process as described in section C above.

The producer's-, user's-, overall accuracies, and kappa (k) (Eq. 2) scores were used as measures of accuracy for each bootstrapped iteration. Kappa is defined as:

$$\kappa = \left(\frac{p_0 - p_e}{1 - p_e} \right) \tag{2}$$

where P_o is the observed proportion of agreement between the observed vs. predicted outcomes and P_e is the expected proportion of agreement. The value of kappa range from -1 to +1, with -1 indicating perfect disagreement and +1 indicating perfect agreement between the observed and predicted classes.

Insert Fig. 2 and Fig. 3

III. RESULTS

A. Intra-species variability

There was a characteristic pattern of spectral variability from the visible to the NIR (VNIR) spectral regions for all the species (Fig. 2). The lowest and highest coefficients of variation in the VNIR were observed in the blue region (400-460 nm) and around the chlorophyll absorption centre (660-685 nm), respectively (Fig. 2c). The spectral variability decreased from the chlorophyll absorption centre to the red-edge region (695-743 nm) before increasing again in the NIR. A slight decrease in the coefficients of variation was observed after 970 nm, a region associated with leaf water content [21]. The band selection procedure based on the biochemical and biophysical properties of the vegetation was further refined by the observed intra-species spectral variability. The spectral region around the chlorophyll absorption centre was not considered based on the assumption that the high within species variability in this region will impair species separability. The final band selection included 432 nm, 460 nm, 507 nm, 536 nm, 556 nm, 574 nm, 640 nm, 696-744 nm, and 970-989 nm.

High intra-species spectral variability was observed for all ten species under study (Fig. 2 & 3). Nevertheless, there were detectable differences among the species. The lowest intra-species spectral variability were observed in *G. buxifolia* (an evergreen), *P. rotundifolia* (deciduous, but known for its drought resistance and dark green leaves), and *E. natalensis* (an evergreen), as can be seen in Fig. 2 for the coefficient of variation and Fig. 3 for the intra-species SAM. Three deciduous trees, namely *C. apiculatum*, *S. birrea*, and *T. sericea* exhibited the highest variability among the tree species.

Insert Fig 4 and Table 2

B. Classification accuracy

The application of the multiple endmember SAM approach increased the producer's and user's accuracies when compared with the conventional SAM, based on the mean spectra of the training sets (Fig. 4). The two species that showed the lowest intra-species SAM, *G. buxifolia* and *P. rotundifolius*, were the most accurately classified species using the mean spectra of the training set (Fig. 4). The producer's accuracy for the classification based on the mean spectra of the training sets increased with increasing intra-species SAM, with a significant negative Pearson correlation (Pearson $r = 0.81$, $p < 0.0001$) (Fig. 5). The above relationship was not significant for the multiple endmember SAM approach. The high producer's and user's accuracies of the multiple endmember approach caused a significantly ($p < 0.05$) higher overall classification performance (overall percent accuracy = 54.47% ± 3.19 CI; 95% confidence interval) for this approach when compared with the SAM classifier involving the mean spectra of the training set (overall accuracy = 20.47% ± 0.94 CI) (Table 2). The kappa statistic yielded similar trends (Table 2).

Insert Fig. 5 and Table 3

C. The effect of band selection on the classification accuracy

Although the SAM classification involving the full range of bands in the VNIR yielded a higher overall accuracy (mean overall accuracy = 54.47%) when compared to those involving bands of known spectral features of biochemical and biophysical properties (mean overall percent accuracy = 53.15%), the difference was not statistically significant (Table 3). Of the 72 bands, 30 were selected through the mathematical procedure, in the following in the order of importance: 706.0 nm, 762.7 nm, 696.6 31 nm, 668.2 nm, 677.7 33 nm, 687.1 nm, 715.5 nm, 724.9 nm, 734.4 nm, 743.8 nm, 753.3 nm, 384.8 nm, 394.3 nm, 403.7 nm, 413.1 nm, 422.6 nm, 913.5 nm, 819.3 nm, 828.8 nm, 838.2 nm, 847.6 nm, 857.0 nm, 866.5 nm, 875.9 nm, 885.3 nm 894.7 nm, 904.1 nm, 1016.8 nm, 922.9 nm, 932.3 nm, and 941.7 nm. Specific contiguous spectral regions could be identified, viz, blue (384.8-422.6 nm), red-edge (668.2-762.7 nm), and NIR (819.3-904.1 nm, 913.5-941.7 nm and 1016.8 nm). The red-edge region was determined to be the most important spectral range for the discrimination of the tree species. The bands selected via the mathematical procedure produced the highest overall accuracies (overall percent accuracy = 0.57%, kappa = 0.50) among the different bands sets (Table 3). Finally, the difference between the results for all bands and bands selected by the mathematical procedure was statistically significant ($p < 0.001$).

IV. DISCUSSION AND CONCLUSIONS

The results of this canopy-level study generally corroborated findings from our earlier leaf-level study [14]. This study highlights four important phenomena: (i) intra-species spectral variability affects the discrimination of savanna tree species with the SAM classifier, (ii) the effect of intra-species spectral variability on the discrimination of

savanna tree species can be reduced by adopting a multiple endmember (nearest neighbour) approach, (iii) the classification accuracy is affected by the representativeness of the training endmembers, and (iv) physiology-based band selection might be useful for classification of savanna tree species.

Intra-species variability was higher for deciduous and less drought resistant tree species than for evergreen trees or deciduous, but drought resistant species (e.g., *P. rotundifolius*). The images used in this study were collected in the month of May when most of the trees in the region were already in an advanced stage of senescence accompanied by leaf shedding. Therefore, background reflectance might have increased the confusion between tree species, particularly for the deciduous species, given the differences in the stage and rate of leaf senescence across the region. It is recommended that, if single-date imagery is to be used for species mapping, it should be acquired at an optimal time when there is minimal intra-species difference in leaf biochemical and biophysical features, but maximum differences between species. Such an optimal time can be determined from analysis of multi-temporal spectral data. Alternatively, global vegetation classifications have relied on seasonal changes in multi-temporal data [26], [27], [28]. However, seasonal dynamics have rarely been explored with hyperspectral data [29], [30]. This is disconcerting, given that other studies have demonstrated the importance of seasonal variations in the spectral response of chaparral in discriminating such species [26], [29]. We recommend further research along the avenue of multi-temporal assessment for savanna tree species.

A bootstrapping approach was adopted in this study for choosing the training spectra. The best band combination yielded an overall percent accuracy of 64.07% (or kappa = 0.58) for the multiple endmember SAM classifier. However, the high variance of the overall percent accuracy (standard deviation = 6.82% or range = 19.6%), suggested that the quality of the training sample has a distinct impact on classification results. Therefore, it is important that multiple endmembers for each species should be truly representative of the intra- and inter-species variation (e.g., edaphic, topographic, age, etc. variation) that exists for the population. Field sampling methods for collecting endmember spectra should ensure that the intra-species variability is adequately captured. This is relevant particularly for the Kruger National Park, given the high intra-species variability across the landscape due to differences in rainfall and soil quality within relatively short distances. If a large training set is collected, it is furthermore important to sub-select a small portion of them, such that a high classification performance of the nearest neighbour rule is achieved. This is necessary in order to minimise computer space requirements for storing the complete set of training data and the high computational cost for the evaluation of new targets. In addition to an adequate field sampling approach, the bootstrapping approach can be adopted to obtain the best training subset for the classification.

Band selection improved the performance of the multiple endmember SAM classifier. Bands selected by the mathematical approach that maximise inter-species SAM proved superior to bands based on known biochemical and biophysical spectral properties. The red-edge region, controlled by both leaf chlorophyll amounts and leaf mass or stacking, appeared to be the most important region for discriminating between the tree species. It should be noted that the CAO spectral data was limited to the VNIR. Inclusion of the shortwave infrared bands (1200-2500 nm) might provide additional bands that could

improve classification, especially in cases where inter-species differences in leaf moisture regimes exist. Finally, minimisation of the spectral dimension reduces the constraints imposed by a large sample size towards the application of a non-parametric classifier, e.g., SAM in the case of hyperspectral data [7], [18].

This study has demonstrated the need for proper evaluation of endmember spectral variability when classifying tree species in a savanna environment. We believe that a multi-temporal approach will be essential to improve classification results in this environment, albeit with keen consideration of the aforementioned spectral variability.

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 Dr. Izak Smith, A. Ramoelo, B. Majeke, R. Main, Dr. Barend Erasmus, Dr. Konrad Wessels, and Dr. C. Munyati for their field efforts and research inputs.

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] P. E. Dennison and D. A. Roberts, "Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE," *Remote Sensing of Environment*, vol. 87, pp. 123-135, 2003.

[6] P. S. Thenkabail, E. A. Enclona, M. S. Ashton, and B. Van Der Meer, "Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications," *Remote Sensing of Environment*, vol. 91, pp. 354-376, 2004.

[7] D. Landgrebe, "On information extraction principles for Hyperspectral data: A white paper," School of Electrical and Computer Engineering, Purdue University, West Lafayette IN 47907-1285, 25.07.1997 1997.

[8] S. Thessler, S. Sesnie, Z. S. Ramos Bendaña, K. Ruokolainen, E. Tomppo, and B. Finegan, "Using k-nn and discriminant analyses to classify rain forest types in a Landsat TM image over northern Costa Rica," *Remote Sensing of Environment*, vol. 112, pp. 2485-2494, 2008.

[9] M. Sobhan, "Species discrimination from hyperspectral perspective," in *Natural Resources*, vol. PhD. Enschede: ITC, 2007.

[10] M. L. Clark, D. A. Roberts, and D. B. Clark, "Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales," *Remote Sensing of Environment*, vol. 96, pp. 375-398, 2005.

[11] B. W. Pengra, C. A. Johnston, and T. R. Loveland, "Mapping an invasive plant, Phragmites australis, in coastal wetlands using the EO-1 Hyperion hyperspectral sensor," *Remote Sensing of Environment*, vol. 108, pp. 74-81, 2007.

[12] E. Belluco, M. Camuffo, S. Ferrari, L. Modenese, S. Silvestri, A. Marani, and M. Marani, "Mapping salt-marsh vegetation by multispectral and hyperspectral remote sensing," *Remote Sensing of Environment*, vol. 105, pp. 54-67, 2006.

[13] 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.

[14] M. A. Cho, R. Mathieu, and P. Debba, "Multiple endmember spectral-angle-mapper (SAM) analysis improves discrimination of savanna tree species."

presented at Proceedings of 2009 IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Grenoble, France, 2009.

[15] F. J. Triepke, C. K. Brewer, D. M. Leavell, and S. J. Novak, "Mapping forest alliances and associations using fuzzy systems and nearest neighbor classifiers," *Remote Sensing of Environment*, vol. 112, pp. 1037-1050, 2008.

[16] 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.

[17] P. Paclík and R. P. W. Duin, "Dissimilarity-based classification of spectra: computational issues," *Real-Time Imaging*, vol. 9, pp. 237-244, 2003.

[18] J. Hwang, S. Lay, and A. Lippman, "Nonparametric multivariate density estimation: A comparative study," presented at IEEE Transactions on Signal Processing, 1994.

[19] N. Keshava, "Angle-Based Band Selection for Material Identification in Hyperspectral Processing," presented at Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX. Proceedings of SPIE, vol 5039, 2003

[20] G. P. Asner, D. E. Knapp, T. Kennedy-Bowdoin, M. O. Jones, R. E. Martin, J. Boardman, and R. F. Hughes, "Invasive species detection in Hawaiian rainforests using airborne imaging spectroscopy and LiDAR," *Remote Sensing of Environment*, vol. 112, pp. 1942-1955, 2008.

[21] P. J. Curran, "Remote sensing of foliar chemistry," *Remote Sensing of Environment*, vol. 30, pp. 271-278, 1989.

[22] J. A. Gamon, J. Penuelas, and C. B. Field, "A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency," *Remote Sensing of Environment*, vol. 41, pp. 35-44, 1992.

[23] G. A. Blackburn, "Quantifying Chlorophylls and Caroteniods at Leaf and Canopy Scales: An Evaluation of Some Hyperspectral Approaches," *Remote Sensing of Environment*, vol. 66, pp. 273-285, 1998.

[24] A. A. Gitelson, Y. Zur, O. B. Chivkunova, and M. N. Merzlyak, "Assessing carotenoid content in plant leaves with reflectance spectroscopy," *Photochemistry and Photobiology*, vol. 75, pp. 272-281, 2002.

[25] M. A. Cho and A. K. Skidmore, "A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method," *Remote Sensing of Environment*, vol. 101, pp. 181-193, 2006.

[26] P. E. Dennison and R. D.A., "The effects of vegetation phenology on endmember selection and species mapping in southern California chaparral," *Remote Sensing of Environment*, vol. 87, pp. 295-309, 2003.

[27] R. S. DeFries and J. R. G. Townshend, "NDVI-derived land cover classifications at a global scale," *International Journal of Remote Sensing*, vol. 15, pp. 3567-3586, 1994.

[28] T. R. Loveland, B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Yang, and J. W. Merchant, "Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. ," *International Journal of Remote Sensing*, vol. 21, pp. 1303-1330, 2000.

[29] D. A. Roberts, R. O. Green, and J. B. Adams, "Temporal and spatial patterns in vegetation and atmospheric properties from AVIRIS," *Remote Sensing of Environment*, vol. 62, pp. 223-240, 1997.

[30] M. Garcia and S. L. Ustin, "Detection of interannual vegetation responses to climatic variability using AVIRIS data in a coastal savanna in California. ," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 1480– 1490, 2001.

[31] E. Schmidt, M. Lotter, and W. McCleland, *Trees amd shrubs of Mpumalanga and Kruger National Park*. Johannesburg, South Africa: Jacana Media, 2002.

List of Tables

TABLE 1
TREE SPECIES ATTRIBUTES [31]

TABLE 2
STATISTICS FOR THE OVERALL SAM CLASSIFICATION ACCURACY OF TEN SAVANNA SPECIES FOR 20 BOOTSTRAPPED SAMPLES OF THE TRAINING (1/3) AND TEST (2/3) DATA, USING THE MEAN SPECTRA OF THE TRAINING SPECTRA OF EACH SPECIES AS REFERENCE ENDMEMBERS AND ALL TRAINING SPECTRA OF EACH SPECIES AS REFERENCE ENDMEMBERS.

TABLE 3
EFFECT OF BAND SELECTION ON THE OVERALL CLASSIFICATION ACCURACY FOR THE MULTIPLE ENDMEMBER SAM CLASSIFIER OF TEN SAVANNA SPECIES FOR 20 BOOTSTRAPPED SAMPLES OF THE TRAINING (1/3) AND TEST (2/3) DATA

TABLE 1
TREE SPECIES ATTRIBUTES [31]

Tree species	Characteristic features	Phenology
Acacia nigrescens (Knob thorn)	Medium to large tree to 30 m. Twice-compound leaves. Important browsing tree for game	Deciduous
Combretum apiculatum (Red bushwillow)	Small to medium tree, 3-9 m. Broadly ovate leaves.	Deciduous
Combretum imberbe (leadwood)	Medium to large tree, 7-20 m, with a spreading canopy. Obovate to oval small leaves (25-60 x 10-30 mm). Leaves grey-green above, distinctly paler beneath, giving tree a greyish appearance	Deciduous
Dichrostachys cinerea (small-leaved sickle-bush)	Shrub or small rounded tree to 7 m. twice-compound leaves, clustered on side shoot.	Deciduous
Gymnosporia buxifolia (Common Spikethorn)	Small erect tree usually 3-4 m. alternate and clustered leaves	Evergreen
Euclea natalensis (Hairy guarri)	Shrub to medium-sized tree, 2-10 m. Obovate-oblong (60-130 x 15-40 mm) leaves.	Evergreen
Lonchocarpus capassa (Apple leaf)	Rounded tree, to 18 m, large Imparipinnate compound leaves, glossy green above and greyish-green below.	Semi-deciduous
Pterocarpus rotundifoilius (rounded bloodwood)	A large round, woody shrub or tree to 10 m. leaves compound imparipinnate, dark green above and paler below. Tree is noted for withstanding drought and remains conspicuously green when other trees are already senesced. Over grazing causes coppicing.	Deciduous
Sclerocarya birrea (Marula)	Medium to large tree, to 18 m. leaves clustered at tips of branches, 3-7 terminal ovate leaflets.	Deciduous
Terminalia sericea (silver cluster leaf)	Small to medium-sized tree, 4-7 m. leaves crowded at end of branches, upper surface bluish-green, distinctively paler below. Foliage diagnostically blue-grey at a distance; densely covered in silvery hairs.	Deciduous

TABLE 2
STATISTICS FOR THE OVERALL SAM CLASSIFICATION ACCURACY OF TEN SAVANNA SPECIES FOR 20 BOOTSTRAPPED SAMPLES OF THE TRAINING (1/3) AND TEST (2/3) DATA, USING THE MEAN SPECTRA OF THE TRAINING SPECTRA OF EACH SPECIES AS REFERENCE ENDMEMBERS AND ALL TRAINING SPECTRA OF EACH SPECIES AS REFERENCE ENDMEMBERS.

Endmember description	Minimum		Maximum		Mean		Standard deviation	
	%	Kappa	%	Kappa	%	Kappa	%	Kappa
Mean of reference endmember spectra	16.08	0.08	23.87	0.15	20.47	0.13	2.06	0.02
All reference endmember spectra	44.47	0.36	64.07	0.58	54.47	0.48	6.82	0.08

TABLE 3
EFFECT OF BAND SELECTION ON THE OVERALL CLASSIFICATION ACCURACY FOR THE MULTIPLE ENDMEMBER SAM CLASSIFIER OF TEN SAVANNA SPECIES FOR 20 BOOTSTRAPPED SAMPLES OF THE TRAINING (1/3) AND TEST (2/3) DATA

Spectral band selection	Minimum		Maximum		Mean		Standard deviation	
	%	Kappa	%	Kappa	%	Kappa	%	Kappa
All VNIR spectral bands	44.47	0.36	64.07	0.58	54.47	0.48	6.82	0.08
Known spectral features of plants	39.95	0.30	62.31	0.56	53.15	0.46	6.38	0.07
Bands selected via a mathematical procedure	0.47	0.39	0.65	0.59	0.57	0.50	0.06	0.07

List of Figures

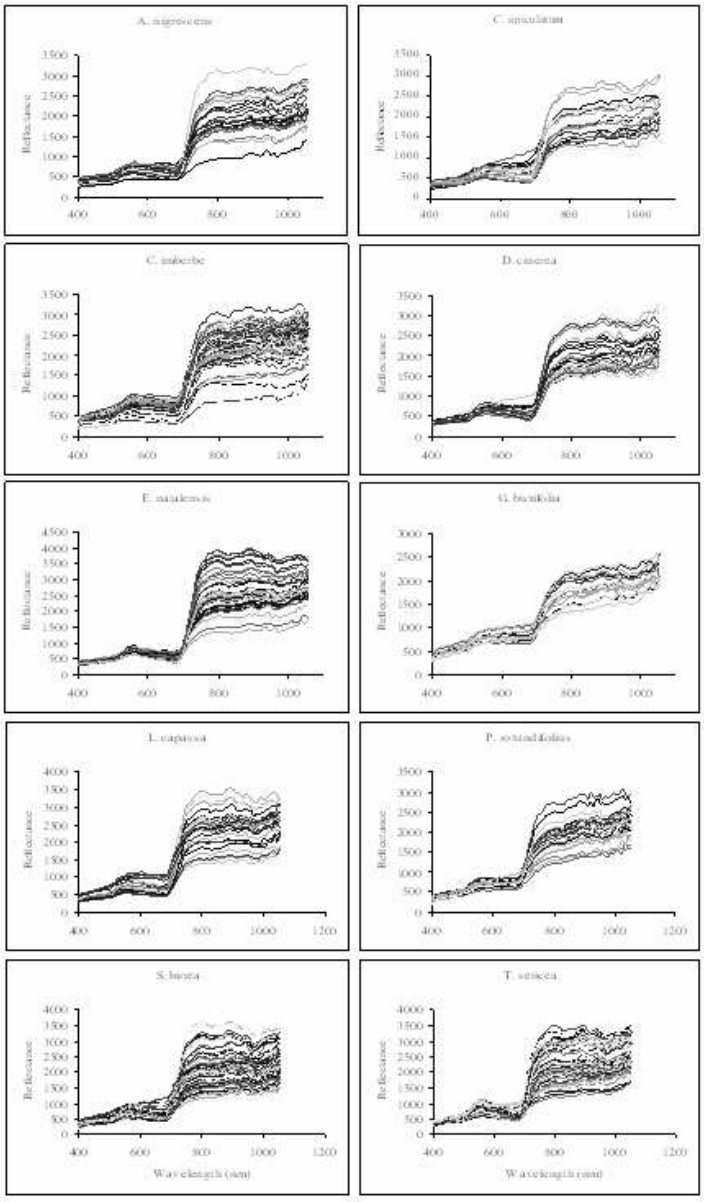
Fig. 1. The spectral profiles of the ten tree species used in the study: 54 *Acacia nigrescens*, 26 *Combretum apiculatum,* 88 *Combretum Imberbe,* 34 *Dichrostachys cinerea,* 46 *Euclea natalensis,* 21 *Gymnosporia buxifolia,* 36 *Lonchocarpus capassa,* 35 *Pterocarpus rotundifolius,* 116 *Sclerocarya birrea*, *and* 71 *Terminalia sericea*

Fig. 2. The mean spectra (A), standard deviation (B) and coefficient of variance of the ten tree species used in the study: *Acacia nigrescens* (AN), *Combretum apiculatum* (CA), *Combretum Imberbe* (CI), *Dichrostachys cinerea* (DC), *Euclea natalensis* (EN), *Gymnosporia buxifolia* (GB), *Lonchocarpus capassa* (LC), *Pterocarpus rotundifolius* (PR), *Sclerocarya birrea* (SB)*, and Terminalia sericea* (TS)

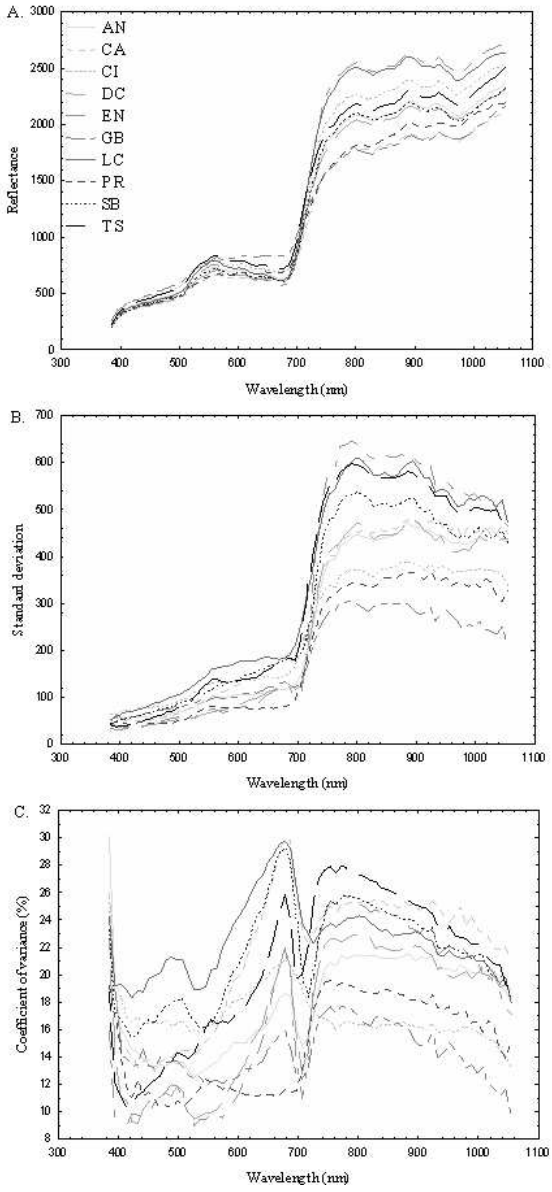
Fig. 3. Intra-species similarity derived from intra-species SAM of ten tree species: *Acacia nigrescens* (AN), *Combretum apiculatum* (CA), *Combretum Imberbe* (CI), *Dichrostachys cinerea* (DC), *Euclea natalensis* (EN), *Gymnosporia buxifolia* (GB), *Lonchocarpus capassa* (LC), *Pterocarpus rotundifolius* (PR), *Sclerocarya birrea* (SB)*, and Terminalia sericea* (TS)

Fig. 4. Producer's (left) and user's accuracy (right) of ten savanna species for 20 bootstrapped samples of the training (1/3) and test (2/3) data, using the mean spectra of the training spectra of each species as reference endmembers (A) and all training spectra of each species as reference endmembers (B). *Acacia nigrescens* (AN), *Combretum apiculatum* (CA), *Combretum Imberbe* (CI), *Dichrostachys cinerea* (DC), *Euclea natalensis* (EN), *Gymnosporia buxifolia* (GB), *Lonchocarpus capassa* (LC), *Pterocarpus rotundifolius* (PR), *Sclerocarya birrea* (SB)*, and Terminalia sericea* (TS)

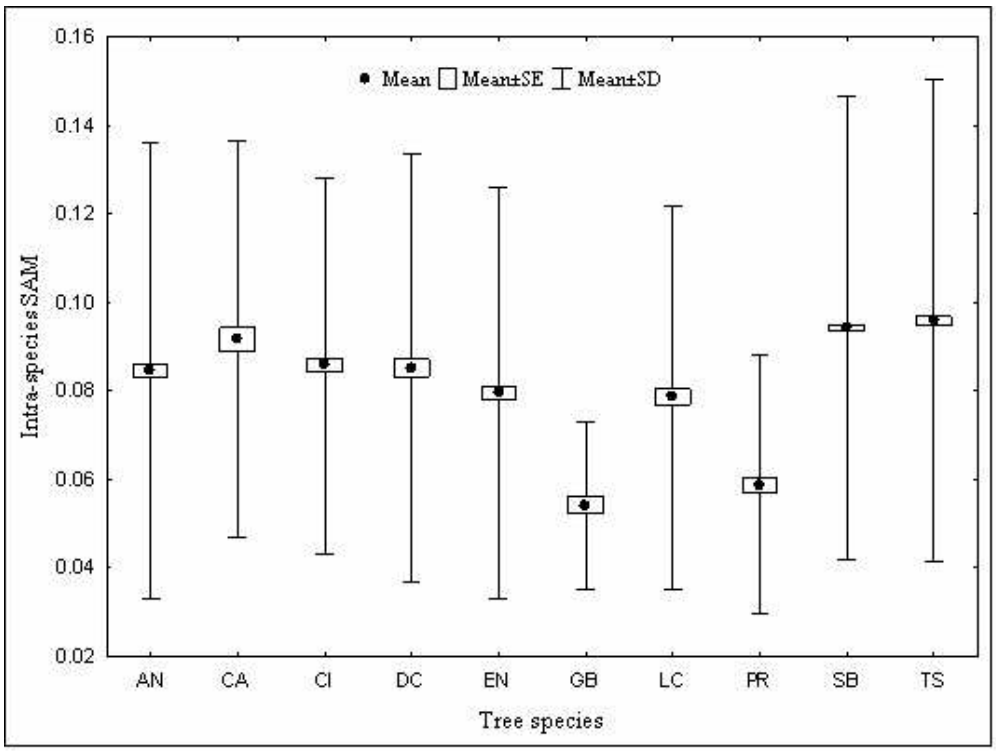
Fig. 5. The correlation between the producer's accuracy (i) or user's accuracy (ii) and intra-species spectral angle measure (SAM), of ten savanna species for 20 bootstrapped samples of the training (1/3) and test (2/3) data, using the mean spectra of the training spectra of each species as reference endmembers (A) and all training spectra of each species as reference endmembers (B).



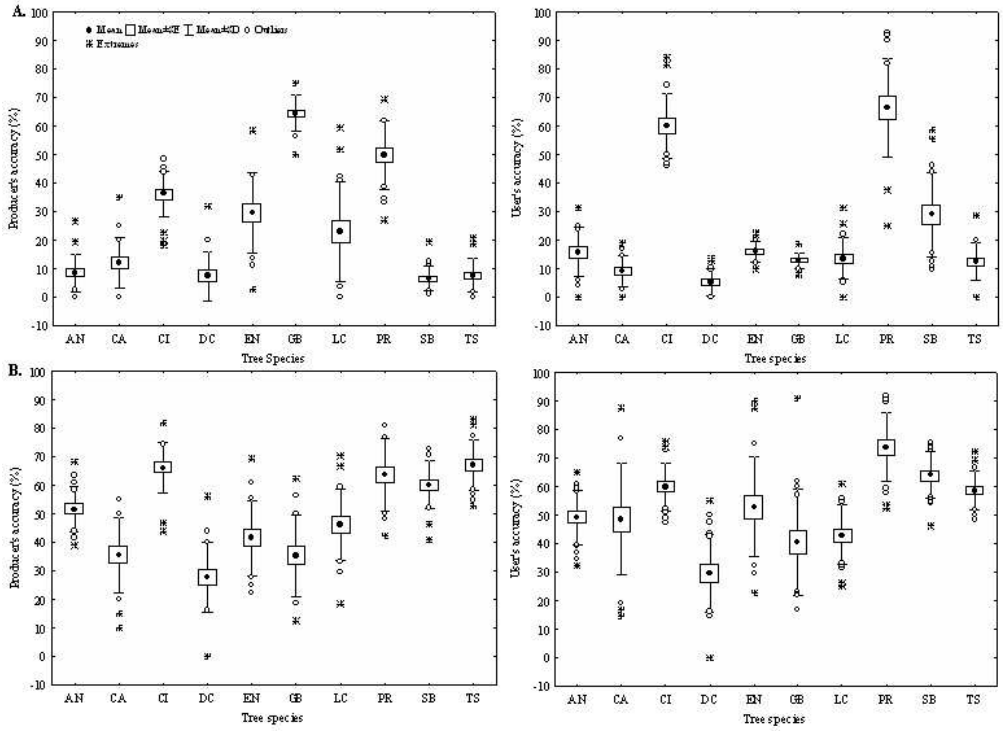
119x193mm (96 x 96 DPI)



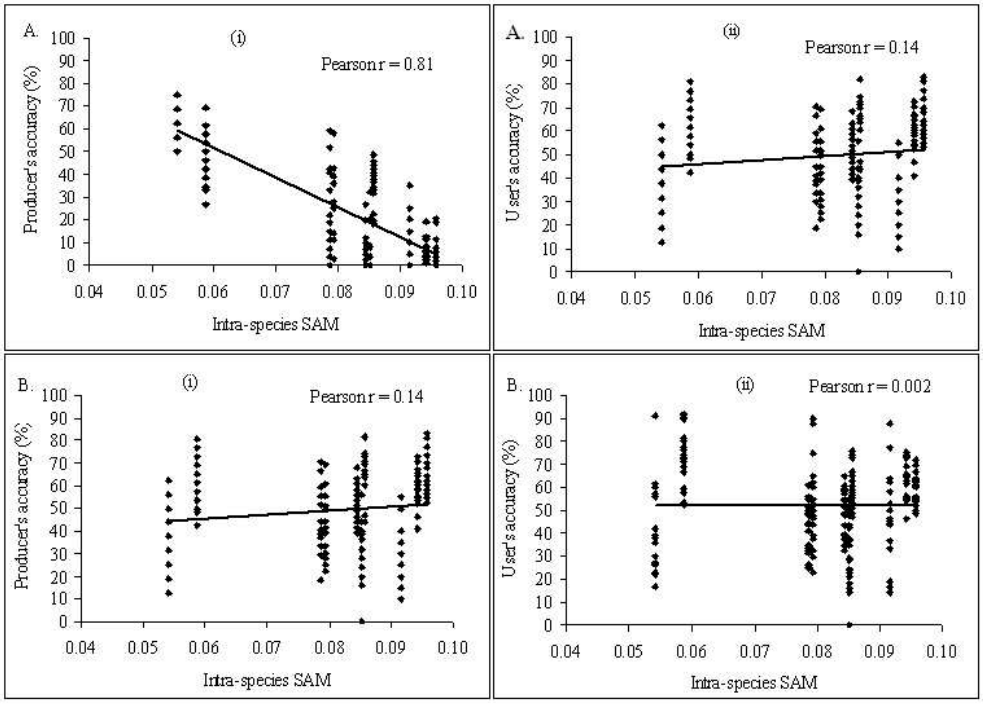
112x244mm (96 x 96 DPI)



142x108mm (96 x 96 DPI)



204x152mm (96 x 96 DPI)



204x145mm (96 x 96 DPI)

Moses A. Cho received the B.Sc. degree in Natural sciences (option: botany) from the University of Yaoundé, Cameroon in 1991. He received the M.Sc. degree in biodiversity conservation in 2001 from the Greenwich University, UK. He obtained his Ph.D. degree from the Wageningen University and The International Institute for Geoinformation Science and Earth Observation, The Netherlands. He is currently a senior researcher at The Council for Scientific and Industrial Research (CSIR), South Africa. He is involved in the development of hyperspectral algorithms for modelling vegetation biochemistry and biophysical properties, and for discriminating vegetation associations. His research interest in the near future is the use of spaceborne hyperspectral imagery for monitoring vegetation condition and biodiversity at a regional scale.

Renaud Mathieu received a Dipl. of Engineer in Agricultural Sciences from the Ecole Supérieure d'Agriculture de Purpan (ESAP), France, in 1990 and a MSc in Applied Remote Sensing from Cranfield University, UK in 1991. He obtained a Ph.D. degree in Geographic Information Sciences from the University of Marne-la-Vallée, France. His is currently a principle researcher at the Council for Scientific and Industrial Research (CSIR), South Africa, and leads the Ecosystems Earth Observation research group. His research interests focus on the application of remote sensing technologies to soil and water conservation, natural resource and wildlife management, environment and agriculture.

Pravesh Debba (1969) was born in Durban, KwaZulu-Natal, South Africa. He received the B.Sc. degree (Mathematics and Statistics) and B.Sc. (Hons) (Statistics) from the University of Durban-Westville, Durban, KwaZulu-Natal, South Africa in 1991 and 1992 respectively. He received the M.Sc. degree in Biostatistics from Limburgs Universitair Centrum, Diepenbeek, Limburg, Belgium in 1998. He received the Ph.D. degree from the Wageningen University and The International Institute for Geoinformation Science and Earth Observation (ITC), The Netherlands. He is currently a principal statistical researcher at The Council for Scientific and Industrial Research (CSIR), South Africa. His research interest is on applying statistical methods in the fields of remote sensing and GIS.

Laven Naidoo received the B.Sc. (Hons) in Environmental Science from the University of Cape Town in 2007. He is current working as an intern researcher with the Council for Scientific and Industrial Research (CSIR), Ecosystem-Earth Observation unit and pursuing the M.Sc. degree on mapping bush encroachment species in and around the Kruger National Park, South Africa using structural and spectral metrics with the University of Cape Town.

Jan van Aardt obtained a B.Sc. Forestry degree (biometry and silviculture specialization) from the University of Stellenbosch, Stellenbosch, South Africa in 1996. This was followed by a Hons. Forestry degree (remote sensing and Geographical Information Systems (GIS) specialization), also from the University of Stellenbosch in 1998. He initiated a M.S. Forestry degree at Virginia Polytechnic Institute and State University, Blacksburg, Virginia in August 1998 with the aid of a Fulbright scholarship. This degree focused on remote sensing (hyperspectral research) and was obtained in August 2000. Jan received his Ph.D. Forestry degree in August 2004, also from Virginia Polytechnic Institute and State University. Hyperspectral

and structural (lidar) sensing of natural resources forms the core of his efforts, which vary between vegetation structural and system state assessment. Jan is currently an associate professor in the Center for Imaging Science at the Rochester Institute of Technology, New York, where he serves as director of the Laboratory for Imaging Algorithms and Systems.

Gregory Asner received his Ph.D. in ecology and remote sensing from the University of Colorado, Boulder in 1998. He currently serves on the faculty in the Department of Global Ecology, Carnegie Institution and in the Department of Environmental Earth System Science at Stanford University. His research aims to understand connections between ecosystems, resource use and climate change. His remote sensing work centers on the use of new technologies for studies of ecosystem structure, function and biodiversity in the context of conservation and management. He is the director of the Carnegie Airborne Observatory at the Carnegie Institution.

70x94mm (72 x 72 DPI)

37x52mm (300 x 300 DPI)

270x203mm (72 x 72 DPI)

112x84mm (72 x 72 DPI)

225x169mm (72 x 72 DPI)

423x158mm (96 x 96 DPI)