

Assessing the effect of seasonality on leaf and canopy spectra for the discrimination of an alien tree species, *Acacia mearnsii*, from co-occurring native species using parametric and non-parametric classifiers

At pot plant level

This chapter is based on Cecilia Masemola, Moses Azong Cho, Abel Ramoelo, 2019. Assessing the effect of seasonality on leaf and canopy spectra for the discrimination of an alien tree species, *Acacia mearnsii*, from co-occurring native species using parametric and non-parametric classifiers. (Accepted for publication in IEEE: Transactions on Geoscience and Remote Sensing)

Abstract

The tree *A. mearnsii* is native to south-eastern Australia but has become an aggressive invader in many countries. In South Africa, it is a significant threat to the conservation of biomes. Detecting and mapping its early invasion is critical. The current ground based methods to map *A. mearnsii* are accurate but are neither economical nor practical. Remote sensing provides accurate and repeatable spatial information on tree species. The potential of remote sensing technology to map *A. mearnsii* distributions remains poorly understood, mainly due to a lack of knowledge on the spectral properties of *A. mearnsii* relative to co-occurring native plants. We investigated the spectral uniqueness of *A. mearnsii* compared to co-occurring native plant species within the South African landscape. We explored full-range (400–2,500 nm), leaf and canopy hyperspectral reflectance of the species. The spectral reflectance was collected bi-weekly from 23 December 2016 and 31 May 2017 using the Field Spec[®] three-field spectroradiometer. We conducted a time series analysis, to assess the effect of seasonality on species discrimination. For comparison, two classification models were employed: parametric interval extended canonical variate discriminant (iECV-DA) and non-parametric random forest-discriminant classifiers (DRF). The results of this study suggest that phenology plays a crucial role in discriminating between *A. mearnsii* and native species. The RF classifier discriminated *A. mearnsii* with slightly higher accuracies (from 92% to 100%) when compared with the iECV-DA (from 85% to 93%). The study showed the potential of RS to discriminate between *A. mearnsii* and co-occurring plant species.

Keywords: *A. mearnsii* extended canonical variates analysis, Random Forest, invasive tree species classification, linear discriminant analysis, and leaf and canopy reflectance.

2.1 Introduction

Invasive alien plant (IAP) is of concern in ecological studies (Luque *et al.* 2014). Invasion poses significant threats to the ecological integrity of terrestrial and aquatic ecosystems (Pyšek *et al.* 2012; Rejmánek and Richardson 2013). The International Union for Conservation of Nature: Invasive Species Specialist Group (IUCN-ISSG) (ISSG 2013) classifies Australian *Acacia* species among the world's 100 worst IAP (Luque *et al.* 2014). A native of south-eastern Australia, *Acacia mearnsii* (black wattle) has become an aggressive invader in many countries (Liu *et al.* 2016). For example, it is a significant invader in the montane rainforest in Rwanda (Seburanga 2015a). Moreover, (Boudiaf *et al.* 2013a) reported it to be the main invader in indigenous cork oak forests in Algeria. In South Africa, it is the most significant threat to the conservation of biomes as it has aggressively invaded grasslands (Yapi 2013; Yapi *et al.* 2018), indigenous forests (Richardson and Rejmánek 2011), and watercourses (Impson *et al.* 2008). The main environmental impacts of *A. mearnsii* include biodiversity loss; specifically by modifying the structure and composition of terrestrial and riparian inhabitants (Pyšek *et al.* 2012; Rejmánek and Richardson 2013) reduction of catchment and river water flows (Moyo and Fatunbi 2010; Le Maitre *et al.* 2016a) the functioning of ecosystems by changing the nitrogen cycle (Lee *et al.* 2017) and the intensification of wildfires (Impson *et al.* 2008). Detecting and mapping its early invasion is critical for an effective management strategy.

Early invasion and rapid response require accurate, consistent and timely information on species distribution. Information on IAP species distribution has been successfully used to model risks and impacts of invasions at landscape (Vila and Ibez 2011; Balch *et al.* 2013) and regional scales (Bradley 2014b). Understanding spatial patterns of *A. mearnsii*, information of its occurrence is required. Currently, *A. mearnsii* mapping is based on ground survey methods and spatial interpolation techniques that approximate its presence at the un-sampled points (Richardson *et al.* 2011). The ground-based approach is accurate but is neither cost-effective nor practical. This approach also precludes accurate mapping of the species over a vast region (West *et al.* 2017). In recent years, ecologists have embraced RS technology to map invasive species occurrences (Boudiaf *et al.* 2013b; Luque *et al.* 2014; Weisberg *et al.* 2017; Lee *et al.* 2017; de Sa *et al.* 2017).

RS provides a non-invasive and non-destructive means of obtaining continuous spatial coverage of the target species' distribution (West *et al.* 2017). Over the years, RS data have been used to discriminate between invasive and native plant species in various ecosystems, for example, river Estuary (Müllerová *et al.* 2017), coastal dune ecosystem (Somers and Asner 2012) and tropical

rainforest (Andrew and Ustin 2006). However, not all invasive species can be mapped using RS [30]. Remote sensing mapping of invasive species depends on their spectral separability from other native species in a heterogeneous species environment.

Previous studies showed that the spectral separability of species depend on the assumption that each species has a unique spectral signature, controlled by their distinctive structural and biochemical features (Hunt *et al.* 2005; Asner 2008; Clark and Roberts 2012a). For example, (Asner 2008) reported separability between native and exotic trees based on their spectra reflectance differences in Hawaiian rainforest. Their study linked spectral separability of the species to variations in leaf pigment, nutrient and structural constituents. A study by (Jiménez and Díaz-Delgado 2015a) showed that absorption features for distinguishing native Mediterranean dune species correlate with the biochemical properties of the species (e.g. pigments, water, lignin and cellulose). According to (Hunt *et al.* 2005), highly pubescent (hairy) plants reflected incident light energy very differently from the less hairy surrounding vegetation. Other research reported a spectral distinction between nitrogen-fixing and non-fixing trees (Asner 2008; Somers and Asner 2012).

Acacia mearnsii, like other invasive species, possesses functional traits that may be different from those of native species. For example, *A. mearnsii* has been reported to have powerful competitive advantages due to their high efficiency in the acquisition and use of nutrients specifically nitrogen (Le Maitre *et al.* 2011). The species has a rapid growth rate, sizeable aboveground biomass and associated leaf area when compared to indigenous vegetation (Morris *et al.* 2011). Moreover, the species has pronounced hairy leaves, which have been found to reflect incident light energy much differently from the less hairy surrounding vegetation (Hunt, E. R. *et al.* 2005). Although spectral discernment of Australian Acacias from native species has been accomplished (Taylor and Kumar 2013; Lehmann *et al.* 2015; Jiménez and Díaz-Delgado 2015a; Große-Stoltenberg *et al.* 2016a), the spectral separability between *A. mearnsii* and co-occurring species has not been meticulously investigated. There is thus no comprehensive regional spatial data of the occurrence of the species in South Africa, even though *A. mearnsii* has been rated to be the most aggressive alien species (Henderson 2007; Fu and Jones 2013). Therefore, the question is whether *A. mearnsii* could be discriminated from co-occurring native species based on leaf and canopy spectral information like other Australian Acacias.

Hyperspectral data is known to provide detailed spectral information related to species-level chemical-structural properties (Asner 2008). The adjacent bands allow the detection of subtle spectral differences between species that are otherwise masked by broadband sensors (Clark and Roberts 2012a; Somers and Asner 2013b; Ferreira *et al.* 2016). Hence, hyperspectral data have been

used as an operational tool for early detection and modelling of future invasion risks (Féret and Asner 2013; Baldeck, C. A. *et al.* 2014). In Hawaii, RS tools have been used to develop invasive species monitoring strategies (Asner *et al.* 2008). Spatially explicit *A. mearnsii* mapping models are hindered by the dearth of information concerning its spectral separability, what constitutes its spectral separability relative to adjacent species, and the appropriate spectral resolution required to provide accurate distribution maps (Morais and Freitas 2015a; Marchante *et al.* 2015b; Müllerová *et al.* 2017). Several studies have shown that comprehensive spectral libraries could give insights into the separability of target species (Huang and Asner 2009; Große-Stoltenberg *et al.* 2016). Subsequently, spectral libraries could be used for the prediction of future invasions and identifying priority areas for conservation (Huang and Asner 2009; Große-Stoltenberg *et al.* 2016a). This highlights the need for understanding the spectral characteristics of *A. mearnsii* when compared to that of co-occurring native species.

Although hyperspectral data have the prospects for quantifying spectral properties of plant species, its application precludes large area mapping due to a small swath width of the collected data (Huang and Asner 2009). Moreover, lower revisit frequencies limit the chance to capture the target species' spatial and phenological changes (Huang and Asner 2009). Due to the rapid growth (Doran and Turnbull 1997; Maslin and McDonald 2006; Seburanga 2015a; Yapi *et al.* 2018) and spread (Mukwada *et al.* 2016) of *A. mearnsii*, hyperspectral sensors can hinder early detection and timely monitoring prospects. However, *A. mearnsii* tends to form large stands and patches (Strydom *et al.* 2017), which makes it possible to detect the species using high spatial and satellite sensors with frequent revisiting time such as Multispectral Instrument (MSI) on-board Sentinel-2 (Ng *et al.* 2017; Mallinis *et al.* 2018) and Operational Land Imager (OLI) on-board Landsat-8 satellite (Wang 2016; Ng *et al.* 2017). The availability of space-borne sensors such as Sentinel-2 and Landsat-8 Operational Land Imager (OLI) could provide an opportunity to distinguish *A. mearnsii* at the landscape level. However, only a few studies have investigated the potential to map *A. mearnsii* using multispectral satellite data (Mararakanye *et al.* 2017). For this reason, we extended the research by including the multispectral dimension to the analysis.

The present study aimed to explore:

- (i) whether *A. mearnsii* is spectrally separable from the co-existing native species;
- (ii) the best spectral wavelength regions for discriminating *A. mearnsii* from its co-occurring native species in South Africa using leaf and canopy spectral reflectance;









- (iii) the optimal phenological period to separate *A. mearnsii* from its co-occurring species;
- (iv) the potential of Landsat-8 OLI and Sentinel-2 spectral band configurations for discriminating *A. mearnsii* from co-occurring plant species

2.2 Methods and Materials

2.2.1 Experimental setup and sampling

The research method chosen to meet the objectives of the study was an outdoor experiment with potted plants. The co-occurring species include *Dombeya tileacea*, *Olea africana*, *Dombeya rotundifolia*, *Euclea crispa*, *Vachellia karroo* and *Vachellia xanthophloea*. *A. mearnsii* and the co-occurring species show differences in their foliage structures as depicted in Table 2.1. Approximately 1-metre tall species were purchased from Nkosi Indigenous Plant Species Nursery based in KwaZulu-Natal, South Africa. The potted plants were left to grow outdoors on the campus of the Council for Scientific and Industrial Research (CSIR) for three months (1 September 2016 to 22 December 2016) before the start of the spectral data collection. The trees were randomly placed to allow the same distribution of energy and other resources. Due to the limited rainfall during this period, we watered the plants once a week from 1 September to 21 December 2016. The experiment took into consideration the soil type of study area to be surveyed during data collection.

Table 0.1. Plant species used to explore spectral separability of *A. mearnsii* from native species

| Species name | Picture | Main characteristics |
|-------------------------------|---|---|
| <i>A. mearnsii</i> |  | Evergreen tree, 6–20 m high. Fast-growing leguminous (nitrogen-fixing) tree. The leaves dark are olive-green branchlets with all parts finely hairy. Leaflets short (1.5–4 mm). Flowering Aug–Sept. |
| <i>Vachellia karroo</i> |  | A deciduous tree (7–12 m). <i>Vachellia karroo</i> leaves pinnate pairs leaflets. Flowering Nov–April. |
| <i>Vachellia xanthophloea</i> |  | A deciduous tree (7–12 m). Leaves pinnate pairs leaflets. Flowering November–April. |
| <i>Euclea crispa</i> |  | Variable short shrub to the medium tree (8–20 m), Flowering December–May. |
| <i>Dombeya tiliacea</i> |  | Scrambling shrub tree (10 m). Spiralled, ovate leaves. Flowering March–August. |
| <i>Dombeya rotundifolia</i> |  | A small deciduous tree (5–10 m). Spiralled irregular lobed, dark green leaves. Flowering July–September. |
| <i>Olea Africana</i> |  | Tree, up to 14 m tall in the forest. Opposite, decussate, shiny, leathery leaves. Flowering October–January. |
| <i>Celtis Africana</i> |  | A deciduous tree (30 m), smooth and slightly leathery leaves. Flowering August–October. |

2.2.2 Canopy level spectra measurements

We conducted a total of nine sets of canopy radiance measurements on a biweekly basis from 23 December 2016 to 31 May 2017. They match the start and the end of the growing seasons in South Africa. The canopy measurements were carried out between 10:00 and 14:00 on cloud-free days. During the sampling campaign, we collected both leaf and canopy reflectance of the selected species between 350 and 2500 nm using an ASD FieldSpec FR 3 Spectroradiometer at 1 nm bandwidth (Analytical Spectral Devices Inc., Boulder, USA). We positioned the fibre optic (FOV 25°) at nadir and a height of approximately 30 cm above the undisturbed individual tree canopy.

Consequently, the field of view at the canopy level was circular, with a radius of 13.3 cm, and the field of view area was covered entirely by leaves to ensure standardised measurements. Furthermore, we eliminated the interference of the background (grass and bare ground) reflectance by placing the pot on a black sheet (board) [49]. We used a ladder for the canopy measurements to ensure that the entire canopy was covered to account for the canopy spectral variability. In order to get the canopy spectral data of each plant, we randomly took six radiance readings and calculated the average to get one canopy reflectance. Overall we collected 80 x 6 measurements per survey. To reduce the effects of changing atmospheric and solar conditions, the reflectance of a Spectralon white reference panel was recorded every 10-15 measurements. The reflectance of the individual tree was converted using a reference measurement for each sample by dividing the reflected target radiance by the irradiance of the white Spectralon® panel.

2.2.3 Leaf level spectra measurement

After every canopy measurement, ten leaves per tree were randomly harvested and taken to a dark laboratory room with the walls and the ceiling coated with the black material. The dark room was used to ensure stable atmospheric and uniform illumination conditions [50-52]. We placed the leaves on a non-reflective black background surface to avoid the impact of external illumination. However, the leaflets of the pinnate leaves of *A. mearnsii*, *V. Karroo* and *V. Xanthophloea* (refer to Fig. 1) were smaller than the sensor FOV, and their reflectance measurements would therefore not be truly representative of the leaf morphology and biochemistry because the leaf area per unit surface would also impact the measured reflectance.

Consequently, we counteracted this effect by stacking the leaflets to simulate a continuous layer of leaves (Clark *et al.* 2005b). The fibre optic with a FOV of 25° was attached to the leaf clip and placed in a nadir position from approximately 4 cm above the leaves. However, in case of bigger leaves, no stacking of the leaves was done during spectra data collection. Capturing leaf spectral properties of the species, five measurements were collected per leaf by repositioning the leaf clip at five different positions for each scan. The reflectance of the individual plant was obtained by averaging the collected spectra reflectance per plant. The resultant spectral database included that of *A. mearnsii* and seven grown native tree species.

2.2.4 Spectral reflectance pre-processing

The pre-processing of spectral reflectance was conducted using the Field Spectroscopy Facility (FSF) Post-Processing Toolbox (MATLAB toolbox) (Robinson and Mac Arthur 2012).

- (i) the toolbox allowed for the exclusion of outliers caused by measurement errors and atmospheric interference;
- (ii) the 350–399 nm bands were not included in the analysis, thus limiting the spectral range to the traditional visible (VIS) to shortwave infrared (SWIR) (400 nm to 2500 nm);
- (iii) in the case of canopy spectra reflectance, we removed SWIR ranges with high noise that were identified through literature and visual inspection, that is, 1350–1460 nm and 1790–1960 nm (Große-Stoltenberg *et al.* 2016b).

After applying the pre-processing, 1759 canopy bands were left for analysis. Lastly, we eliminated sensor noise by using a moving Savitzky-Golay filter (Savitzky and Golay 1964a) with nine-point window size and second polynomial order. Furthermore, we explored various spectral transformation algorithms to evaluate the impact of spectral transformation on species discrimination. We considered the following methods: multiplicative scatter correction (MSC) (Naes *et al.* 1990), standard normal variation (SNV) (Barnes, R. J. *et al.* 1993) and first derivatives. The performance of the transformed data was then compared with that of the untransformed spectral dataset.

2.2.5 Simulation of Landsat-8 OLI and Sentinel-2 wavelengths

To investigate the potential of mapping *A. mearnsii* at the regional level using spaceborne sensors, we simulated Landsat-8 OLI and Sentinel-2 MSI data based on canopy spectral measurements using the respective pre-defined spectral response functions provided by the package 'hsdar' - CRAN.R-project.org (Lehnert *et al.* 2018) of Rattle statistical software (R Core Team 2018). The resultant centre wavelengths were: Landsat 8 OLI (442.96 nm; 482.04 nm; 561.41 nm; 654.59 nm; 864 nm; 1609 nm and 2201 nm) and Sentinel-2 (490 nm; 560 nm; 665 nm; 705 nm; 740 nm; 785 nm; 842 nm, 1601 nm and 2190 nm).

2.3 Classification of plant species samples from leaf and canopy spectra Data

The classification between *A. mearnsii* and sampled native species was based on multivariate Interval Extended Canonical Variable Analysis (iECVA) (Nørgaard *et al.* 2006a) and tree-based Random Forest discriminant analysis (RF-DA) (Lemmond *et al.* 2008; Jones 2015) methods. Fig. 2 summarises the methodology applied in this work. The analysis was carried out in Matlab using the ECVA toolbox available at <http://www.models.life.ku.dk/source/>. On the other hand, RF-DA applied for the classification between *A. mearnsii* and sampled native species were carried out using Matlab Fathom Toolbox by (Jones 2015). The toolbox is available at <http://www.marine.usf.edu/user/djones/>.

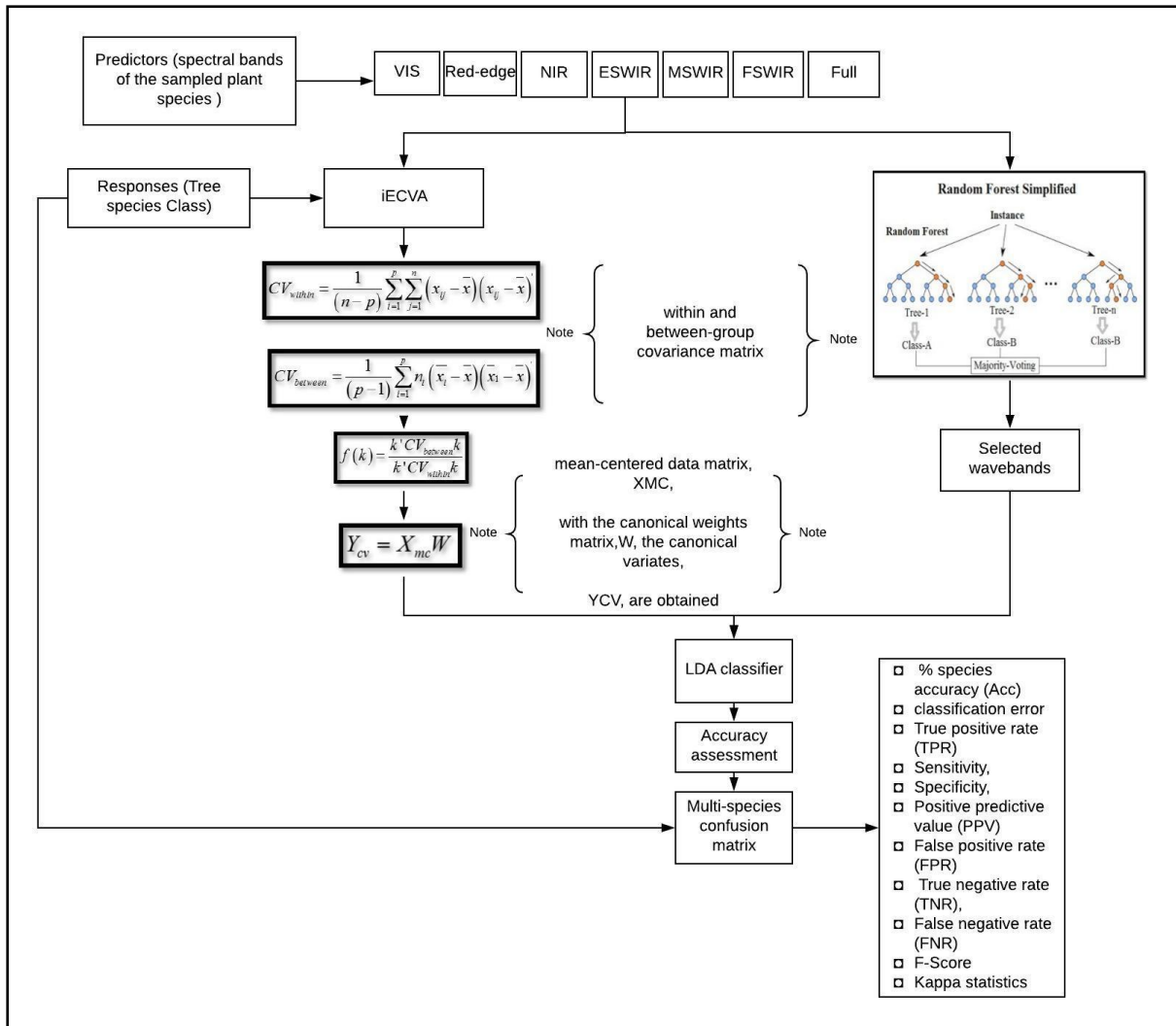


Figure 0.1 The proposed method for distinguishing between *A. mearnsii* and sampled native species using leaf and canopy spectral reflectance.

2.3.1 Interval Extended canonical variate analysis-discriminant analysis

The iECVA method (Nørgaard *et al.* 2006) is a modified classical canonical variates analysis (CVA). The modification was done to mitigate the shortcomings of standard CVA discriminant analysis methods. Unlike standard CVA, iECVA (Nørgaard *et al.* 2006) can handle high dimensional dataset (Nørgaard *et al.* 2006a). The iECVA is a supervised method that finds multivariate directions that separate species classes, and, subsequently, reduces the dimensionality of the predictors. The iECVA (Nørgaard *et al.* 2006) reduce data dimensionality using embedded interval partial least square (iPLS) (Norgaard *et al.* 2000) concepts. The concept of iPLS is used to select critical spectral regions to separate the

species. Briefly, the iECVA technique focuses on the absorption area that contains the critical information that discriminates between species classes. The iECVA uses the mathematical foundation of CVA explained below.

CVA is a supervised classification method that maximises the ratio of the within-group and the between-group covariance using the defined optimisation criterion (4). The method assumes that the matrix, X (n , m) represents the spectral reflectance to be categorised into the samples to target species, where n (samples from different species) and m (number of wavelengths) per species. The standard CVA finds the within (1) and between (2) covariate matrices on the assumption that all species are subject to the same variability.

$$CV_{within} = \frac{1}{(n-p)} \sum_{i=1}^p \sum_{j=1}^n (x_{ij} - \bar{x})(x_{ij} - \bar{x})' \quad \text{Equation 0.1}$$

As mentioned in [49] x_{ij} is the j^{th} sample in the i^{th} group in the column format.

$$CV_{between} = \frac{1}{(p-1)} \sum_{i=1}^p n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \quad \text{Equation 0.2}$$

In the case where CV_{within} is non-singular, the eigenvalue problem is possible and is calculated as shown in (5). However, for a singular CV_{within} , left multiplication by the inverse of CV_{within} is impossible, and this becomes problematic when CVA is used for calibration of multi-collinearity data as it tends to break down.

$$f(k) = \frac{k' CV_{between} k}{k' CV_{within} k} \quad \text{Equation 0.3}$$

$$CV_{within}^{-1} CV_{between} k = \lambda k \quad \text{Equation 0.4}$$

The breakdown problem of CVA in multi-collinearly data has been elucidated by the formulation of iECVA (Nørgaard *et al.* 2006). The iECVA assumes that the two species showed in (1) and (2) can be rewritten according to (Duda *et al.* 1973) in which the directions of the multi-group analysis are expressed as in (5) and then transformed into multivariate regression problem as in Equation (6).

$$(\bar{x}_1 - \bar{x}_2)(\bar{x}_1 - \bar{x}_2)' k = \lambda CV_{within} k \quad \text{Equation 0.5}$$

$$Y_{cv} = X_{mc} W \quad \text{Equation 0.6}$$

Where Y_{cv} represents the columns of the differences between each group mean and the overall mean, X_{mc} is between species covariance, W is the weights, while e is the residual matrix.

Subsequently, an iPLS was implemented between covariance matrices and the species class relationship. iPLS (Nørgaard *et al.* 2000) is a recursive algorithm that builds partial least-squares (PLS) models using user-defined intervals. The resultant accuracy of the interval models then compared with that of the global model (in this study using the full spectral range). The iPLS method identifies features within the spectral regions that intensity variations between species class and assign weights. The weights are sorted in descending order and introduced into (3) for an optimisation process. During optimisation, the weight with the lowest value is left out before the application of the classifier. The iPLS (Nørgaard *et al.* 2000) is utilised to ensure that the space covered by the retained k-1 weights cover the full space of the solution (Nørgaard *et al.* 2006a). The canonical variates are obtained by multiplying the mean-centred data matrix with the canonical weights matrix. The resulting canonical variates were therefore used as an input for an LDA (Nørgaard *et al.* 2006b) to distinguish between *A. mearnsii* sampled native species. The discriminant function for the canonical variates is shown in (8).

$$L_i(t) = \log(\pi_i) - \frac{1}{2}(t - \bar{t}_i)' S_{within, Y_{CV}}^{-1} (t - \bar{t}_i) + \log \|S_{within, Y_{CV}}\| \quad \text{Equation 0.7}$$

According to (Nørgaard *et al.* 2006a) i represent a class index (1; .. ;g), t is a column vector with canonical variates, for the sampled species to be categorised, \bar{t}_i is the mean vector of the canonical variates for species class i , S_{within} is within-group covariance matrix and Y_{CV} represent the canonical variates obtained through multiplication of the mean-centred data matrix (1) with the canonical weights matrix. The detailed description of this method is described in (Nørgaard *et al.* 2006a).

2.3.2. Random forest-discriminant analysis

In this work, iECVA has been compared to the tree-based RF-DA. RF is the most used technique for tree discrimination. It is used for feature selection, thus providing a better understanding of the spectral information variation among species. In contrast to parametric iECVA, the RF is a non-parametric decision tree based technique. The technique uses the majority vote of the ensemble of trees to identify the species class. The values of the number of variables that are randomly sampled as candidates at each split ($mtry$) and the number of trees to grow ($nree$) were identified based on algorithm tuning strategy. We searched for the optimal $mtry$ band the $nree$ values using random search and grid search strategies. Subsequently, the most accurate value for $mtry$ was ten. These parameters were tuned

because of their importance in RF. The rest of the RF parameters were based on default values used by (Jones 2015).

We trained the RF model using bootstrapped leaf and canopy reflectance corresponding to the species. To avoid the influence of both between-class and within-class disparities during classification, we iterated the model 100 times during fitting as recommended by (Li *et al.* 2016). During training, the feature selection and removal of the most correlated spectral wavebands were based on embedded feature selection techniques within RF. According to Saw *et al.* (2015), the selection of the features is based on the variables (wavebands) importance yielded by random forest. This approach identifies essential variables based on randomisation (Saw *et al.* 2015) and estimations of out-of-bag error (Saw *et al.* 2015). The detailed description of the method can be found in (Saw *et al.* 2015).

Similar to iECVA, the selected variables were then used as an input to the DA classification model also found in Fathom Toolbox for Matlab (Jones 2015). In classification problems, an imbalanced dataset may lead to the inadequate identification of the minority class. To avoid the problems related to the majority class on the classifier we used an equal number of samples per species. Moreover, we examined the relative importance of different parts of the spectrum for distinguishing between *A. mearnsii* and native species. Both leaf and canopy spectral reflectance were divided into visible (VIS, 400-650nm), Red-edge (RE, 651-750 nm), near-infrared (NIR, 751-1300nm), early-shortwave-infrared (ESWIR, 1301-1460 nm), mid-shortwave-infrared (MSWIR, 1451-1789 nm) and far shortwave-infrared (FSWIR, 1901-2449 nm) sub-regions.

2.4. Evaluation of Classifiers Performance

Performances of these spectral regions were then assessed and compared to each other and with that of the model built using the full spectra data also known as a global model using a multi-class confusion matrix (MCCM). The MCCM evaluates the model based on per species accuracy (Acc) in percentage as shown in (8), classification error, true positive rate, sensitivity, specificity, positive predictive value (PPV) (9), false positive rate, true negative rate, false negative rate (FNR), F-Score and Kappa statistics (10). As explained in (Zhu *et al.* 2010), sensitivity represents the proportion of actual positives that are correctly identified by a classification model, whereas specificity is the percentage of the true negatives correctly identified by an iECVA-LDA model. The numerical values of sensitivity represent the

probability that iECVA-LDA (Nørgaard *et al.* 2006a) model could identify individual species. The higher the values of sensitivity, the higher the chance that the model will discriminate that particular species from others. On the other hand, the specificity represents the probability of the iECVA-LDA to distinguish a specific species without giving false positive results. Also, the model calculates per species accuracy from the proportion of both true positive and true negative in the selected population. The maximum numerical values of 1.0 or 100% demonstrate that the species is strongly separable from other species (Zhu *et al.* 2010). It is noteworthy that high sensitivity or specificity does not necessarily imply that the accuracy of the classification is high as well (Zhu *et al.* 2010). As a result, we also used the coefficient statistic to measure if there is an actual agreement between predicted and observed species (Zhu *et al.* 2010). According to Cohen, we can interpret kappa statistics as follows: values $\leq 0\%$ are an indication of no agreement, 20–40% is considered fair, while values between 41–60% are moderate, 61–80% is substantial, and values between 81 and 100 indicate perfect agreement (Landis and Koch 1977). The probability that the classifiers discriminated the *A. mearnsii* better than the random chance has been demonstrated with the Z-score and associated P-value.

$$\% Accuracy = \frac{\text{obtained result} - \text{expected result}}{\text{expected result}} * 100 \quad \text{Equation 0.8}$$

Where %Accuracy, is the percentage of correctly classified species class out of all classes.

$$PPV = \frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity}) \times (1 - \text{prevalence})} \quad \text{Equation 0.9}$$

According to (Fielding and Bell 1997; Lurz *et al.* 2001), sensitivity is conceptually identical to the Producer's Accuracy. Moreover, assesses the probability that the species will be classified an *A. mearnsii* if it is *A. mearnsii*. Whereas, Specificity is a measure of the probability that a species class will not be predicted as *A. mearnsii* if it is not *A. mearnsii*.

$$Kappa = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)} \quad \text{Equation 0.10}$$

where, $\text{Pr}(a)$ and $\text{Pr}(e)$ represent the probability of species classification success and the probability of classification success due to chance, respectively.

2.5 Results

2.5.1 The spectral separability between *Acacia mearnsii* and co-occurring native species at the leaf level

The various spectral transformations explored yielded comparable results for distinguishing between *A. mearnsii* and native species. Therefore, only the results observed from SNV transformed spectra are presented. Generally, *A. mearnsii* exhibited increased separability compared to sampled native species. Figure 3 and Table I, indicate that the difference between these species was during the transition from peak productivity to senescence in South Africa (Table 2.1, March and April). The distribution patterns of the tree species in the two-dimensional canonical variate space of their leaf spectra reflectance displayed a tight grouping of *A. mearnsii* and a rather substantial overlap among native species.

The iECVA-DA and RF-DA classification for the species produced the best result in the VIS, RE, NIR and ESWIR spectral regions (Table 2.1). From Table 2.1, iECVA-DA and RF-DA yielded accuracies and kappa coefficient that ranged from 85–100% and 93-to100 percentage, respectively (Table 2.1). In comparison, RF-DA classifier yielded somewhat slightly higher accuracies and kappa coefficient that ranged from 93-to 100%. Surprisingly we observed a significant separation between *A. mearnsii* and native species in May with RF-DA (Table 2.1). Analysis of the full spectra data showed a significant ($p=0.0002$) difference between *A. mearnsii* and the native species. The full spectra distinguished between *A. mearnsii* and native species with slightly higher accuracies than that produced with MSWIR and FSWIR spectral bands (Table 2.1). The strong performance of full spectra data implies that both classifiers can deal with high spectral variability and multi-collinearity that comes with hyperspectral data. Overall, there was great spectral confusion among the native species. Hence, iECV-DA and RF-DA classifiers produced significantly lower overall accuracies and kappa statistics that ranged from 31–52% and 20–51%, respectively (Table 2.1). A statistical comparison of the classifiers suggests that there is no significant difference between the iECVA-DA and RF-DA tree-based models ($Z=0.335$, $p = 0.369$).

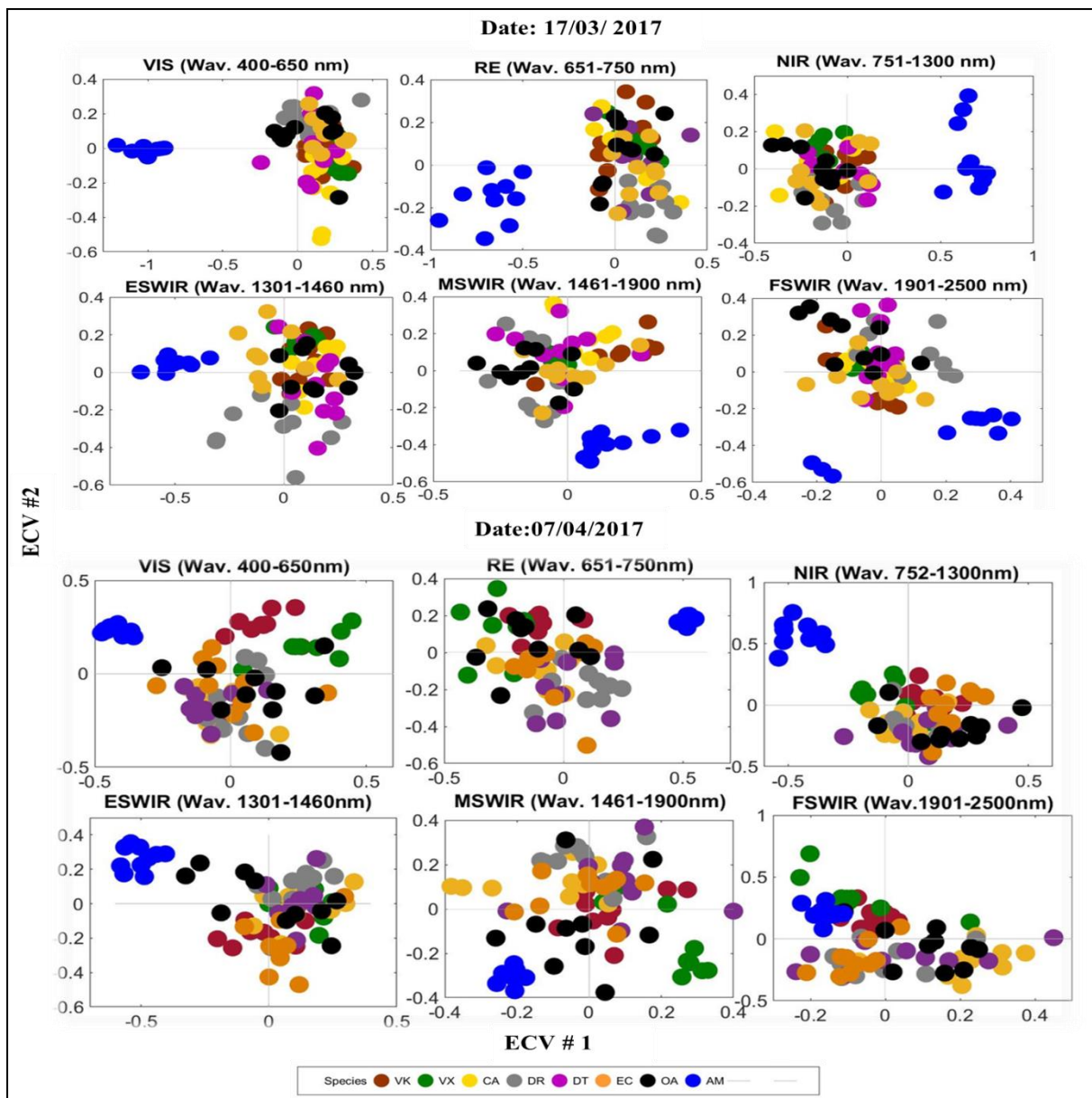


Figure 0.2 Leaf scale distribution pattern of the studied plant species in Extended Canonical Variable (ECV#1) versus (ECV#2) two-dimensional space of their leaf narrowband spectral regions during the optimal separability period of *A. mearnsii*. The markers are DT= *Dombeya tileacea*, OA=*Olea africana*, DR=*Dombeya rotundifolia*, EC=*Euclea crispa*, VK=*Vachellia karroo* and VX=*Vachellia xanthophloea*. ECV#1 ECV#1 and ECV#2 = first and second extended canonical variates, respectively.

Table 0.2 Leaf scale statistical metrics in percentage, for each sampled species class, yielded by each explored spectral range using iECVA-LDA models. The bold results represent the period where *A. mearnsii* is highly discriminated.

| Date | 23/12/2016 | | 13/01/2017 | | 03/02/2017 | | 24/02/2017 | | 17/03/2017 | | 07/04/2017 | | 28/04/2017 | | 12/05/2017 | | 31/05/2017 | |
|---|------------|-------|------------|-------|------------|-------|------------|-------|-------------|-------------|-------------|-------------|------------|-------|------------|-------|------------|-------|
| | RF | iECVA | RF | iECVA | RF | iECVA | RF | iECVA | RF | iECVA | RF | iECVA | RF | iECVA | RF | iECVA | RF | iECVA |
| Accuracy metrics | | | | | | | | | | | | | | | | | | |
| Visible spectral region (VIS = 400-650 nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 55% | 53% | 68% | 67% | 78% | 60% | 89% | 83% | 93% | 91% | 96% | 95% | 82% | 74% | 83% | 67% | 57% | 27% |
| PPV (AM) | 94% | 68% | 90% | 71% | 100% | 68% | 97% | 78% | 100% | 98% | 100% | 100% | 99% | 69% | 100% | 72% | 97% | 60% |
| Kc (AM) | 0.51 | 0.38 | 0.56 | 0.47 | 0.70 | 0.50 | 0.82 | 0.48 | 0.90 | 0.87 | 0.92 | 0.90 | 0.82 | 0.48 | 0.80 | 0.36 | 0.52 | 0.32 |
| OAcc | 36% | 38% | 34% | 42% | 39% | 46% | 59% | 52% | 51% | 49% | 50% | 35% | 60% | 29% | 40% | 26% | 24% | 0% |
| Okc | 0.31 | 0.29 | 0.39 | 0.34 | 0.32 | 0.27 | 0.53 | 0.34 | 0.43 | 0.37 | 0.43 | 0.42 | 0.54 | 0.27 | 0.29 | 0.22 | 0.24 | 0.17 |
| Red edge spectral region (RE = 650-750 nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 63% | 39% | 73% | 44% | 67% | 58% | 67% | 67% | 91% | 90% | 93% | 91% | 91% | 77% | 81% | 67% | 64% | 0% |
| PPV (AM) | 56% | 46% | 97% | 60% | 97% | 66% | 94% | 58% | 97% | 96% | 95% | 99% | 95% | 55% | 91% | 76% | 67% | 25% |
| Kc (AM) | 0.50 | 0.33 | 0.73 | 0.40 | 0.69 | 0.44 | 0.58 | 0.57 | 0.86 | 0.87 | 0.90 | 0.89 | 0.85 | 0.32 | 0.70 | 0.27 | 0.62 | 0.10 |
| OAcc | 39% | 57% | 45% | 47% | 35% | 42% | 44% | 48% | 45% | 31% | 39% | 52% | 66% | 46% | 32% | 43% | 19% | 41% |
| Okc | 0.29 | 0.37 | 0.37 | 0.42 | 0.30 | 0.42 | 0.34 | 0.28 | 0.37 | 0.35 | 0.37 | 0.46 | 0.58 | 0.33 | 0.25 | 0.39 | 0.24 | 0.22 |
| Near Infrared spectral region (NIR = 751-1300 nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 27% | 83% | 33% | 76% | 46% | 75% | 60% | 85% | 97% | 96% | 100% | 98% | 83% | 71% | 89% | 63% | 46% | 78% |
| PPV (AM) | 90% | 52% | 90% | 51% | 94% | 69% | 40% | 79% | 90% | 98% | 96% | 100% | 100% | 88% | 100% | 55% | 94% | 66% |
| Kc (AM) | 0.18 | 0.66 | 0.22 | 0.65 | 0.45 | 0.70 | 0.30 | 0.73 | 0.92 | 0.85 | 0.96 | 0.92 | 0.90 | 0.53 | 0.85 | 0.48 | 0.45 | 0.37 |
| OAcc | 24% | 26% | 25% | 36% | 31% | 38% | 25% | 33% | 25% | 33% | 24% | 42% | 45% | 23% | 28% | 38% | 28% | 31% |
| Okc | 0.24 | 0.18 | 0.20 | 0.27 | 0.25 | 0.22 | 0.25 | 0.13 | 0.20 | 0.27 | 0.20 | 0.36 | 0.37 | 0.27 | 0.21 | 0.23 | 0.16 | 0.17 |
| Early short-wave infrared (ESWIR = 1301-1460 nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 0.33 | 0.51 | 0.54 | 0.67 | 0.64 | 0.60 | 0.44 | 0.79 | 0.93 | 0.91 | 0.95 | 0.98 | 0.62 | 0.83 | 0.77 | 0.71 | 0.60 | 0.13 |
| PPV (AM) | 0.90 | 0.65 | 0.96 | 0.56 | 0.96 | 0.68 | 0.92 | 0.82 | 0.96 | 1.00 | 0.94 | 1.00 | 0.97 | 0.90 | 1.00 | 0.65 | 0.94 | 0.48 |
| Kc (AM) | 0.22 | 0.57 | 0.55 | 0.62 | 0.62 | 0.68 | 0.34 | 0.68 | 0.90 | 0.90 | 0.91 | 0.90 | 0.56 | 0.63 | 0.71 | 0.57 | 0.54 | 0.45 |
| OAcc | 0.29 | 0.38 | 0.31 | 0.27 | 0.28 | 0.43 | 0.29 | 0.25 | 0.31 | 0.33 | 0.28 | 0.42 | 0.43 | 0.26 | 0.26 | 0.48 | 0.23 | 0.44 |
| Okc | 0.23 | 0.25 | 0.24 | 0.19 | 0.25 | 0.30 | 0.18 | 0.24 | 0.24 | 0.27 | 0.24 | 0.28 | 0.35 | 0.12 | 0.22 | 0.33 | 0.25 | 0.29 |
| Mid shortwave Infrared region (MSWIR = 1461- 1900 nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 56% | 30% | 90% | 67% | 62% | 50% | 86% | 75% | 94% | 92% | 94% | 93% | 69% | 49% | 89% | 40% | 64% | 30% |
| PPV (AM) | 93% | 65% | 99% | 69% | 97% | 72% | 95% | 77% | 99% | 99% | 96% | 99% | 99% | 65% | 100% | 44% | 96% | 36% |
| Kc (AM) | 0.46 | 0.74 | 0.89 | 0.89 | 0.65 | 0.80 | 0.69 | 0.84 | 0.89 | 0.84 | 0.90 | 0.88 | 0.66 | 0.53 | 0.82 | 0.31 | 0.62 | 0.29 |
| OAcc | 34% | 37% | 48% | 32% | 26% | 35% | 28% | 43% | 48% | 52% | 26% | 46% | 36% | 44% | 36% | 45% | 33% | 25% |
| Okc | 0.28 | 0.39 | 0.43 | 0.64 | 0.21 | 0.48 | 0.26 | 0.61 | 0.43 | 0.53 | 0.21 | 0.58 | 0.29 | 0.59 | 0.32 | 0.59 | 0.28 | 0.45 |
| Far short-wave infrared (FSWIR = 1901-2500 nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 64% | 50% | 73% | 44% | 69% | 33% | 73% | 53% | 83% | 88% | 94% | 91% | 82% | 43% | 82% | 25% | 44% | 17% |
| PPV (AM) | 96% | 57% | 97% | 51% | 99% | 48% | 97% | 68% | 97% | 88% | 96% | 95% | 99% | 57% | 99% | 35% | 92% | 19% |
| Kc (AM) | 0.62 | 0.39 | 0.73 | 0.40 | 0.76 | 0.28 | 0.73 | 0.51 | 0.77 | 0.87 | 0.88 | 0.89 | 0.78 | 0.34 | 0.84 | 0.23 | 0.34 | 0.12 |
| OA | 44% | 26% | 41% | 32% | 31% | 35% | 40% | 42% | 41% | 51% | 26% | 46% | 43% | 42% | 41% | 40% | 30% | 39% |
| Okc | 0.36 | 0.34 | 0.32 | 0.16 | 0.24 | 0.13 | 0.30 | 0.19 | 0.32 | 0.34 | 0.21 | 0.33 | 0.34 | 0.25 | 0.33 | 0.17 | 0.20 | 0.23 |
| Full spectral (400nm-2500nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 67% | 66% | 67% | 63% | 69% | 77% | 86% | 84% | 95% | 90% | 100% | 99% | 90% | 77% | 80% | 62% | 69% | 26% |
| PPV (AM) | 90% | 99% | 97% | 58% | 99% | 64% | 95% | 78% | 100% | 95% | 100% | 90% | 99% | 68% | 97% | 72% | 93% | 89% |
| Kc (AM) | 0.58 | 0.48 | 0.54 | 0.57 | 0.66 | 0.62 | 0.79 | 0.64 | 0.86 | 0.87 | 0.95 | 0.94 | 0.85 | 0.40 | 0.76 | 0.31 | 0.58 | 0.29 |
| OAcc | 28% | 35% | 24% | 45% | 28% | 36% | 38% | 25% | 47% | 35% | 20% | 45% | 42% | 25% | 36% | 37% | 31% | 32% |
| Okc | 0.34 | 0.34 | 0.36 | 0.16 | 0.27 | 0.13 | 0.23 | 0.19 | 0.33 | 0.34 | 0.24 | 0.33 | 0.33 | 0.25 | 0.24 | 0.17 | 0.26 | 0.23 |

¹ AM = *A. mearnsii*, CA=*Celtis Africana*, DT=*Dombeya tileacea*, OA=*Olea Africana*, DR=*Dombeya rotundifolia*, EC=*Euclea crispa*, VK=*Vachellia karroo* and VX=*Vachellia xanthophloea*. Where Acc=accuracy, Kc=Kappa coefficient, OAcc=Overall accuracy, Okc=Overall kappa coefficient, PPV=Positive predictive value.

2.5.2 Canopy level spectral separability of *Acacia mearnsii* among co-occurring native species

At canopy level, similar spectral dissimilarities as that of leaf spectral data were observed between *A. mearnsii* and native species with an overall best discrimination accuracy between March and April (Table 2.2). The ECV plots (Fig. 2.4) showed a clear separation between *A. mearnsii* and native species. Interestingly, upscaling from leaf to canopy level slightly increased the spectral variation between *A. mearnsii* and native species in the VIS, RE, NIR and ESWIR reflectance bands (Table II), irrespective of the classifier used.

The iECVA-DA and RF-DA accuracies and Kappa statistics ranged from 92–100% and 93–100%, respectively (Table 2.2). Contrary to iECVA-DA, RF-DA produced slightly higher accuracies in RE, NIR and ESWIR reflectance during the senescence-winter transition period, April-May (Acc (AM) = 83-95%, Kcc (AM)=0.75 -0.83).

Overall accuracies were significantly low which could be attributed to the strong spectral reflectance confusion between native species as demonstrated in Fig. 2.4. The use of the whole wavelength range (400–2,500 nm) yielded significant ($p=0.0001$) high accuracy for *A. mearnsii*, regardless of classifier employed. The consistency of the results from iECVA-DA and RF-DA strongly suggests that leaf biochemical properties are essential traits for the separability of *A. mearnsii*. As opposed to iECV-DA, the spectral separability between *A. mearnsii* and native species was also clearly expressed in May (Table 2.1 and Table 2.2) with RF-DA at both leaf and canopy level.

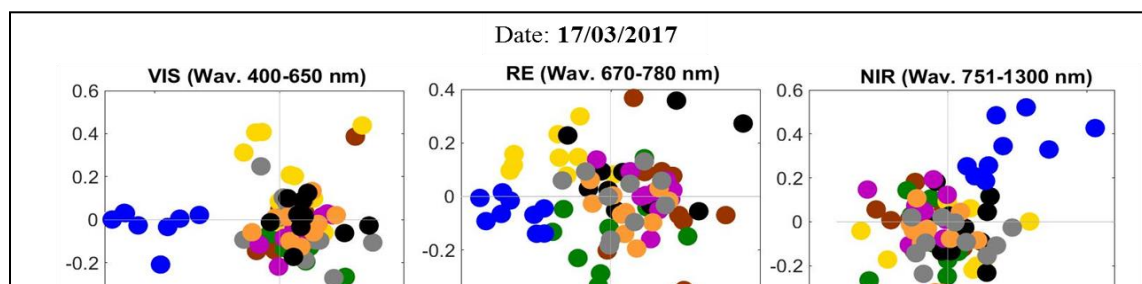


Table 0.3 Canopy scale statistical metrics for each sampled species class yielded by each explored canopy spectral range using iECVA-LDA models. The bold results represent the period where *A. mearnsii* is highly discriminated.

| Date | 23/12/2016 | | 13/01/2017 | | 03/02/2017 | | 24/02/2017 | | 17/03/2017 | | 07/04/2017 | | 28/04/2017 | | 12/05/2017 | | 31/05/2017 | |
|---|------------|-------|------------|-------|------------|-------|------------|-------|-------------|-------------|-------------|-------------|------------|-------|------------|-------|------------|-------|
| | RF | iECVA | RF | iECVA | RF | iECVA | RF | iECVA | RF | iECVA | RF | iECVA | RF | iECVA | RF | iECVA | RF | iECVA |
| Accuracy metrics | | | | | | | | | | | | | | | | | | |
| Visible spectral region (VIS = 400-650 nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 55% | 53% | 68% | 67% | 78% | 60% | 89% | 83% | 93% | 91% | 96% | 95% | 82% | 74% | 83% | 67% | 57% | 27% |
| PPV (AM) | 94% | 68% | 90% | 71% | 100% | 68% | 97% | 78% | 100% | 98% | 100% | 100% | 99% | 69% | 100% | 72% | 97% | 60% |
| Kc (AM) | 0.51 | 0.38 | 0.56 | 0.47 | 0.70 | 0.50 | 0.82 | 0.48 | 0.90 | 0.87 | 0.92 | 0.90 | 0.82 | 0.48 | 0.80 | 0.36 | 0.52 | 0.32 |
| OAcc | 36% | 38% | 34% | 42% | 39% | 46% | 59% | 52% | 51% | 49% | 50% | 35% | 60% | 29% | 40% | 26% | 24% | 0% |
| Okc | 0.31 | 0.29 | 0.39 | 0.34 | 0.32 | 0.27 | 0.53 | 0.34 | 0.43 | 0.37 | 0.43 | 0.42 | 0.54 | 0.27 | 0.29 | 0.22 | 0.24 | 0.17 |
| Red edge spectral region (RE = 650-750 nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 63% | 39% | 73% | 44% | 67% | 58% | 67% | 67% | 91% | 90% | 93% | 91% | 91% | 77% | 81% | 67% | 64% | 0% |
| PPV (AM) | 56% | 46% | 97% | 60% | 97% | 66% | 94% | 58% | 97% | 96% | 95% | 99% | 95% | 55% | 91% | 76% | 67% | 25% |
| Kc (AM) | 0.50 | 0.33 | 0.73 | 0.40 | 0.69 | 0.44 | 0.58 | 0.57 | 0.86 | 0.87 | 0.90 | 0.89 | 0.85 | 0.32 | 0.70 | 0.27 | 0.62 | 0.10 |
| OAcc | 39% | 57% | 45% | 47% | 35% | 42% | 44% | 48% | 45% | 31% | 39% | 52% | 66% | 46% | 32% | 43% | 19% | 41% |
| Okc | 0.29 | 0.37 | 0.37 | 0.42 | 0.30 | 0.42 | 0.34 | 0.28 | 0.37 | 0.35 | 0.37 | 0.46 | 0.58 | 0.33 | 0.25 | 0.39 | 0.24 | 0.22 |
| Near Infrared spectral region (NIR = 751-1300 nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 27% | 83% | 33% | 76% | 46% | 75% | 60% | 85% | 97% | 96% | 100% | 98% | 83% | 71% | 89% | 63% | 46% | 78% |
| PPV (AM) | 90% | 52% | 90% | 51% | 94% | 69% | 40% | 79% | 90% | 98% | 96% | 100% | 100% | 88% | 100% | 55% | 94% | 66% |
| Kc (AM) | 0.18 | 0.66 | 0.22 | 0.65 | 0.45 | 0.70 | 0.30 | 0.73 | 0.92 | 0.85 | 0.96 | 0.92 | 0.90 | 0.53 | 0.85 | 0.48 | 0.45 | 0.37 |
| OAcc | 24% | 26% | 25% | 36% | 31% | 38% | 25% | 33% | 25% | 33% | 24% | 42% | 45% | 23% | 28% | 38% | 28% | 31% |
| Okc | 0.24 | 0.18 | 0.20 | 0.27 | 0.25 | 0.22 | 0.25 | 0.13 | 0.20 | 0.27 | 0.20 | 0.36 | 0.37 | 0.27 | 0.21 | 0.23 | 0.16 | 0.17 |
| Early short-wave infrared (ESWIR = 1301-1460 nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 0.33 | 0.51 | 0.54 | 0.67 | 0.64 | 0.60 | 0.44 | 0.79 | 0.93 | 0.91 | 0.95 | 0.98 | 0.62 | 0.83 | 0.77 | 0.71 | 0.60 | 0.13 |
| PPV (AM) | 0.90 | 0.65 | 0.96 | 0.56 | 0.96 | 0.68 | 0.92 | 0.82 | 0.96 | 1.00 | 0.94 | 1.00 | 0.97 | 0.90 | 1.00 | 0.65 | 0.94 | 0.48 |
| Kc (AM) | 0.22 | 0.57 | 0.55 | 0.62 | 0.62 | 0.68 | 0.34 | 0.68 | 0.90 | 0.90 | 0.91 | 0.90 | 0.56 | 0.63 | 0.71 | 0.57 | 0.54 | 0.45 |
| OAcc | 0.29 | 0.38 | 0.31 | 0.27 | 0.28 | 0.43 | 0.29 | 0.25 | 0.31 | 0.33 | 0.28 | 0.42 | 0.43 | 0.26 | 0.26 | 0.48 | 0.23 | 0.44 |
| Okc | 0.23 | 0.25 | 0.24 | 0.19 | 0.25 | 0.30 | 0.18 | 0.24 | 0.24 | 0.27 | 0.24 | 0.28 | 0.35 | 0.12 | 0.22 | 0.33 | 0.25 | 0.29 |
| Mid shortwave Infrared region (MSWIR = 1461- 1900 nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 56% | 30% | 90% | 67% | 62% | 50% | 86% | 75% | 94% | 92% | 94% | 93% | 69% | 49% | 89% | 40% | 64% | 30% |
| PPV (AM) | 93% | 65% | 99% | 69% | 97% | 72% | 95% | 77% | 99% | 99% | 96% | 99% | 99% | 65% | 100% | 44% | 96% | 36% |
| Kc (AM) | 0.46 | 0.74 | 0.89 | 0.89 | 0.65 | 0.80 | 0.69 | 0.84 | 0.89 | 0.84 | 0.90 | 0.88 | 0.66 | 0.53 | 0.82 | 0.31 | 0.62 | 0.29 |
| OAcc | 34% | 37% | 48% | 32% | 26% | 35% | 28% | 43% | 48% | 52% | 26% | 46% | 36% | 44% | 36% | 45% | 33% | 25% |
| Okc | 0.28 | 0.39 | 0.43 | 0.64 | 0.21 | 0.48 | 0.26 | 0.61 | 0.43 | 0.53 | 0.21 | 0.58 | 0.29 | 0.59 | 0.32 | 0.59 | 0.28 | 0.45 |
| Far short-wave infrared (FSWIR = 1901-2500 nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 64% | 50% | 73% | 44% | 69% | 33% | 73% | 53% | 83% | 88% | 94% | 91% | 82% | 43% | 82% | 25% | 44% | 17% |
| PPV (AM) | 96% | 57% | 97% | 51% | 99% | 48% | 97% | 68% | 97% | 88% | 96% | 95% | 99% | 57% | 99% | 35% | 92% | 19% |
| Kc (AM) | 0.62 | 0.39 | 0.73 | 0.40 | 0.76 | 0.28 | 0.73 | 0.51 | 0.77 | 0.87 | 0.88 | 0.89 | 0.78 | 0.34 | 0.84 | 0.23 | 0.34 | 0.12 |
| OA | 44% | 26% | 41% | 32% | 31% | 35% | 40% | 42% | 41% | 51% | 26% | 46% | 43% | 42% | 41% | 40% | 30% | 39% |
| OKc | 0.36 | 0.34 | 0.32 | 0.16 | 0.24 | 0.13 | 0.30 | 0.19 | 0.32 | 0.34 | 0.21 | 0.33 | 0.34 | 0.25 | 0.33 | 0.17 | 0.20 | 0.23 |
| Full spectral (400nm-2500nm) | | | | | | | | | | | | | | | | | | |
| Acc (AM) | 67% | 66% | 67% | 63% | 69% | 77% | 86% | 84% | 95% | 90% | 100% | 99% | 90% | 77% | 80% | 62% | 69% | 26% |
| PPV (AM) | 90% | 99% | 97% | 58% | 99% | 64% | 95% | 78% | 100% | 95% | 100% | 90% | 99% | 68% | 97% | 72% | 93% | 89% |
| Kc (AM) | 0.58 | 0.48 | 0.54 | 0.57 | 0.66 | 0.62 | 0.79 | 0.64 | 0.86 | 0.87 | 0.95 | 0.94 | 0.85 | 0.40 | 0.76 | 0.31 | 0.58 | 0.29 |
| OAcc | 28% | 35% | 24% | 45% | 28% | 36% | 38% | 25% | 47% | 35% | 20% | 45% | 42% | 25% | 36% | 37% | 31% | 32% |
| Okc | 0.34 | 0.34 | 0.36 | 0.16 | 0.27 | 0.13 | 0.23 | 0.19 | 0.33 | 0.34 | 0.24 | 0.33 | 0.33 | 0.25 | 0.24 | 0.17 | 0.26 | 0.23 |

AM = *A. mearnsii*, CA=*Celtis Africana*, DT=*Dombeya tileacea*, OA=*Olea Africana*, DR=*Dombeya rotundifolia*, EC=*Euclea crispa*, VK=*Vachellia karroo* and VX=*Vachellia xanthophloea*. Where Acc=accuracy, Kc=Kappa coefficient, OAcc=Overall accuracy, Okc=Overall kappa coefficient, PPV=Positive predictive value.

2.5.3 Up-scaling *in situ* canopies spectral to Sentinel-2 and Landsat-8 OLI bands

Table III shows results from using simulated Landsat 8 OLI and Sentinel-2 MSI data. The results of Sentinel-2 MSI show comparable results with that of Landsat 8 OLI regarding discriminating between *A. mearnsii* and sampled native species. Both spectral datasets indicated high discrimination of the *A. mearnsii*. Sentinel-2 MSI yielded a percentage Acc (AM) of 95.52 and Kcc (AM) of 0.90, whereas Landsat-8 OLI spectral data achieved % Acc = 93.38 and Kcc of 0.88. Although both classifiers selected the same variables, RF discriminated *A. mearnsii* with slightly lower Acc and Kcc (AM) values (Table 2.3). Both classifiers selected the same best predictor variables (Table 2.3), where SWIR and NIR showed to be important for Sentinel-2 MSI and Landsat 8 OLI (Table 2.3). Furthermore, red-edge (783 nm) and red-edge (707 nm) bands were selected as best the predictor variables for Sentinel-2 MSI. Furthermore, Red (654 nm) was selected for Landsat 8 OLI.

Table 0.4 Statistical metrics in percentage for each sampled species class yielded by Sentinel 2 MSI and Landsat-8 OLI spectral bands using iECVA-LDA models. All selected bands are significant at p -value < 0.05. Acc= accuracy, Kc=Kappa coefficient, OAcc=Overall accuracy, Okc= Overall kappa coefficient, PPV= Positive predictive value and AM=*A. mearnsii*, RF=Random forest-discriminant analysis and iECV-DA= interval extended canonical variates -discriminant analysis.

| Models | RF | | iECVA | |
|-----------------------|--|--|--|---|
| | S2-MSI | L8-OLI | S2-MSI | L8-OLI |
| Accuracy metrics | | | | |
| Acc(AM) | 92% | 90% | 95% | 93% |
| PPV(AM) | 100% | 94% | 100% | 97% |
| Kc(AM) | 0.89 | 0.89 | 0.90 | 0.88 |
| OAcc | 44% | 45% | 62% | 57% |
| Okc | 0.31 | 0.40 | 0.56 | 0.46 |
| Important wavelengths | Red (665 nm), RE (783 nm), RE (707nm), NIR (835.6 nm) and SWIR (1610 nm) | Red (654.59 nm), NIR (864nm) and SWIR-2200nm | Red (665 nm), RE (783 nm), RE (707nm), NIR (835.6 nm) and SWIR (1610 nm) | Red (654.59nm), NIR (864nm) and SWIR-(2200nm) |

RE=Red-edge, NIR=near-infrared, SWIR=Shortwave-infrared, S2-MSI= Sentinel-2 Multi-spectral Imager and L8-OLI=Landsat-8 Operational Land Imager

2.6 Discussion

This study investigated the separability of *A. mearnsii* from co-occurring native species using both leaf and canopy hyperspectral reflectance. This study demonstrated for the first time the utility of the iECVA-LDA classifier for classification of vegetation species. We also compared the ability of the iECVA-LDA classifier to discern species with that of RF-DA classifier. The results from our study suggest that *A. mearnsii* is distinguishable from native species. However, seasonal shifts played an important part in separating between species. Our results corroborate those from (Asner 2008) and (Somers and Asner 2012) studies that have found the high spectral separation between invasive and native species during certain phenological periods.

In this study, the most significant spectral differences between *A. mearnsii* and other species was observed for the March and April reflectance data at both leaf and canopy levels. In Southern Africa, March to April is the transition period from vegetation peak productivity to senescence (Madonsela *et al.* 2017) and (Cho *et al.* 2010). The high separability of *A. mearnsii* during the summer-autumn transition period could be attributed to the fact that the species is an evergreen leguminous species (Somers and Asner 2012; Tomlinson *et al.* 2013) and (Madonsela *et al.* 2017). According to (Tomlinson *et al.* 2013), the canopy of evergreen species remains relatively stable throughout the growing cycle. Thus during leaf fall and senescence evergreen species tends to exhibit a more traditional leaf strategy with higher leaf biomass and longer leaf lifespan when compared to deciduous plants.

Consequently, high biomass enhances the spectral separability between evergreen and deciduous tree, particularly during leaf senescence period (Bai *et al.* 2015). As reported by (Tomlinson *et al.* 2013), tree species that vary in leaf habit (deciduous and evergreen) have different leaf chemical and physiological leaf traits. Likewise (Somers and Asner 2012) and (Madonsela *et al.* 2017) point spectral disparities between deciduous and native species to variations in structural and biochemical constituents during the transition period from March to April (peak productivity to senescence).

2.6.1. Optimal spectral regions for species differences

The full spectrum (400–2500 nm) has been widely used for discriminating vegetation species (Jiménez and Díaz-Delgado 2015b). Among explored hyperspectral regions, VIS, RE, NIR and ESWIR from both leaf and canopy levels were significant in this study. The performance observed from the regions above confirmed the results of (Somers and Asner 2012) and (Ferreira *et al.* 2016) in which infrared region (680-1080 nm), VIS (400-650nm) and SWIR (1300-2500 nm) was the best regions for discriminating between native and invasive Hawaii tree species. At the leaf level, the RE spectral range did not contribute as much as VIS, NIR and ESWIR to distinguish between *A. mearnsii* from the native species. However, the RE extracted from canopy spectra performed better than the MSWIR (1461-1789 nm) and FSWIR (1952-2449 nm). Substantial differences in the VIS, NIR and ESWIR point to high disparities in biochemical and other properties that control spectral signature in these regions between *A. mearnsii* and the native species. *A. mearnsii* is known to be a nitrogen-fixing plant that is often manifested by higher leaf nitrogen content as compared to non-fixing species.

Strong dissimilarities between the species in the SWIR spectral regions could suggest high nitrogen variations between species because it is highly correlated with N-H and C-H vibrations of proteins. Moreover, SWIR was found to be necessary for the discrimination of species based on leaf tannin content, which has been shown to vary between invasive Australian acacias and non-invasive species. For example, (Lehmann *et al.* 2015) discriminated *Acacia longifolia* using four SWIR wavelength regions (1360–1450 nm and 1630–1740 nm) due to their high correlation with tannin concentration. According to (Ferwerda and Jones 2006; Blackburn 2007) the dissimilarities at the SWIR could be attributed to the fact that wavelengths related to tannin are linked to the molecular vibration such as bending and broadening of C-H, C-O and O-H bonds and their overtones. Substantiating the studies by (Cho *et al.* 2007; Somers and Asner 2012; Ferreira *et al.* 2016), we observed disparities in discrimination abilities between leaf and canopy spectral. As in (Somers and Asner 2012)) the canopy level spectra showed a higher discrimination performance compared to leaf spectra. The difference in discrimination abilities among spectral regions, notably VIS and RE at leaf and canopy levels, highlights the significance of scaling up the leaf level spectral to the canopy level for species discrimination, as emphasised by (Cho *et al.* 2008), (Clark *et al.* 2005a), (Kalacska *et al.* 2007; Lehmann *et al.* 2015). Also, (Clark and Roberts 2012b) showed improvement in differentiating species from canopy spectra, as compared to the leaf level. Moreover, (Cho *et al.* 2008) linked the disparities

between leaf and canopy RE spectra in discriminating species to the fact that canopy reflectance affords extra information, such as leaf orientation, leaf clumping and colour of twigs. On the other hand low separability in MSWIR and FSWIR spectral domains could be suggesting that the differences in leaf and canopy water content not contribute significantly to spectral separability.

The identified spectral regions from this study could be useful in providing scientific guidance for the development of tailor-made species mapping platforms like UAV (Müllerová *et al.* 2017; Weil *et al.* 2017). These platforms have been shown to offer a cost-effective solution to address and monitor an invasion challenge (Niphadkar and Nagendra 2016).

2.6.2 Species discrimination based on simulated Sentinel-2 MSI and Landsat-8 OLI

In general, *A. mearnsii* is separable from sampled native species with both Landsat 8 OLI and Sentinel-2 MSI. Both Landsat-8 OLI and Sentinel-2 simulated data could discriminate over 90% of *A. mearnsii* among co-occurring species. This could provide an opportunity to locate the species regularly at the regional scale. However, Sentinel-2 MSI seems to offer more information for discrimination of *A. mearnsii* as compared to Landsat-8. Our results substantiate those of (Sothe *et al.* 2017) in which Sentinel-2 classified sub-tropical forest better than Landsat-8 OLI. Similar to studies by (Ramoelo *et al.* 2015; Ramoelo and Cho 2018) and (Adelabu *et al.* 2014), the current study found critical spectral regions for optimal discrimination of *A. mearnsii* in the RE, NIR and SWIR for Sentinel-2 and NIR and SWIR for Landsat-8 OLI. Likewise, (Adelabu *et al.* 2014) demonstrated improved species discrimination accuracy because of the inclusion of RE and SWIR for Sentinel-2 data. However, numerous studies have noted that the spatial resolution of these sensors might complicate species mapping application, among other things due to spatial variability in the canopy.

2.6.3 Comparison of iECVA-DA and RF-DA models

Both models were able to reduce the redundancy of the hyperspectral data while retaining the most useful information to carry out the discrimination classification. However, the results are shown in this study indicate that the classification results obtained from the RF-DA are slightly better than those obtained from the iECVA-DA classifier. The performance of RF-DA (both scales) corroborates with other studies of plant species discrimination (Prosperre *et al.* 2014).

2.7 Conclusion

The results of the study highlighted the following about the discrimination of *A. mearnsii*.

- i. Invasive *A. mearnsii* can be distinguished from other species at leaf and canopy level using spectral reflectance
- ii. Spectral differences related to senescence phenology had a stronger effect on the separability between *A. mearnsii* and native species.
- iii. Spectral regions associated with biochemical properties are essential in discriminating the species. For instance, there was high spectral separability in RE and visible spectral regions at leaf and the canopy scales.
- iv. A high spectral separability was yielded by using non-parametric RF-DA classifiers. The classifier was able to distinguish between *A. mearnsii* and sampled species even during the Southern Africa senescence-winter transition period (April-May).
- v. Sentinel-2 has advantages compared to Landsat 8- OLI for discriminating between *A. mearnsii* and sampled species.

This study can contribute to initiatives aimed at managing and monitoring invasive *A. mearnsii*. This research provides critical information towards an understanding of the spectral differences between *A. mearnsii* and co-occurring native species. The findings could be useful for understanding vital spectral bands required to delineate and monitor *A. mearnsii* in a mixed species environment effectively.