

Article

Plant Species Discrimination in a Tropical Wetland Using *In Situ* Hyperspectral Data

Kurt Prospere *, Kurt McLaren † and Byron Wilson †

Department of Life Sciences, University of the West Indies, Mona, Kingston 7, Jamaica;
E-Mails: kurt.mclaren@uwimona.edu.jm (K.M.); byron.wilson@uwimona.edu.jm (B.W.)

† These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: kurt.prospere@gmail.com;
Tel.: +1-876-461-8098; Fax: +1-876-977-1075.

Received: 27 June 2014; in revised form: 17 August 2014 / Accepted: 2 September 2014 /

Published: 10 September 2014

Abstract: We investigated the use of full-range (400–2,500 nm) hyperspectral data obtained by sampling foliar reflectances to discriminate 46 plant species in a tropical wetland in Jamaica. A total of 47 spectral variables, including derivative spectra, spectral vegetation indices, spectral position variables, normalized spectra and spectral absorption features, were used for classifying the 46 species. The Mann–Whitney *U*-test, paired one-way ANOVA, principal component analysis (PCA), random forest (RF) and a wrapper approach with a support vector machine were used as feature selection methods. Linear discriminant analysis (LDA), an artificial neural network (ANN) and a generalized linear model fitted with elastic net penalties (GLMnet) were then used for species separation. For comparison, the RF classifier (denoted as RFa) was also used to separate the species by using all reflectance spectra and spectral indices, respectively, without applying any feature selection. The RFa classifier was able to achieve 91.8% and 84.8% accuracy with importance-ranked spectral indices and reflectance spectra, respectively. The GLMnet classifier produced the lowest overall accuracies for feature-selected reflectance spectra data (52–77%) when compared with the LDA and ANN methods. However, when feature-selected spectral indices were used, the GLMnet produced overall accuracies ranging from 79 to 88%, which were the highest among the three classifiers that used feature-selected data. A total of 12 species recorded a 100% producer accuracy, but with spectral indices, and an additional 8 species had perfect producer accuracies, regardless of the input features. The results of this study suggest that the GLMnet classifier can be used,

particularly on feature-selected spectral indices, to discern vegetation in wetlands. However, it might be more efficient to use RFa without feature-selected variables, especially for spectral indices.

Keywords: tropical wetland; feature selection; spectral vegetation indices; derivative spectra; generalized linear model fitted with elastic net penalties; feature selection

1. Introduction

Over the last decade, leaf spectral reflectance has been used successfully to discriminate plant species found in various habitat types/ecosystems [1–4]. In particular, *in situ* hyperspectral measurements greatly assist with the discrimination process by allowing contiguous spectral data to be analyzed statistically. The manipulation of small, but often significant differences in the spectral curves using methods, such as continuum removal, permits the identification of different vegetation types [1,5].

Species differentiation has also been achieved with univariate and multivariate approaches, which include parametric and non-parametric analysis of variance [6–8], discriminant analysis [9] and classification and regression tree-based techniques [10]. These methods can be used individually or several methods can be combined to achieve hyperspectral feature reduction. Mather and Koch [11] described the two main categories of feature reduction: feature extraction and feature selection. Feature selection seeks to reduce the original data set to a subset of features that retain information required to better separate the classes, while excluding highly correlated and redundant features from the classification analysis [12]. Contrastingly, feature extraction applies a transformation that allows the definition of a small set of new features that contain the majority of information contained in the original data set [13].

Band selection techniques have been used to select informative bands in hyperspectral data from spectra collected in different bio-types and to discriminate vegetation types or species [13–16]. However, intra-species variations due to age differences [17], micro-climate [18], edagilogs, topography [19], phenology [20], illumination [20], precipitation and other environmental factors [20] can influence the biophysical and biochemical constituents of a leaf [1,20], and have limited the success of these band-based techniques.

Consequently, hyperspectral variables and indices that are known to be related to foliar pigments have been used to discriminate between canopy species across different landscapes. Spectral indices are mathematical transformations of spectral reflectance that can be used to improve the accuracy of vegetation signals [21]. These indices can also be used to differentiate plant species that differ in canopy structure and/or biochemical composition (example: [2,22]). Therefore, a suite of hyperspectral metrics that indicate vegetation chemical and structural properties are often used for species discrimination [22].

Such species discrimination may prove advantageous to natural resource managers tasked with monitoring vegetation invasions, for example. Hyperspectral data may prove invaluable in landscapes characterized by a high degree of heterogeneity, fragmentation and high biodiversity, such as those

present in the Black River Lower Morass (BRLM) in Jamaica. The sustainable management of different ecosystem types within this wetland requires a detailed understanding of vegetation species distribution and the ability to identify plants accurately and efficiently at the landscape level [23]. Moreover, wetlands inventory and the monitoring of vegetation species quality and distribution are important for management, but can be impeded by the marsh and seasonally-to-permanently flooded conditions. To this end, field spectrometry has been used to characterize the reflectance of vegetation types *in situ* and for the scaling-up of measurements from the leaf to crown scales [9] and within laboratory settings [24].

We therefore attempted to build hyperspectral libraries from the most indicative vegetation species found in remnant swamp forest patches and morass areas in the largest wetland in Jamaica, giving priority to capturing reflectance from endemic vegetation wherever possible. Both reflectance spectra and spectral indices were used in an attempt to discriminate between sampled species. Generalized linear models fitted with an elastic net regularization (GLMnet) is an untested technique for tree discrimination using hyperspectral data. The performance of this supervised learning classifier was compared to linear discriminant analysis (LDA) and neural networks (ANN), two algorithms commonly employed for such tasks. Therefore, the objectives of this investigation were to: (1) compare the performance of LDA, ANN and GLMnet techniques in identifying wetland vegetation species; (2) compare the effectiveness of spectral features and indices derived from five feature selection procedures; and (3) examine the analysis capability of hyperspectral data for identifying vegetation found in different niches in the BRLM. We also compared the ability of the random forest tree-based classifier (RF) to discern species using non-feature-selected reflectance spectra and spectra indices.

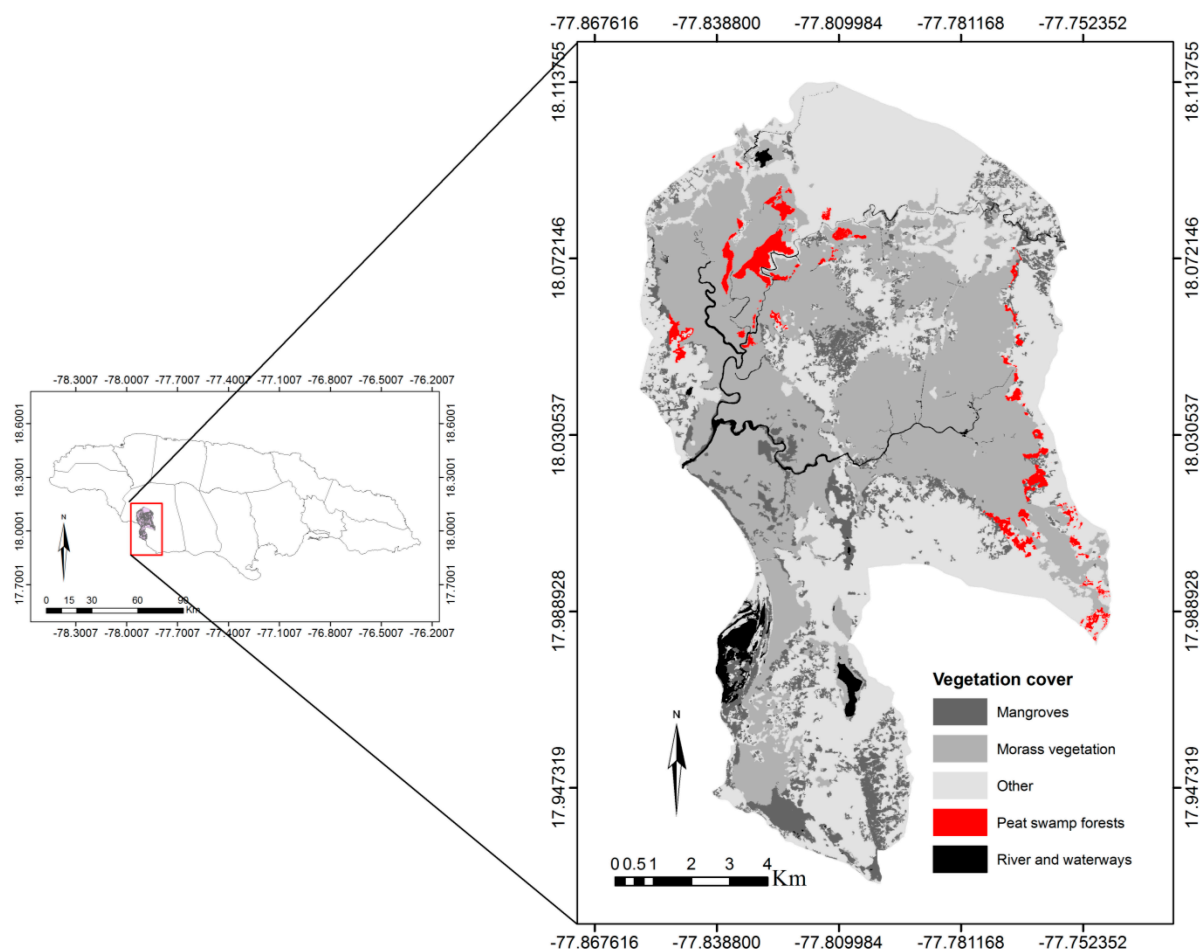
2. Methods

2.1. Study Site

Located in the parish of St. Elizabeth, Jamaica, the Black River Lower Morass (coordinates 18.189553°N and 77.683307°W) has an approximate area of 6,075 ha and is the largest wetland on the island. The BRLM was declared a wetland of international importance in 1998 by the Convention on Wetlands of International Importance, RAMSAR [25]. The boundary of the RAMSAR site was used to define the limits of the study area (Figure 1), and within the study site, there are patches of swamp and mangrove forests, morass, varied types of herbaceous wetland habitats and several raised limestone islands. Five human-residential communities are found on different limestone islands within the confines of the RAMSAR boundaries and at least 6 communities border the area. The 2011 population census estimated that approximately 20,000 persons live in those communities [26].

2.2. Sample Acquisition

To investigate the ability of reflectance patterns of leaves to effectively discriminate plant species, we attempted to collect 20 healthy, sun-lit leaves from at least 15 individual plants per species. This was not always achieved, due to availability, access, leaf position in the canopy (shading), fungal growth and poor leaf health. However, we tried to ensure that the number of sampled plants allowed for spectral variation and a wide spatial distribution between individuals. The list of species sampled and the number of training and test spectra used for the subsequent data analyses are given in Table 1.

Figure 1. The Black River Lower Morass study site, with the official RAMSAR boundary.**Table 1.** Plant species used in this study and the number of spectra used as training and testing samples for data analysis and classification († endemic plant).

Vegetation Species [Family]	Code	Functional Group	Sampled Plants	Spectra	
				Train	Test
<i>Alpinia allughas</i> (Retz.) Roscoe [Zingiberaceae]	ALAL	Herb	27	100	35
<i>Adenanthera pavonina</i> L [Fabaceae]	ADPA	Tree	15	200	100
<i>Amphitecna latifolia</i> (P. Mill) A. H. Gentry [Bignoniaceae]	AMLA	Tree	23	340	51
<i>Andira inermis</i> (W. Wright) Kunth ex. DC. [Fabaceae]	ANIN	Tree	18	225	45
<i>Annona glabra</i> L. [Annonaceae]	ANGL	Shrub-tree	11	128	48
<i>Arundo donax</i> L. [Gramineae]	ARDO	Tall grass	18	75	15
<i>Avicennia germinans</i> L. [Avicenniaceae]	AVGE	Tree	22	270	60
<i>Calophyllum calaba</i> L. [Calophyllaceae]	CACA	Tree	20	225	75
<i>Calyptanthes chytraculia</i> (L) Sw. [Myrtaceae]	CACH	Shrub-tree	18	225	45
<i>Calyptronoma occidentalis</i> (Sw.) H. E. Moore [Arecaceae]	CAOC †	Palm	19	298	80
<i>Casearia guianensis</i> (Aubl.) Urb. [Flacourtiaceae]	CAGU	Shrub-tree	5	20	5
<i>Casearia aculeate</i> Jacq. [Salicaceae]	CAAC	Shrub-tree	15	143	52
<i>Cecropia peltata</i> Miq. [Cecropiaceae]	CEPE	Tree	5	12	8
<i>Chrysobalanus icaco</i> [Chrysobalanaceae]	CHIC	Shrub-tree	20	237	80
<i>Cladium mariscus</i> (L.) Pohl. <i>ssp. jamaicense</i> (Crantz) Kük [Cyperaceae]	CLMA	Graminoid	30	220	110

Table 1. Cont.

Vegetation Species [Family]	Code	Functional Group	Sampled Plants	Spectra	
				Train	Test
<i>Coccoloba longifolia</i> Fisch [Polygonaceae]	COLO [†]	Shrub-tree	21	80	25
<i>Crinum americanum</i> L. [Amaryllidaceae]	CRAM	Forb/herb	25	80	20
<i>Crudia spicata</i> (Aubl.) Willd. [Fabaceae]	CRSP	Tree	19	225	60
<i>Eichhornia diversifolia</i> (Vahl) Urb. [Pontederiaceae]	EIDI	Hydrophyte	26	60	18
<i>Elaeis guineensis</i> Jacq. [Arecaceae]	ELGU	Palm	10	104	26
<i>Eugenia brownie</i> Urb. [Myrtaceae]	EUBR [†]	Shrub	8	55	33
<i>Ficus maxima</i> Mill. [Moraceae]	FIMA	Tree	12	135	45
<i>Ficus pertusa</i> L.f. [Moraceae]	FIPE [†]	Epiphyte	6	24	12
<i>Grias cauliflora</i> L. [Lecythidaceae]	GRCA	Tree	17	135	18
<i>Guettarda argentea</i> Lam. [Rubiaceae]	GUAR	Tree	26	300	90
<i>Gynerium sagittatum</i> (Aubl.) P. Beauv [Poaceae]	GYSA	Tall grass	20	75	25
<i>Hibiscus elatus</i> Sw. [Malvaceae]	HIEL	Tree	20	300	100
<i>Ipomoea tiliacea</i> (Willd.) Choisy [Convolvulaceae]	IPTI	Climber	66	50	16
<i>Lonchocarpus domingensis</i> (Turp. ex. Pers) DC [Fabaceae]	LODO	Tree	20	300	100
<i>Melaleuca quinquenervia</i> (Cav.) S. T. Blake [Myrtaceae]	MEQU	Tree	26	300	90
<i>Nectandra patens</i> (Sw.) Griseb [Lauraceae]	NEPA	Shrub-tree	18	150	30
<i>Persicaria glabrum</i> (Willd) M. Gomez [Polygonaceae]	PEGL	Forb	26	80	24
<i>Philodendron lacerum</i> (Jacq.) Schott [Araceae]	PHLA	Epiphyte	10	42	18
<i>Rhizophora mangle</i> L [Rhizophoraceae]	RHMA	Tree	22	319	32
<i>Roystonea princeps</i> (Becc.) Burret [Arecaceae]	ROPR	Palm	15	208	32
<i>Sabal maritima</i> (Kunth) Burret [Arecaceae]	SAMA	Palm	20	80	20
<i>Sagittaria lancifolia</i> [Alismataceae]	SALA	Emergent	20	60	20
<i>Symphonia globulifera</i> L.f. [Clusiaceae]	SYGL	Tree	23	300	45
<i>Syngonium auritum</i> (L.) Schott [Araceae]	SYAU	Climber	20	45	15
<i>Syzygium jambos</i> [Myrtaceae]	SYJA	Shrub-tree	10	120	30
<i>Tabebuia angustata</i> Britton [Bignoniaceae]	TAAN [†]	Tree	20	223	75
<i>Terminalia latifolia</i> Sw. [Combretaceae]	TELA	Tree	17	165	22
<i>Thalia geniculata</i> L. [Marantaceae]	THE	Forb/herb	27	60	21
<i>Tillandsia fasciculata</i> Sw. [Bromeliaceae]	TIFA	Epiphyte	17	50	33
<i>Trichocentrum luridum</i> (Lindl.) M. W. Chase & N. H. Williams [Orchidaceae]	TRLU	Epiphyte	11	14	6
<i>Typha domingensis</i> Pers. [Typhaceae]	TYDO	Emergent	23	80	12
Total	46	–	887	6,937	1,922

Leaves at different positions in the canopy may have unique spectral attributes, due to differences in photosynthetic properties. Therefore, we stratified our sampling of leaves according to tree height. Specifically, we sampled leaves from the well-lit top, middle and low foliage branches from the crowns of individual plants. The remoteness of our site, coupled with the prevalent soft peat substrate in the morass prohibited the use of ladders, cranes or any other mechanism to access the canopy tops safely. Leaves were collected for immediate processing by either climbing trees, or by using a telescopic pole with a clipper attached to one end, or a combination of the two methods. Consequently, with the exception of *Rhizophora mangle* (RHMA) (which was easy to climb), samples were usually obtained from trees ≤ 10 m in height.

In the field, foliar reflectance was measured immediately after the leaf was cut. However, on days when inclement weather prevented immediate reflectance measurement, clipped samples were transported to a field shelter within 45 minutes, where their reflectance was measured immediately. The leaves were handled in a manner consistent with Liu *et al.* [16], thereby ensuring leaf freshness. We assessed differences in the reflectance pattern of leaves stratified according to canopy heights (the independent variable with three levels using a one-way ANOVA) for each species and found that they were not statistically different ($F_{2,38} = 35.4$; $p < 0.19$). Additionally, the spectra of leaves that were clipped and transported were not significantly different when compared with the spectra of leaves analyzed *in situ* ($F_{2,35} = 30.1$; $p < 0.18$). We therefore pooled the spectra of leaves from the different canopy-strata and those obtained both *in situ* and *ex situ* to calculate the mean reflectance curve for each tree.

2.3. Collection of Reflectance Spectra

During October, 2012, reflectance spectra for 46 plant species (Table 1) were collected with a full-range analytical spectral device (ASD) (Fieldspec®4, Analytical Spectral Devices, Boulder, CO, USA). The ASD instrument consists of three separate spectrometers and covered a spectral range of 350 nm to 2,500 nm. These spectrometers ranged from 350 to 975 nm, 976 to 1,770 nm and 1,771 to 2,500 nm, with spectral resolutions of 3 nm at 700 nm and 10 nm at 1,400 and 2,100 nm, respectively. A leaf-clip assembly was interfaced with the ASD Plant Probe, with an internal Halogen light source, and a 1.5-m fiber optic cable was used with an effective 25° field of view. The leaf-clip holds the target sample in place without removing the leaf from a tree and excludes ambient light, ensuring a constant geometry for the light source and foreoptics.

Fitted to the leaf-clip was a two-sided, rotating head, each embedded with a background panel. The black panel face (less than 5% reflectance) was used for reflectance measurements, and the white panel face was fitted with a 99% reflectance panel (Spectralon, Labsphere, North Sutton, NH, USA). A dark current correction was performed every ten minutes to eliminate instrument noise from spectral measurements, and a white reference measurement was taken to convert leaf radiance to percent reflectance. The spectrometer was programmed to give a spectra with 10 spectral averaging, to obtain reliable mean and variance estimates.

Each leaf, depending on its surface area, was measured in 3 to 5 (3 for most species) spots along an axis perpendicular to the main leaf vein [27]. For smaller leaves, for example *Adenanthera pavonina* (ADPA) and *Avicennia germinans* (AVGE), the field of view occupied over 50% of the surface area of the leaf. In such cases, single-leaf repetitions were not possible; therefore, spectra from other leaflets on the same bi-pinnate stalk were treated as the replicate per leaf. For larger leaves (e.g., *Alpinia allughas* (ALAL)), a minimum of 3 spots per leaf was used. Palm leaves (for example, *Roystonea princeps* (ROPR) and *Calyptronoma occidentalis* (CAOC)) were measured on a central axis parallel with the length of the leaves, as per Benoit *et al.* [27].

2.4. Preprocessing

The preprocessing procedure employed by Pu [2] was used during this study. Reflectance values below 400 nm and above 2,400 nm were truncated and curve smoothing was applied to the remaining

bands by using a simple average over blocks of five neighboring bands. Spectral curves were then normalized by dividing the curve by its mean reflectance, which reduces intraspecies spectral variability by suppressing illumination differences [28]. Thus, the spectral reflectance curve ρ_i , was replaced with $(\rho_i/((1/k)\sum \rho_i))$, where k represents the total bands of the spectral reflectance. The normalized data were then used in all subsequent analyses. All statistical analyses were conducted using the R statistical package Version 3.0.2 [29].

2.5. Extraction of Spectral Variables

The 47 spectral variables used for classifying the sampled species are given in Table 2. These include derivative spectra, spectral vegetation indices, spectral position variables, normalized spectra and spectral absorption features extracted from the *in situ* hyperspectral measurements. Continuum removed spectra features were not included in this study, because they have been reported as ineffective for plant species discrimination (e.g., [1,2]).

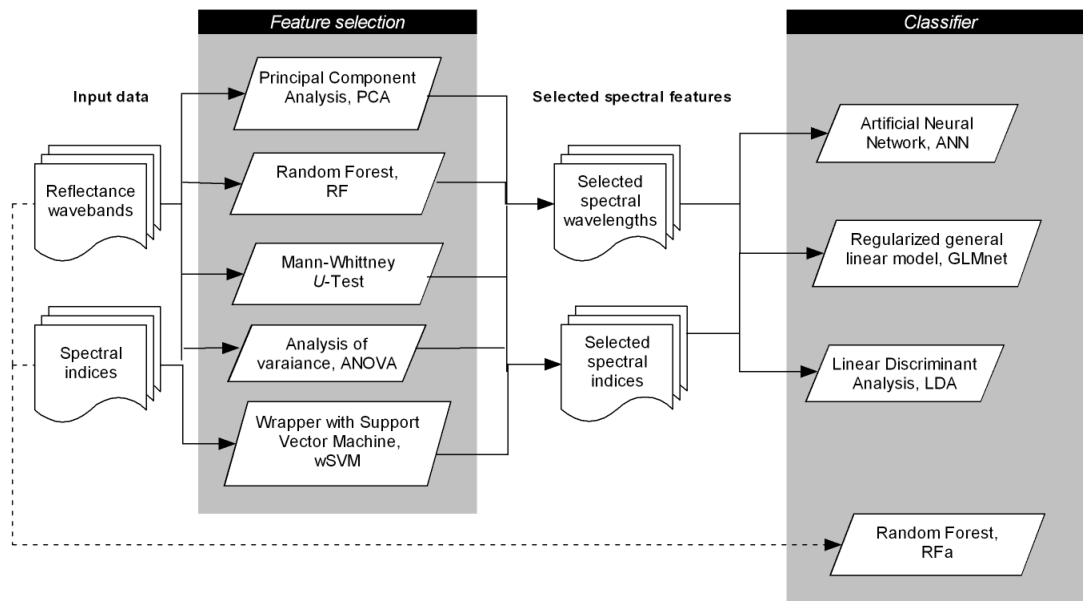
Table 2. Summary of the 7 spectral indices derived from the *in situ* hyperspectral measurements for this study.

Name	Index	Comments	Reference
Normalized Difference Vegetation Index, NDVI	$(\rho_{831} - \rho_{667})/(\rho_{831} + \rho_{667})$	Multiple narrow band versions of this index reported for various applications.	Merton [30]
Photochemical Reflectance Index, PRI	$(\rho_{531} - \rho_{570})/(\rho_{531} + \rho_{570})$	Varies with vegetation type, photosynthetic capacity and efficiency of radiation use; water stress indicator.	Gamon <i>et al.</i> [31]
Gitelson and Merzlyak chlorophyll 1 and 2,	GM1 = $(\rho_{750})/(\rho_{550})$ GM2 = $(\rho_{750})/(\rho_{700})$	Measurement of chlorophyll content in plant leaves.	Gitelson and Merzlyak [32]
Lichtenthaler Indices 1 to 3	Lic1 = $(\rho_{800} - \rho_{680})/(\rho_{800} + \rho_{680})$ Lic2 = $(\rho_{440})/(\rho_{690})$ Lic3 = $(\rho_{440})/(\rho_{740})$	Detection of stress in green vegetation leaves using UV-laser induced fluorescence bands.	Lichtenthaler <i>et al.</i> [33]
Simple Ratio Pigment Index, SRPI	$(\rho_{430})/(\rho_{680})$	Based on the carotenoid/chlorophyll-a content.	Peñuelas <i>et al.</i> [4]
Normalized Phaepophytin Index, NPQI	$(\rho_{415} - \rho_{435})/(\rho_{415} + \rho_{435})$		Barnes <i>et al.</i> [34]
Normalized Pigment Chlorophyll Ratio Index, NPCI	$(\rho_{680} - \rho_{430})/(\rho_{680} + \rho_{430})$	Varies with total pigments/chlorophyll.	Peñuelas <i>et al.</i> [35]
Greenness Index, G	$(\rho_{554})/(\rho_{677})$	Indicator of prolonged vegetation stress due to changes in canopy structure, but not due to biochemical constituents.	
Structure Intensive Pigment Index, SIPI	$(\rho_{445} - \rho_{800})/(\rho_{680} - \rho_{800})$	Semi-empirical estimation of the carotenoids:chlorophyll a ratio.	Peñuelas and Filella [36]
Simple Ratio, SR	$(\rho_{774})/(\rho_{677})$	Indicator of prolonged vegetation stress due to changes in canopy structure.	Gong <i>et al.</i> [37]; Chen [38]
Reflectance at 550 nm, R550	(ρ_{550})	Chlorophyll content	Thomas and Gausman [39]

Table 2. Cont.

Name	Index	Comments	Reference
Reflectance at 680 nm, R680	(ρ_{680})	Chlorophyll content	Thomas and Gausman [39]
Water Index, WI	$(\rho_{900})/(\rho_{970})$	Water status	Peñuelas <i>et al.</i> [40]
Cellulose Absorption Index, CAI	$0.5(\rho_{2000} + \rho_{2200}) - \rho_{2100}$		Nagler <i>et al.</i> [41]
Moisture Stress Index, MSI	$(\rho_{1600})/(\rho_{820})$	Water status	Rock <i>et al.</i> [42]
Normalized Difference Water Index, NDWI	$(\rho_{860} - \rho_{1240})/(\rho_{860} + \rho_{1240})$	Water status	Gao [43]
Disease Water Stress Index, DWSI	$(\rho_{802} + \rho_{547})/(\rho_{1657} + \rho_{682})$	Water status	Gálvao <i>et al.</i> [44]
3-Band ratio at 975 nm, RATIO975	$2 * \rho_{960-990}/(\rho_{920-940} + \rho_{1090-1110})$	Water status	Pu <i>et al.</i> [45]
3-Band ratio at 1,200 nm, RATIO975	$2 * \rho_{1180-1220}/(\rho_{1090-1110} + \rho_{1265-1285})$	Water status	Pu <i>et al.</i> [45]
Leaf Chlorophyll Index, LCI	$(\rho_{850} - \rho_{710})/(\rho_{850} + \rho_{680})$	Chlorophyll content; not affected by scatter variation in the spectra and is essentially a function of pigment absorption.	Datt [46]
Chlorophyll Index, SGA	$(\rho_{750} + \rho_{705})/(\rho_{750} + \rho_{705} - 2 * \rho_{445})$	Chlorophyll content	Sims and Gamon [47]
Chlorophyll Index, SGB	$(\rho_{750} - \rho_{445})/(\rho_{705} - \rho_{445})$	Chlorophyll content	Sims and Gamon [47]
DattA	$(\rho_{780} - \rho_{710})/(\rho_{780} - \rho_{680})$	Chlorophyll content	Datt [46]
Water Index at 1,180 nm, WI.1180	$(\rho_{900})/(\rho_{1180})$		Sims and Gamon [48]
Maximum 1st derivative of 10 different regions within the spectra: A.1D, B.1D, C.1D, D.1D, E.1D, F.1D, G.1D, H.1D, I.1D, J.1D	A.1D: (495–550 nm); B.1D: (550–650 nm); C.1D: (680–780 nm); D.1D (970–1090 nm); E.1D (1,110–1,205 nm); F.1D (1,205–1,285 nm); G.1D (1,310–1,450 nm); H.1D (1,455–1,640 nm); I.1D (1,845–1,925 nm); J.1D (1,925–2,200 nm)	Pigments absorption, water, cellulose, starch and lignin absorption; refer to Table 2 in Pu <i>et al.</i> (2004) for a full description.	Pu <i>et al.</i> [49]
Corresponding spectral positions of the maximum 1st derivatives: A.WP, B.WP, C.WP, D.WP, E.WP, F.WP, G.WP, H.WP, I.WP, J.WP		Pigments absorption, water, cellulose, starch and lignin absorption; refer to Table 2 in Pu <i>et al.</i> (2004) for a full description.	Gong <i>et al.</i> [37]; Pu <i>et al.</i> [49]

An outline of the steps used in this study is shown in Figure 2. All relevant reflectance spectra and spectral indices were randomly split into training (approximately 70% of the data) and test data (approximately 30%) sets. Five different feature selection methods were used to isolate spectra and identify indices that were most effective at discriminating among species. The training and test data sets were subsequently modified to include only those indices and spectra identified during the respective feature selection procedures. Each feature-selected training data set was then used for training the 3 different classifier models. The test data sets were then used to assess the classifier's performance by comparing their overall accuracy and kappa statistic. All of the training spectra and indices were respectively classified with a random forest classifier, without applying any prior feature selection.

Figure 2. Flowchart showing a summary of the process followed in this study.

2.6. Feature Selection and Discrimination Procedure

Feature selection by means of recursive feature elimination uses a feature-ranking criterion to produce a list of features arranged by their discriminatory ability and therefore provides a means by which the most parsimonious model can be selected. For this study, a statistically-based criterion that estimated the importance of the waveband features was used to separate features. Spectra and indices were first standardized to ensure that they had a mean of 0 and a standard deviation of 1 before feature selection. A support vector machine algorithm with a 10-fold cross-validation was implemented using the R package e1071 (Version 1.6-1) [50], coupled with the caret package [51] to select the ‘optimal’ number of variables for the reflectance spectra and spectral indices data sets that produced the highest accuracy. A feature set of 15 indices and 20 spectra yielded the most accurate results; therefore, the top 15 and top 20th ranked spectral indices and reflectance spectra, respectively, per feature selection procedure, were used for subsequent species separations. To allow for comparability, the same random number seed (100) was set prior to modeling.

We then used recursive feature elimination that included resampling from the caret package [51] using a random forest (RF) algorithm in the R package randomForest [52] and a wrapper using a support vector machine (wSVM) feature selection method implemented in the R package FSelector [53]. The RF approach is an embedded method of feature selection that uses recursive partitioning, producing an ensemble of classification trees that are calculated on random subsets of the data [54]. For each resampling iteration, the algorithm partitions the data into training and hold-back sets via resampling. The model is then trained on the training sets using all predictors and subsequently performs a prediction on the held-back samples. Variable importance or rankings were then calculated by keeping the S_i most important variables for each subset size S , for which $i = 1 \dots S$. The training data set was used to train the model using S_i predictors, and the held-back samples were used to make predictions. The performance profile over S_i using the held-back samples was then calculated, and the appropriate number of predictors was determined.

2.6.1. *U*-Test

A Mann–Whitney *U*-test was used to determine if the variance of reflectance between tree species was greater than within tree species [1]. A non-parametric test was chosen because hyperspectral data are not independent and are not normally distributed [55]. Unequal sample sizes did not affect the statistical test, because the number of samples was sufficiently large [56]. The *U*-test was used to test the hypothesis that there was no significant difference between the median reflectance of each individual waveband between pairs of tree species [1]. Schimdt and Skidmore [1] described the null hypothesis for *N* vegetation types and *J* spectral bands per reflectance measurement as:

$$H_0 : \eta_n(i) = \eta_{n+1}(i) \quad (1)$$

where *i* is the spectral band, η_n is the median reflectance for vegetation type number $n = 1, 2, 3, \dots, (n-1)$ and $i = 1, 2, 3, \dots$. For this study, the maximum frequency was 1,035 (${}_{46}C_2 = (46 \cdot (46-1))/2$). The hypothesis was therefore tested 1,035 times for all possible combinations of the 46 species at the significance levels of $\alpha' = 0.05$ and $\alpha'' = 0.01$, with Bonferroni adjustments in both cases.

2.6.2. Analysis of Variance, ANOVA

Adjacent hyperspectral wavebands show high correlations, and it is therefore not efficient or reliable to include all measured bands for analyses [57]. Most classification algorithms underperform when reflectance values for highly correlated data are used during model training. To minimize such effects, we performed a paired one-way ANOVA for all possible combinations of species in the training data set across the 47 spectral indices. The resultant probability value obtained from the ANOVA analysis provided an index of importance for the tested index that was used to discern between tree species [57].

The ANOVA was used to eliminate redundant spectral bands from the analysis. However, because adjacent bands are not independent, an ANOVA could not be used to test within-band differences. We recorded the frequencies of each spectral variable (reflectance spectra and spectral indices) that the ANOVA identified as having interspecific variation between the paired species at the 95 and 99% probability levels. The frequencies obtained from the one-way ANOVA were then used to select a subset of variables for running the classifier algorithms.

2.6.3. Principal Component Analysis, PCA

PCA is a multivariate statistical technique that can be used to extract information from spectral data and transform the data into a set of orthogonal variables called principal components (PC). Because neighboring bands of hyperspectral data are highly correlated, a PCA was used to transform the original data into its PCs. This reduces irrelevant information from the original inter-correlated variables. The uncorrelated linear combinations (eigenvector weights) of variables in *n*-dimensional space are then chosen to successively extract linear combinations that have corresponding smaller variances. The first PC accounts for the maximum proportion of the variance, and subsequent components account for the next highest proportion of the remaining variance [58]. PCA wavebands were computed using factor loadings (or eigenvectors) for each of the bands and by multiplying the factor loadings by their respective waveband's reflectivity [15].

2.6.4. Random Forest Selector

Random forest comprises a collection of decision tree classifiers [59], in which each tree in the forest has been trained using a bootstrap sample of training data, and a random subset of features is sampled independently from the input features [13]. By omitting a subset of the training data set from the training of each plant species, random forests are better able to examine the contribution and behavior of each predictor (spectral band or index) [60]. The out-of-bag data were used for feature selection by determining the importance of different spectral wavelengths or indices during the classification process, based on a Z score, which was used to assign a significance level.

2.6.5. The Wrapper Approach with a Support Vector Machine: Wrapper-SVM

Wrapper feature selection uses the induction algorithm as a black box method [61] and combines the strength of a traditional search algorithm with the capability of a classifier [7]. Therefore, cross-validation or bootstrapping should account for the variability caused by feature selection when assessing performance. In this study, performance estimates that included feature selection variation were generated using the rfe function (caret package), which induces recursive feature elimination with resampling. Spectral indices and reflectance spectra feature sets that were permitted with the rfe function were 1, 2, 3, 4, 5, 10, 15, 20 and 30, using a support vector machine with a 10-fold cross-validation and 10 repeats of the bootstrap for optimization. The predictors (spectral indices and spectral bands, respectively) were ranked; the less important predictors were sequentially eliminated prior to modeling.

2.7. Classifier Training, Prediction and Accuracy Assessment

Three classifiers, namely a linear discriminant analysis (LDA), an artificial neural network (ANN) and a generalized linear model fitted with elastic net penalties (GLMnet), were used to distinguish the plant species. RF was used as a fourth classifier (denoted as RFa to distinguish it from the same algorithm used for feature selection); however, instead of using the feature selected variables as inputs, the full complement of spectral bands and indices were used. The RFa classifier uses predictor variable importance (measured as the decrease in overall classification accuracy when the variable is permuted in the out-of-bag samples) as the criteria to build the best model for discerning species.

Spectral wavebands and spectral indices with the highest discriminating powers determined from the feature selection procedures were used to create respective training sets. For each training set, a resampling iteration was performed in which specific samples were ‘held-out’, and the model was fitted on the remaining samples. Predictions were subsequently extracted from the ‘held-out’ samples and the average performance on the held-out predictions determined. The predetermined optimal parameter set was used to generate a final model that included all of the training data. For every combination of classification method and training set, a 10-fold cross-validation with 10 repeats of the bootstrap was performed in the caret package. To ensure that the same resampling sets were used, the random seed was set to 100 prior to each model run. Estimations of model performance were conducted on the training set, while the withheld test set was used to evaluate the classifier’s performance. The kappa index and overall accuracy (OAA) were calculated for the algorithms used to

discern between the species (LDA, ANN, GLMnet and RFa), by using each classifier's respective confusion matrix.

2.7.1. Linear Discriminant Analysis (LDA)

LDA is a parametric classifier that has been used to discern vegetation types or to identify tree species (e.g., [2,9,57,62]). LDA uses a pooled within-class covariance matrix and spectral predictor variables from training samples to build the discrimination functions for each class. Hence, the original redundant data are projected to a new orthogonal space oriented along the axis that can maximize the ratio of between-class to within-class variance among training samples [57]. The LDA function within the *caret* package was used to implement the LDA algorithm.

2.7.2. Artificial Neural Networks (ANN)

Artificial neural networks (ANNs) are non-parametric statistical data modeling tools that have been used successfully to discriminate remotely-sensed vegetation (e.g., [2,63]). We used a multilayered perceptron neural model with a fully-connected feed-forward, supervised learning network, trained by the back-propagation algorithm to minimize a quadratic error criterion. In a layered structure, the input to each node is the sum of the weighted outputs of the nodes in the prior layer, which are connected to the input spectral features (the wavelengths and indices selected from feature selection procedure). For the respective spectra and indices for each feature selection method, we used an output layer containing as many neurons as classes into which the samples were differentiated.

Wavebands and indices that were identified by the feature-selection method as being most important for discriminating species were used as the input nodes for the multilayered perception implemented in the MLP function from the RSNNS package, Version 0.4-3 [64]. *RSNNS* implements an R interface to the Stuttgart Neural Network Simulator, SNNS [65]. Five hidden layers with a maximum of 100 iterations were used for learning, with randomized weights as the initialization function. The parameter for the learning function was a standard back-propagation, with a topological order update function.

2.7.3. Generalized Linear Models with an Elastic Net Regularization (GLMnet)

Generalized linear models (GLMs) are mathematical extensions of linear models that do not force data into unnatural scales and thereby allow for non-linearity and non-constant variance structures in data [66]. However, this popular machine learning technique does not penalize for the size of estimated coefficients, therefore limiting its performance. By introducing a penalty term, the elastic net [67] penalizes the size of estimated coefficients by using a combination of two regularization techniques [68], the l_2 regularization (used in ridge regression) and the l_1 regularization, used in lasso [69]. The penalty term, $P_\alpha(\beta)$ is defined as:

$$P_\alpha(\beta) = (1 - \alpha)/2 * |\beta|_2^2 + \alpha |\beta|_1 \quad (2)$$

where P_α is a compromise between the ridge-regression penalty (*Alpha*, $\alpha = 0$) and the lasso penalty ($\alpha = 1$). In applying the l_1 penalty, lasso attempts to achieve a parsimonious solution. This idea has been broadly applied, for example to generalized linear models [69]. Lasso attempts to ensure that

most of the variable coefficients will be shrunk to 0, so the least significant variables are removed from the model. Contrastingly, the ridge penalty shrinks all variables, but not to 0.

For comparisons with the popular LDA and ANN classifiers used for hyperspectral species discrimination, we used an extension of Friedman *et al.* [68] to classify the hyperspectral data in our training data set (both spectral wavebands and indices). A GLM was fitted with an elastic-net regularization via the R package glmnet (Version 1.9-5) [68]. Alpha (α) values of 0, 0.5 and 1 and lambda (λ) values ranging from 0 to 0.05 with increments of 0.01 were used as the tuning grid to select the optimal model. The response type was chosen as Gaussian, and a Newton logistic type was used.

3. Results

3.1. Feature Selection

A support vector machine using recursive feature elimination, a 10-fold cross-validation and 10 repeats of the bootstrap selected the ‘optimal’ number of variables to be taken from the different feature selection methods. Ultimately, a set of 15 spectral indices produced the lowest root mean squared error, RSME (5.81 ± 0.119 standard deviation, SD), while a group of 20 spectra wavelengths produced the lowest RSME of 6.02 ± 0.792 SD.

3.1.1. ANOVA

A subset of the most frequent spectral indices was selected from all of the spectral metrics using a one-way ANOVA. Figure 3 illustrates the frequency distribution of the ANOVA for all of these indices between every paired species, arranged in descending order of frequency. The pre-determined selected number of features meant that the first 15 indices (H.1D to LIC.1 from left to right) were subsetted. Five vegetation indices (VI), Moisture Stress Index (MSI), Cellulose Absorption Index (CAI), NDWI, Disease Water Stress Index (DWSI) and NDVI, were within that subset. The spectral variables H.1D, F.1D, Simple Ratio (SR), R550, C.1D, formed a sub-group of variables that were related to leaf pigment status, notably chlorophyll content of leaves among the different species.

Figure 4a summarizes the ANOVA results of all possible species combination pairs showing species pairs that were statistically different per wavebands. Of the 401 wavebands assessed, 296 were able to discriminate over 800 pairs of species and 392 wavebands showed a discrimination frequency of over 700. The highest frequency was observed in the shortwave infrared range of the spectrum, and relatively low frequencies were noted in the visible range of the spectrum.

3.1.2. Mann–Whitney *U*-Test

A Mann–Whitney *U*-test was used to test the null hypothesis 1,035 times at the adjusted Bonferroni significance levels of 95% and 99% for the spectral variables used in this study. The lowest performing variables were the spectral positions (A-WP to J-WP), corresponding to the maximum first derivative spectra (A-1D to J-1D) (Figure 5). Of the first 15 spectral derivatives (H.1D to SR from left to right) that were selected, 11 (73.3%) were also among the ANOVA-based selected features; these included H.1D, CAI, RATIO1200, WI1180, MSI, F.ID, NDWI, DWSI, R550, RATIO975 and SR.

Figure 3. Frequency distribution of ANOVA for the 47 spectral variables for paired species across the 46 species assessed. Each bar represents the number of paired species between which a spectral variable difference is significant.

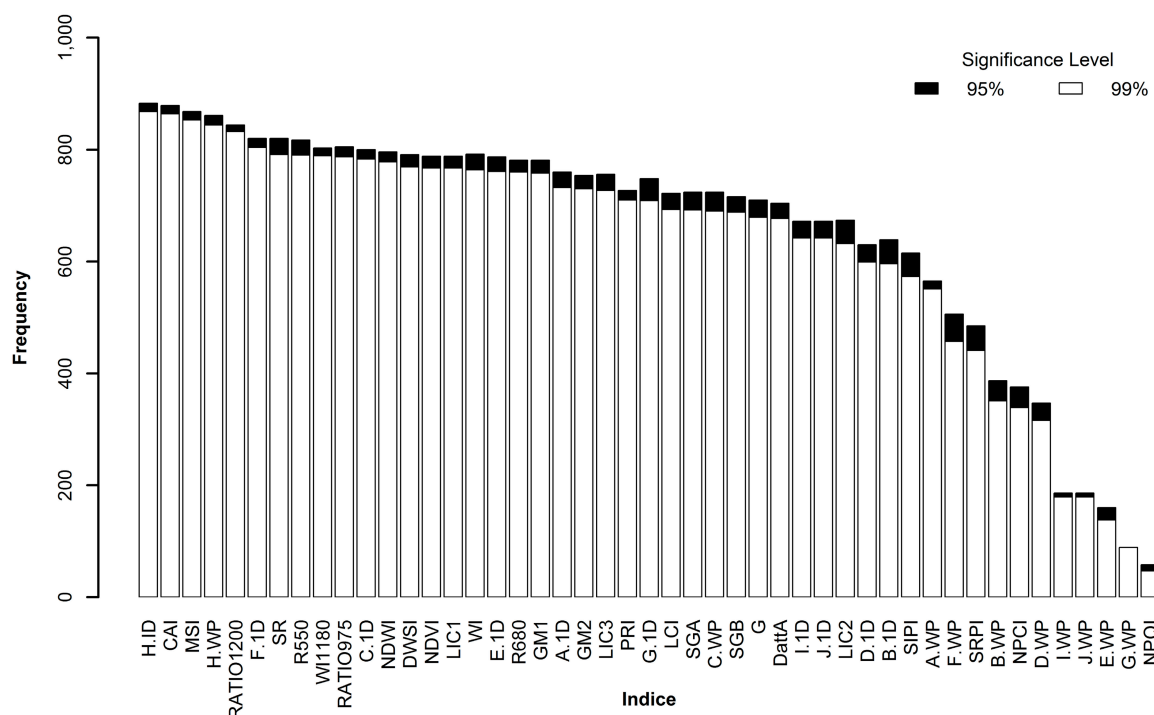
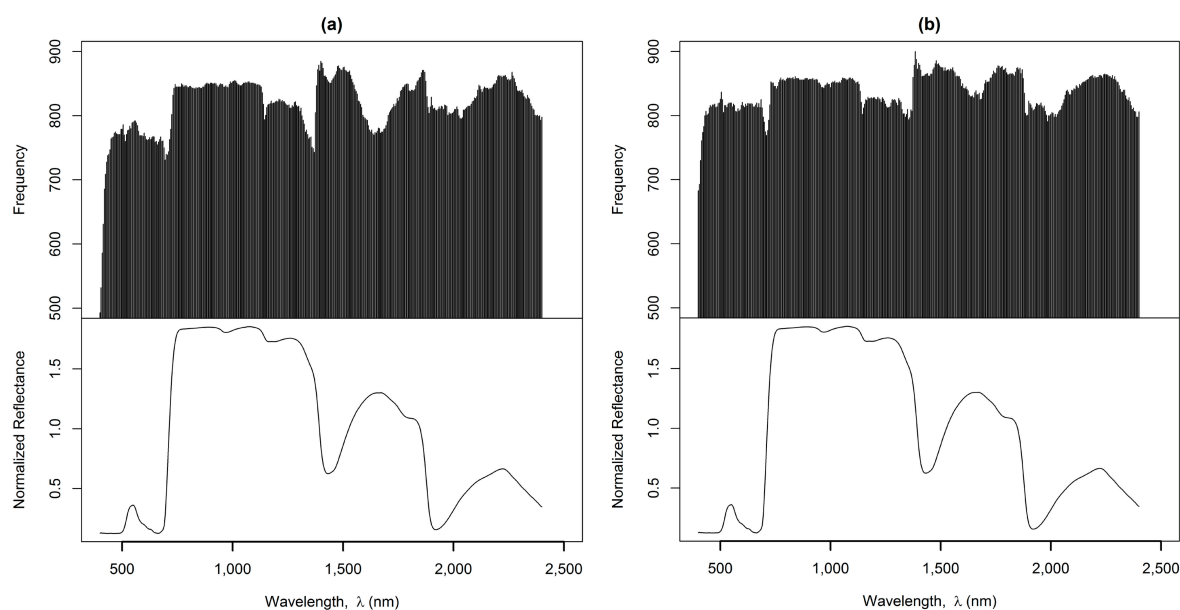
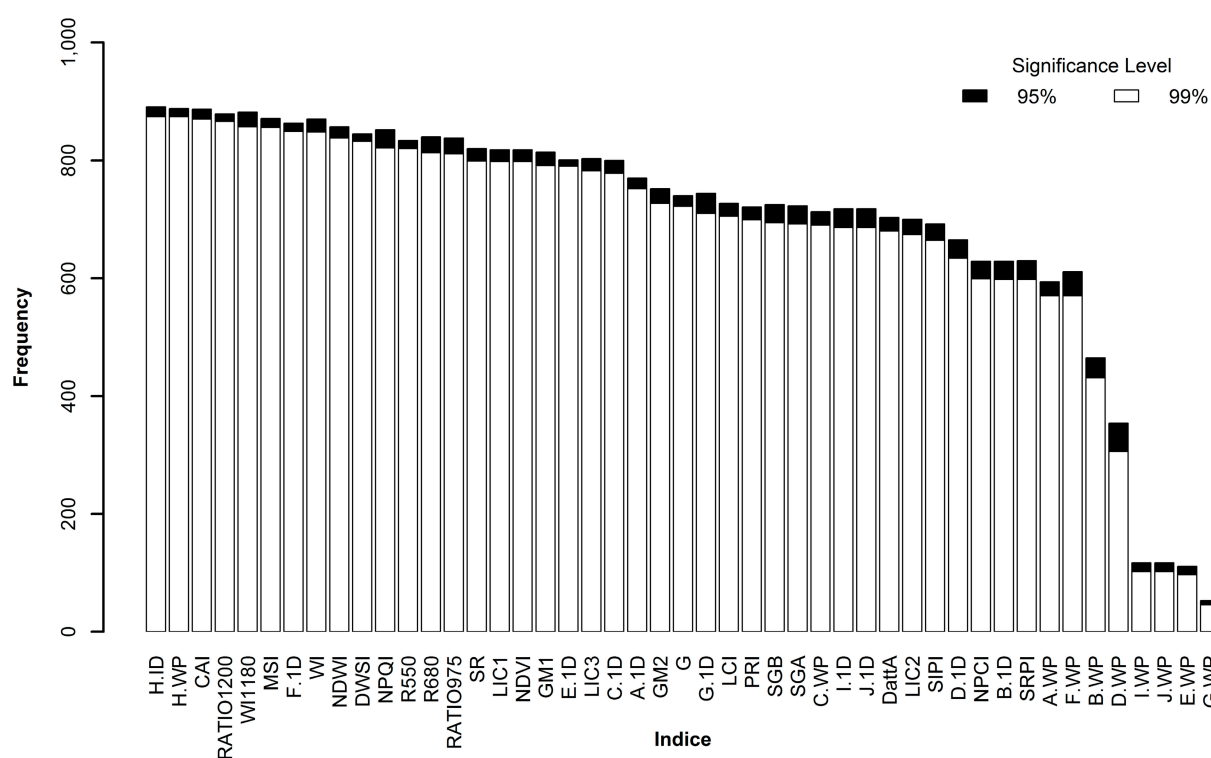


Figure 4. (a) Frequency distribution of ANOVA of all 401 5-nm wavebands from 400 to 2,400 nm (top), for every paired species across the 46 species. Bars represent the number paired species for which the spectral variable difference is significant at the 0.01 probability level. (b) (Top) Frequency plot of statistically significant differences using the Mann–Whitney U -test of all 401 5-nm wavebands from 400 to 2,400 nm, for every paired species across the 46 species. Bars represent the frequency of paired species for which a spectral variable difference is significant at the 9.7×10^{-6} significance level. For both (a) and (b) (bottom), the mean reflectance curve of TELA (*Terminalia latifolia*) to indicate typical vegetation reflectance features.



The results of all possible species combinations and the frequency of species pairs with a statistically significant difference per waveband are shown in Figure 4b, with the mean normalized reflectance of TELA (*Terminalia latifolia*) plotted as a reference for the position of the main features of a typical leaf reflectance curve. At the 99% significance level (Bonferroni adjusted), the wavebands at 1,385, 1,390 and 1,480 nm scored the highest frequencies overall, with the waveband at 1,385 nm recording the highest frequency (900). In contrast, the lowest frequencies (683, 693 and 730) were obtained for the wavebands at 400, 405 and 410 nm, respectively. On average, the *U*-test was able to discriminate 12 additional pairs of species combinations at $\alpha = 0.05$, compared to the $\alpha = 0.01$ significance level. Relatively low frequencies were noted for the infrared plateau compared with the relatively higher frequencies occurring in the shortwave infrared regions (SWIR).

Figure 5. Frequency distribution of the Mann–Whitney *U*-test for the 47 spectral variables for paired species across the 46 species studied. Each bar represents the number of paired species for which a spectral variable difference is significant.



3.1.3. Principal Component Analysis, PCA

The first five PCs explained at least 95% of the variation for all samples within the training spectra data set, and the first eight PCs explained at least 99%. However, for some species, the target variabilities (95% and 99%) were obtained before the stated number of PCs. For example, the first five PCs were able to explain at least 99% of the variability for CEPE (*Cecropia peltata*) and *Trichocentrum luridum* (TRLU).

For reflectance values, the first PC only contained wavebands from the SWIR and near-infrared (NIR) (700 to 1,100 nm) regions. Wavebands from the SWIR region dominated the first PC with a 68% frequency occurrence, and the near-infrared (NIR) (700 to 1,100 nm) accounted for the remaining 32% frequency of occurrence (14 out of 44 species). Conversely, when all of the PCs that accounted

for at least 99% of the variation were pooled, bands from the visible region had a high frequency of occurrence. A summary of these pooled results for the PCA for band selection revealed that three main clusters were distinguishable; one in the visible (400 to 560 nm), another in the red edge slope (680 to 750 nm) and the last in the SWIR (1,340 to 1,545 nm).

3.2. Species Separation

Overall, spectra data produced higher accuracies compared to variable indices, irrespective of the algorithm used (Table 3). Generally, GLMnet produced higher accuracies on the spectral indices when compared with the spectral waveband data. The GLMnet, ANN and LDA classifiers produced overall accuracies that ranged from 52–77%, 74–87% and 83–87%, respectively (Table 3). A one-way ANOVA test confirmed significant differences between the performance of the three classifiers ($F_{2,12} = 9.68$; $p = 0.003$) for spectra reflectance using feature selected wavebands. *Post hoc* comparisons using the Tukey HSD test indicated that the GLMnet ($M = 65.79$, $SD = 11.2$) produced significantly lower accuracies at the 95% CI, when compared with LDA ($M = 85.19$, $SD = 2.07$) and ANN ($M = 81.67$, $SD = 5.9$), respectively. One-way ANOVA also confirmed significant differences between the three classifiers when feature-selected spectral variables were used ($F_{2,12} = 8.70$; $p = 0.004$). ANN produced significantly lower accuracies ($M = 60.12$, $SD = 5.11$) when compared with the GLMnet classifier ($M = 80.41$, $SD = 8.49$) and LDA ($M = 74.39$, $SD = 9.44$), respectively, confirmed with a Tukey HSD test.

Table 3. Summary of species identification accuracies using linear discriminant analysis (LDA), artificial neural network (ANN), generalized linear models with an elastic net regularization (GLMnet) and random forest (RFa) algorithms with spectral wavebands and indices.

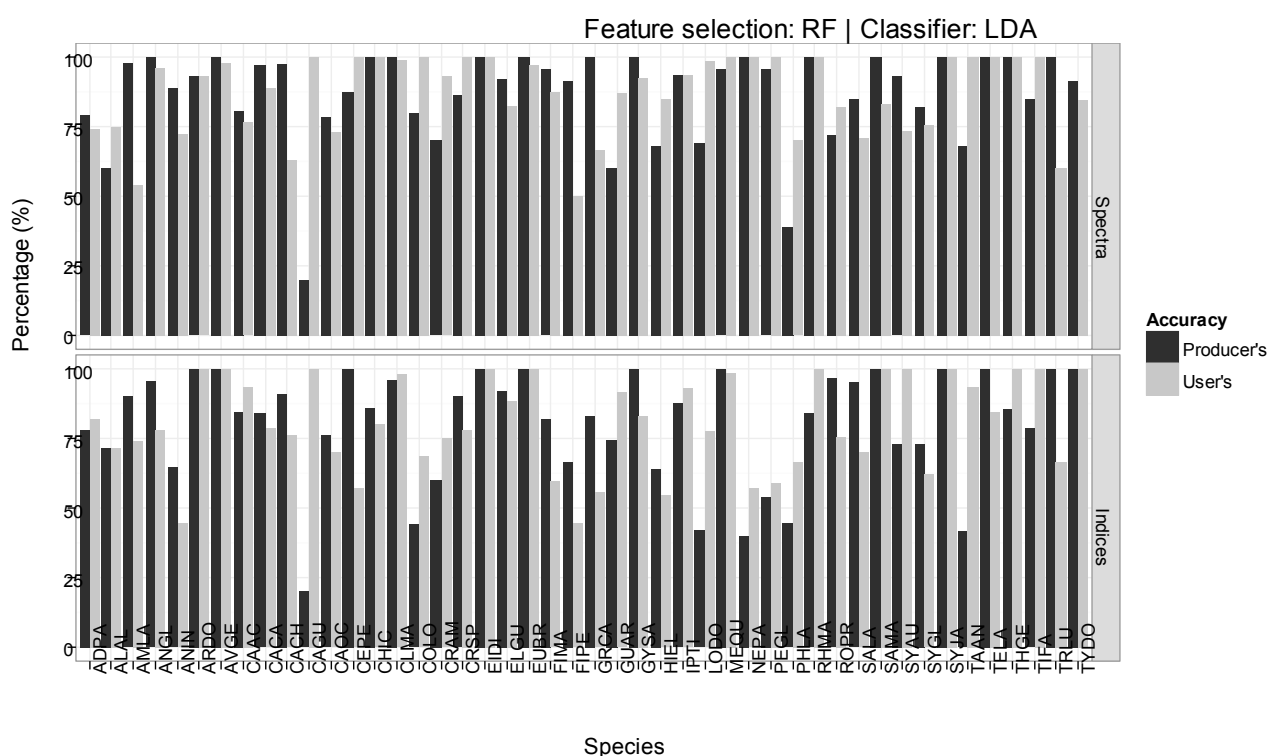
Classifier	Feature Selection	Spectra			Indices		
		OAA (%)	OAA Range (%)	Kappa	OAA (%)	OAA Range (%)	Kappa
LDA	U-test	86.84	85.24–88.32	0.8641	78.62	76.71–80.43	0.7794
	ANOVA	83.40	81.66–85.04	0.8286	78.62	76.71–80.43	0.7794
	PCA	82.57	80.80–84.24	0.8201	57.54	55.29–59.77	0.5619
	RF	86.06	84.43–87.58	0.8561	79.34	77.46–81.13	0.7868
	wSVM	87.07	86.26–87.85	0.8666	77.84	75.91–79.67	0.7712
ANN	U-test	77.06	75.11–78.92	0.7633	63.84	61.65–65.99	0.6269
	ANOVA	73.88	71.86–75.83	0.7305	56.87	54.62–59.10	0.5555
	PCA	83.71	81.99–85.34	0.8319	63.74	61.54–65.89	0.6257
	RF	86.42	84.80–87.92	0.8792	63.48	61.28–65.63	0.6229
	wSVM	87.29	86.48–88.06	0.8688	52.65	50.39–54.91	0.5102
GLMnet	U-test	51.56	49.30–53.82	0.4982	79.19	77.30–80.98	0.7850
	ANOVA	56.58	54.41–58.89	0.5514	85.02	83.34–86.58	0.8452
	PCA	77.26	75.23–79.12	0.765	66.23	64.07–68.35	0.6515
	RF	74.40	72.02–75.99	0.7315	87.62	86.06–89.06	0.8721
	wSVM	69.17	68.06–70.25	0.6811	83.98	82.26–85.59	0.8344
RFa	RFa	84.81	83.12–86.38	0.8430	91.80	90.52–93.01	0.9160

The RFa classifier was able to discern species with 91.8% accuracy when 24 spectral indices were used as input variables (Table 3). Without prior feature selection, the RFa chose 201 wavebands in its optimal model, and achieved an overall accuracy of 84.8% in discerning the 46 plant species. However, in some instances with feature selection, the more parsimonious models yielded higher accuracies (up to a 2.5% higher) compared with the RFa. However, the LDA classifier, using features selected from RF, produced relatively high accuracies and kappa statistics for both spectra and spectral variables. The ANOVA and *U*-test were the only two methods among the different feature-selected methods that selected bands exclusively from the mid-infrared part of the spectra (1,300–2,500 nm).

3.3. Producer's and User's Accuracies

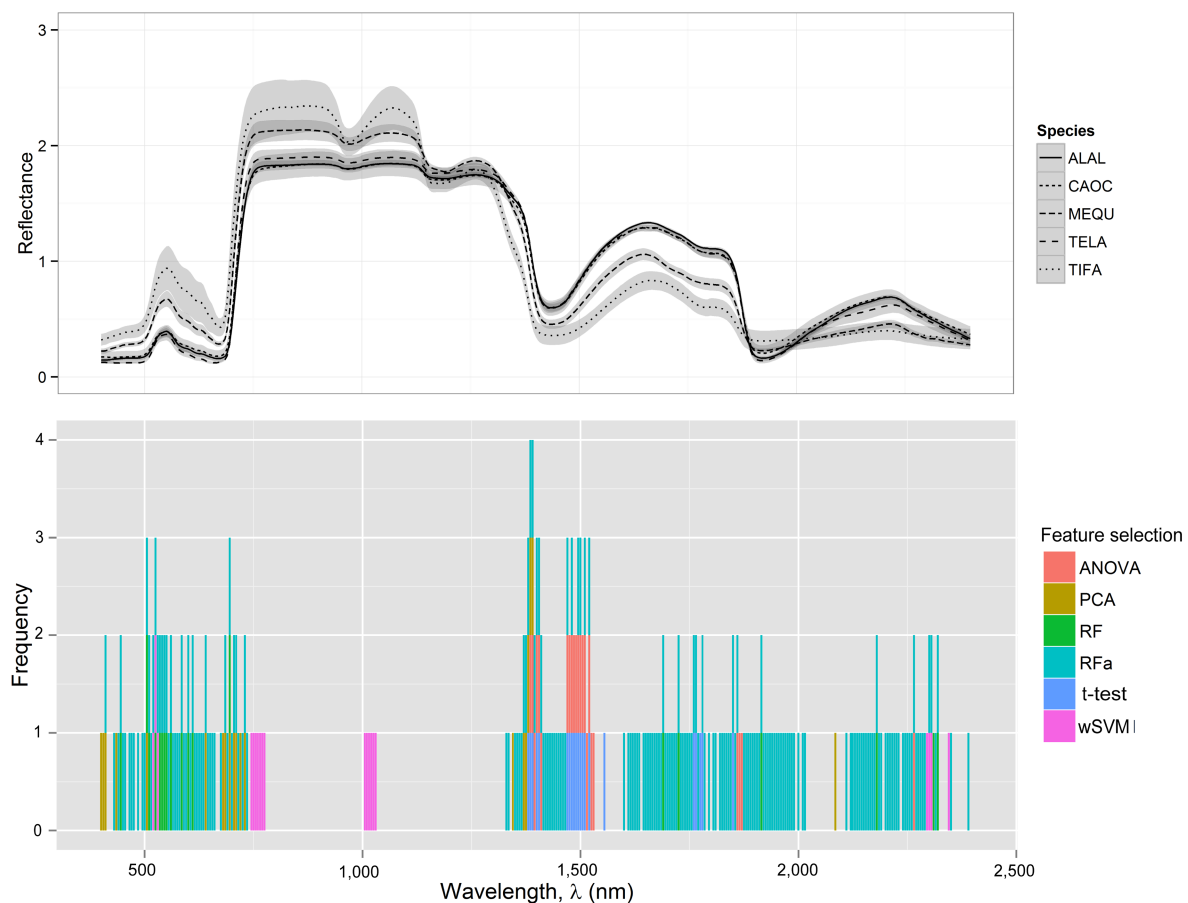
Of the 46 species, 15 were easily separated using RF feature selection on spectra reflectance and the LDA classifier based on their producer's accuracy. A total of 12 species recorded a 100% producer accuracy, using the same combination of feature selection and classification methods, but with spectra-derived variables, and an additional eight species (Figure 6)—AVGE (*Avicennia germinans*), EIDI (*Eichhornia diversifolia*), EUBR (*Eugenia brownie*), GYSA (*Gynerium sagittatum*), SAMA (*Sabal maritima*), SYJA (*Syzygium jambos*), *Terminalia latifolia* (TELA) and *Trichocentrum luridum* (TRLU)—had perfect producer accuracies regardless of input features (spectral reflectance or spectral derivatives). EIDI, EUBR, TELA and TRLU were among the most easily distinguished species, or species that recorded full accuracy percentages at least 50% of the time, irrespective of the combination of classifier, feature selection and input features used. In contrast, there was greater spectral confusion among the epiphytes, PHLA (*Philodendron lacerum*) and FIPE (*Ficus pertusa*), the invasive tree, MEQU (*Melaleuca quinquenervia*), and the native shrub, CAGU (*Casearia guianensis*).

Figure 6. Producer's and user's accuracy for 46 study species using the RF feature selection method and the LDA classifier for spectra and spectra-derived indices.



The wavebands selected using the five spectral discrimination methods (ANOVA, *t*-test, RF, PCA and w-SVM) were merged to determine their frequency of occurrence, and the distribution of these bands along the spectral axis formed several distinct clusters (Figure 7).

Figure 7. Selected regions where discriminating wavebands from the different feature selection methods occurred most frequently (a). The mean spectral reflectance curve for five ‘indicator’ species with different plant forms (b), with their respective 95% standard deviations shown for reference; the species are ALAL (*Alpinia allughas*), a tall grass, and MEQU (*Melaleuca quinquenervia*), a tree, both of which are invasive species. TELA (*Terminalia latifolia*) is a near threatened (International Union for Conservation of Nature ICUN, 1998) broadleaf species. TIFA (*Tillandsia fasciculata*) is a common epiphyte with a confined distribution within the Lower Morass. CAOC (*Calyptronomia occidentalis*) is a native shrub found mainly in the heavily disturbed swamp forest patches.



In the visible spectral range, one cluster comprised spectra selected from at least three feature selection methods. The wavelength range of this region was 400–610 nm. In the red edge region, wavebands formed a cluster at both ends of the slope (680–775 nm). Other clusters were located in the far near-infrared (FNIR) and short-wave infrared (SWIR) segments of the spectrum, with spectra located at 1,380, 1,385 and 1,390 nm being selected by the PCA, ANOVA and *U*-test feature selection procedures, respectively.

4. Discussion

The different feature selection methods used in this study demonstrated varied measures of effectiveness when used with the different classifiers. For the feature-selected spectral wavebands, LDA proved to be the most effective classifier, whereas the regularized GLM correctly discerned more plant species based on their associated spectral indices. Instead of using feature-selected inputs, RFa was able to produce the highest accuracy (92%) with a less parsimonious set of 24 spectral indices. A comparison of results from other studies that used different or similar feature selection methods and the results of this study are found in Tables 4 and 5 for spectral reflectance and spectral indices variables, respectively. The comparisons show that our results were able to identify wavelengths in the three regions of the spectra commonly used for discerning plants based on their foliar spectral reflectance (the visible, NIR and mid-NIR bands).

Table 4. Comparison of selected bands between previous studies and this study used for species discrimination.

Study	Selection Method	No. of Bands	Visible Bands (400–700 nm)	Near-Infrared Bands (700–1,300 nm)	Mid-Infrared Bands (1,300–2,500 nm)
Thenkabail <i>et al.</i> [15]	–	12	490, 520, 550, 575, 660, 675, 700	720, 845, 905, 920, 975	
Schmidt and Skidmore [1]	–	6	404, 628	771	1,398, 1,803, 2,183
Bajwa <i>et al.</i> [70]	Entropy		627–684		
	Derivative		690–705	740–756, 810–825	
	ANN		530–550, 690–710	740–750	
	PCA		690–710		
Vaiphasa <i>et al.</i> [7]	–			720, 1277	1,415, 1,644
Adam and Mutanga [10]				745, 746, 892, 932, 934, 958, 961, 989	
	PCA	20	405, 410, 435, 505, 640, 680, 685, 695	705, 710, 720, 730	1,345, 1,370, 1,375, 1,380, 1,385, 1,390, 2,085
	U-test	21			1,380–1,390, 1,400, 1,405, 1,470–1,510, 1,520, 1,555, 1,760, 1,765, 1,775, 1,780, 1,850
	ANOVA	23			1,385–1,410, 1,470–1,530, 1,860, 1,865, 1,870, 2,265
This study	RF	20	445, 505, 510, 525, 535–560, 585, 600, 610, 695		1,690, 1,725, 1,915, 2,180, 2,310, 2,320
	wSVM	20	520, 525, 530, 745–770	775, 1,005, 1,010, 1,015, 1,020, 1,025, 1,030	2,295, 2,300, 2,305, 2,345
	RFa	201	430, 440, 470, 530–540, 560, 565, 590, 630, 635	705	1,415, 1,725, 2,120, 2,140, 2,160, 2,280, 2,315
		(top 20 listed)			

Table 5. Comparison of selected spectra indices between previous studies and this study used for species discrimination.

Study	Selection Method	VIS-NIR (Photosynthetic Pigments, LAI, Structure, Physiology, Stress)	NIR (Water and Structure)	SWIR (Lignin, Cellulose, Nitrogen, Proteins)
Clark and Roberts [22]	RF	RVSI, GP.W, RE.W, ARI2	WBI, NRI-A	SWIR3-D, SWIR-A, Normalized difference nitrogen index (NDLI)
Pu [2]	ANOVA	A.WP, C.1D,C.WP, A.1D, B.1D, B.WP, NDVI, PRI, SIPI, SR, NPQI, NPQI, R680, R550	DWSI, NDWI, LCI, Absorption feature depth at 975 nm DEP-975, AREA-975, DEP-1200, AREA-1200, WI, RATIO1200, RATIO975	CAI, H.WP, H.1D, E.ID, F.1D
Benoit <i>et al.</i> [27]	-	CI (DattA), DattB, SGA, SGB, SR, PRI, NDVI,	WI.970, WI.1180, NDWI, MSI	CAI
Cho <i>et al.</i> [71]	-	NDVI, CI, PRI, CRI, Vogelman index VOG, GMI		
Kalacska <i>et al.</i> , [72]		R500, mSR-705, mND-705	R800, WBI-970	
This study	PCA	A.WP, B.1D, B.WP, C.WP, G, SR		CAI, D.WP, E.WP, F.WP, G.1D, H.WP, I.1D
	<i>U</i> -test	SR, R550, NPQI, RATIO680	MSI, NDWI, DWSI, RATIO975, WI, WI.1180, RATIO1200	CAI, F.1D, H.1D, H.WP
	ANOVA	C.1D, NDVI, LIC1, R550, SR,	DWSI, NDWI, RATIO975, WI.1180,	CAI, F.1D, H.ID, H.WP
	RF	A.1D, A.WP, C.1D, G, NPQI, PRI	MSI, RATIO975, RATIO1200	CAI, E.1D, F.1D, H.1D
	wSVM	R550, A.1D,B.1D, C.1D	WI, WI.1180, NDWI, DWSI, RATIO1200, RATIO975, MSI	CAI, E.1D, H.1D
	RFa	LIC1, SR, C.WP, A.1D, C.1D, SR, G, LIC3, SIPI, SGB, SGA, R550,	RATIO975, NDWI, WI, DWSI, RATIO1200, MSI	CAI, D.WP, H.WP, E.WP, J.ID D.1D, I.1D

The results from our study suggest that classification performance is improved, at least with the ANN, when bands from different parts of the spectrum are chosen. Lower OAA and kappa values were obtained for the ANN and GLMnet classification when the ANOVA and *U*-test feature-selected sets were used (Table 3). The ANOVA and *U*-test were the only selection methods that isolated bands from a single part of the spectrum (the mid-infrared region). Our results agree with previous studies (e.g., [1,2]) that have illustrated the relative importance of using different parts of the spectrum for species discrimination, particularly the wavebands of the NIR plateau.

4.1. Performance of the Different Classifiers

In this study, the LDA generally outperformed the ANN and GLMnet classifiers, especially with feature-selected reflectance wavebands. The nonlinear ANN can usually handle both parametric and nonparametric data sets, while the LDA is theoretically limited to parametric data sets. *In situ* spectral measurements collected from individual species are assumed to follow a normal distribution [2]; therefore, one would expect non-significant species recognition accuracies if the performance of the LDA and ANN were compared. In fact, accuracies for the two classifiers were not significantly

different for feature-selected spectral waveband data. However, the ANN classifier failed to accurately distinguish between plant species when spectra-derived indices were used. This finding is consistent with other studies that have found ANNs to produce relatively lower classification results with remote sensing data when compared with traditional methods (e.g., [73]).

Although the RFa is a well-known classifier, it failed to top the accuracies attained by the LDA algorithm, even on a less parsimonious set of input features. The LDA simultaneously uses all of the predictor variables to estimate predictor covariance, allowing it to distinguish between classes. Conversely, the RFa distinguishes classes by individually building decision spaces for each explanatory variable at each node level; therefore, there is interdependence between the nodes, and as such, the final classification is ultimately dependent on the decision spaces at higher nodes [23]. This translates to higher misclassification rates, especially for reflectance data, which can be highly variable for different samples of foliar reflectance from one plant species, leading to many different possible splits for the decision tree. However, the RFa algorithm should be robust even in the presence of spectral variability, as the classifier minimizes errors from a single decision tree by selecting random samples, generating hundreds of decision trees and using a majority vote to make the final decision. The relatively large sample size of spectra for some species compared to others (Table 1) may have increased the spectral variability in plants with more foliar spectra measured. This may have caused an overlap in the foliar spectra of some species, making it more difficult for the RFa classifier to distinguish species yielding spectra with a high degree of overlap.

Even when used for feature selection, the RF algorithm was able to produce the highest accuracy (87.6%) when combined with the GLMnet classifier, compared with the LDA and ANN. The GLMnet uses a generalized linear model (GLM) along with an elastic net regularization. The GLM allows for response variables that have error distribution models other than a normal distribution, and the elastic net regularization applies constraints to the lasso and to the ridge parameters [70]. Therefore, we would expect the distribution of the wavebands and spectral indices to have minimal effect on the performance of the GLMnet classifier. Furthermore, the lasso constraints control the selection or removal of variables in the model, while the ridge handles collinear variables. By controlling the relative weighting of these two constraints, the elastic net regularization is able to handle highly correlated data. These are desirable qualities of a classifier when dealing with hyperspectral data, thus we would expect the GLMnet classifier to give consistently high accuracies and its performance to be unaffected by data from different feature-selection methods. However, the findings of this study do not support these assertions. Although the GLMnet gave significantly higher accuracies with spectral indices when compared with spectral wavebands, PCA feature-selected data significantly lowered (in the case of spectral indices) or increased (for reflectance wavebands) the accuracies.

4.2. Reflectance Spectra vs. Spectral Indices

In this study, the spectral metrics out-performed the spectral reflectance data sets only when the GLMnet classifier was applied; but the opposite was true for the LDA and ANN classifiers. When data from all of the feature-selected spectra wavebands were pooled, they were found to aggregate in several clusters along the wavelength axis (Figure 7). Therefore, irrespective of the feature-selection method used, different spectral regions were able to differentiate the tested species. Furthermore, in most cases, there was no overlap among the bands selected using the different feature selection

methods, but instead, neighboring bands had comparable discriminating power when selected by other methods. This is consistent with the results of other studies that used foliar reflectance properties for species discrimination

Indices derived from spectral reflectance and reflectance wavebands were shown in this study to produce different accuracies when used by the same classifier. Spectral differences among species can be affected by illumination, and these differences are better captured with reflectance spectra [22]. However, it can be assumed that illumination did not significantly affect our results, because we used an artificial light source and excluded ambient light. Moreover, spectral data in this study were collected from an illuminated source under controlled conditions, thus eliminating the need to apply filtering or corrections to the raw data. Indices normally minimize brightness variation from band ratios and derivative analyses [22], but shading effects were negligible in this study. Accordingly, chemical absorptions and leaf structure were responsible for species separations.

4.3. Inter- and Intra-Plant Spectral Variability

Spectral variability among individual species can be attributed to differences in internal leaf structure and to leaf biochemical composition [9], most notably water and chlorophyll content [20], epiphyll cover and leaf morphology [74]. However, in this study, the poor performance of the spectral indices may be indicative of pigment absorption in the visible region and water, cellulose, starch and lignin spectral absorption in the near-infrared (NIR) and short wave infrared (SWIR) regions of the spectrum (spectral positions A-WP to J-WP, corresponding to the maximum first derivative spectra A-1D to J-1D). However, the spectral signatures of plant species are also affected by factors that are not limited to age, vitality or physiological characteristics [75]. Inter-species leaf variability can also be attributed to the measurement of bidirectional reflectance, instead of hemispherical reflectance [9]. In this study, most of the spectra metrics selected by the feature selection methods were directly related to leaf chemistry. Therefore, spectral responses due to the concentration of leaf pigments and other bio-chemicals may elucidate greater spectral variation than provided by differences due solely to leaf morphology.

4.4. Implications for Natural Resource Management

We were able to achieve vegetation species discrimination from leaf spectral reflectance using data obtained from controlled illumination at the leaf level. This was an initial step towards the ultimate goal of discriminating and mapping wetland vegetation species and communities using hyperspectral sensors based on an airborne platform at the landscape level. Clark *et al.* [9], Kalacska *et al.* [72], Cho *et al.* [71] and, more recently, Clark and Roberts [22] successfully demonstrated that leaf-level methods can be scaled up to the canopy level to facilitate spectral discrimination of plant/tree species from different types of tropical forest (rain forest, dry forest and mangroves) at the landscape level.

However, discerning species from heterogeneous habitats in wetlands has been accomplished with varying levels of success. Schmidt and Skidmore [1] used hyperspectral remote sensing to map 27 salt-marsh grass and herbaceous plant species, by assessing the canopy level reflectance spectra of several vegetation associations occurring in a Dutch salt-marsh. They were able to demonstrate that separability can be achieved for most plant species from the marshland. However this success has not

been widely replicated, especially for discriminating among tropical wetland species. This can be attributed to the high spectral and spatial variabilities associated with herbaceous wetland vegetation and associated steep environmental gradients, which produce short ecotones and sharp demarcations between vegetation units [10,76].

Furthermore, the reflectance spectra of wetland vegetation canopies are often very similar and can be confused with the reflectance spectra of the underlying soil, hydrologic regime and atmospheric vapor [77]. As a result, for wetlands such as the BRLM, the spatial or spectral optical classification normally employed in remote sensing may result in low classification accuracies. Moreover, the ability of hyperspectral data to effectively distinguish individual species within flooded wetland environments is reduced, because the performance of near to mid-infrared bands are attenuated by the presence of underlying water and wet soil [76,78]. However, several authors (e.g., [1,7,78]) have used the narrow spectral channels offered by hyperspectral data to detect and map the spatial heterogeneity of wetland vegetation.

Hyperspectral sensors have been used for the early detection [79], mapping and monitoring [80] of the introduction and spread of invasive plant species in wetland environments. Wetlands are highly susceptible to plant invasions, which threaten the biodiversity and ecological integrity of such systems [81]. Fourteen of the 46 species used in this study are non-native, and at least two, MEQU and ALAL, are highly invasive to the BRLM. Management of this RAMSAR site would entail landscape mapping to monitor the introduction, presence and spread of such invasive species and to identify the location of single nuclear trees for eradication exercises.

4.5. Limitations

Despite attaining spectral separation of the vegetation tested in this study, there were several limitations to our analyses. First, we did not conduct leaf tissue chemical assays. Therefore, we were unable to relate leaf chemical properties to spectral indices. Ideally, chemical and hyperspectral data should be collected at the same time so that the chemical constituents of the leaves can be correlated to the hyperspectral data. This would minimize pseudo-replication from sites within the wetland that show micro-variations (e.g., vegetation in brackish and saline *vs.* freshwater locations, limestone *vs.* peat substrate). Furthermore, sampling was conducted during one month of the rainy season (October). To account for possible changes in physio-chemical parameters, for example water stress indicators, the collection of spectra, ideally on the same individuals sampled, should have been conducted during a representative month in the dry season (March or April).

Furthermore, several studies have used phenological changes in invasive plants to better discriminate them from non-introduced or native plants (e.g., [80,82,83]). However, the phenologies of the species used in this study were not considered during the sampling and collection of spectra. Clark and Roberts [22] used seven tropical, canopy-emergent species to demonstrate the effectiveness of using spectral metrics derived from leaves, bark and a combination of leaves and bark during phenological changes in the plants, to discriminate between species. In this investigation, we only assessed the foliar reflectance of the vegetation, and ignored branch and bark spectral reflectance.

Although we found feature selection to be effective and RFa efficient (prior feature selection not required) in distinguishing species based on their foliar reflectance, it should be noted that accuracies might decay markedly at coarser spatial scales [9]. In this study, the RFa's accuracy of approximately

92% might be satisfactory for *in situ* species differentiation. However, for remotely-sensed images, atmospheric effects should be corrected or compressed prior to conducting species recognition analyses, especially in wetland environments [9]. Wetland environments require special analytical techniques, because saturation and atmospheric vapor affect the near-infrared region [84]. Atmospheric correction should enhance the spectral separability between species with hyperspectral remote sensing data, but water absorption bands in the mid-infrared region should be considered. Mapping the wetland landscape requires remotely-acquired hyperspectral data collected by sensors placed on airborne or satellite-based platforms. Although we have demonstrated the applicability of using spectral reflectance and reflectance indices with feature selection, and the efficiency of the RFa algorithm, our analyses were limited to leaf-level, *in situ* conditions. If remotely-sensed hyperspectral data collected by sensors placed on airborne or satellite-based platforms are to be used to map and monitor vegetation changes in highly fragile and heterogeneous tropical wetland ecosystems such as the BRLM, the next step is to determine whether these leaf-level methods of spectral discrimination can be scaled up to the canopy and the landscape levels, and whether these methods can be successfully replicated in different tropical wetlands. However, to address the effects of atmospheric conditions and structural and vegetative complexities on remotely-acquired reflectance spectra, additional spectral indices may need to be explored or the effective wavebands used for distinguishing canopies of differing species composition may need to be revised.

5. Conclusions

We presented an application of leaf-level hyperspectral data for species discrimination using five feature selection methods (ANOVA, *U*-tests, PCA, RF and wSVM) and four classifiers (LDA, ANN, GLMnet and RFa) to discriminate among 46 flora species under wetland conditions. Both spectral reflectance and spectral indices were used, and feature selection proved helpful in obtaining parsimonious models that were able to discern between the leaves of different species with accuracies of approximately 88% for waveband and spectral index variables. However, the highest accuracy (92%) was achieved using the RFa classifier with spectral indices, but at the expense of using a less parsimonious model. However, one can question the efficiency of the feature selection step, since this step itself might be computationally expensive, compared to including more features. Nevertheless, the spectral discrimination of invasive plants, grasses, a floating macrophyte, endemic shrubs and trees, as well as both native and non-native climbers and epiphytes, was achieved at the individual foliar level. This result was obtained under controlled data-collection conditions. It is therefore the first step towards the ultimate goal of using hyperspectral remote sensing to discriminate and map the canopies of different vegetation types in a wetland environment, using the BRLM as a focal ecosystem. This will be used to support current initiatives aimed at managing and monitoring invasive flora and monitoring fragile/threatened habitats, such as the remnant fragments of swamp forests in the BRLM.

Acknowledgments

We greatly acknowledge ASD Inc. and the Alexander Goetz Instrument Support Program (AGISP) who provided, via a scholarship, the spectrometer and necessary instrumentation used to collect data for this study. Financial support was also provided through the United Nations Environmental Program

(UNEP), the National Environmental Planning Agency of Jamaica (NEPA), Commonwealth Agricultural Bureaux International (CABI), the Government of Jamaica “Mitigating the threats of alien invasive species in the insular Caribbean” project and the MacArthur Foundation.

Author Contributions

K. Prospere carried out field work, data analysis and is the main author of all sections of the manuscript. K. McLaren secured supporting funding, supervised research and helped with formulating the methodology. B. Wilson provided editorial advice, is the co-principal investigator and helped with formulating the methodology.

Conflicts of Interest

The authors declare no conflict of interest.

References

- Schmidt, K.S.; Skidmore, A.K. Spectral discrimination of vegetation types in a coastal wetland. *Remote Sens. Environ.* **2003**, *85*, 92–108, DOI: 10.1080/01431160152609245.
- Pu, R. Broadleaf species recognition with in situ hyperspectral data. *Int. J. Remote Sens.* **2009**, *30*, 2759–2779, DOI:10.1080/01431160802555820.
- Allard, D.; D’Or, D.; Froidevaux, R. An efficient maximum entropy approach for categorical variable prediction. *Eur. J. Soil Sci.* **2011**, *62*, 381–393, DOI: 10.1111/j.1365-2389.2011.01362.x.
- Peñuelas, J.; Baret, F.; Filella, I. Semi-imperical indices to assess carotenoids/chlorophyll, a ratio from leaf spectral reflectance. *Photosynthetica* **1995**, *31*, 221–230.
- Abdel-Rahman, E.M.; Ahmed, F.B.; Van den Berg, M. Estimation of sugarcane leaf nitrogen concentration using *in situ* spectroscopy. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, S52–S57, DOI: 10.1016/j.jag.2009.11.003.
- Mutanga, O.; Skidmore, A.K.; van Wieren, S. Discriminating tropical grass (*Cenchrus ciliaris*) canopies grown under different nitrogen treatments using spectroradiometry. *ISPRS J. Photogramm. Remote Sens.* **2003**, *57*, 263–272, DOI: 10.1016/S0924-2716(02)00158-2.
- Vaiphasa, C.; Ongsomwang, S.; Vaiphasa, T.; Skidmore, A.K. Tropical mangrove species discrimination using hyperspectral data: A laboratory study. *Estuarine, Coastal and Shelf Science* **2005**, *65*, 371–379, DOI: 10.1016/j.ecss.2005.06.014.
- Artigas, F.J.; Yang, J. Spectral discrimination of marsh vegetation types in the New Jersey meadowlands, USA. *Wetlands* **2006**, DOI: 10.1672/0277-5212(2006)26[271:SDOMVT]2.0.CO;2.
- Clark, M.L.; Roberts, D.A.; Clark, D.B. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sens. Environ.* **2005**, *96*, 375–398, DOI: 10.1016/j.rse.2005.03.009.
- Adam, E.; Mutanga, O. Spectral discrimination of papyrus vegetation *Cyperus Papyrus* in swamp wetlands using field spectrometry. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 612–620, DOI: 10.1016/j.isprsjprs.2009.04.004.
- Mather, P.M.; Koch, M. *Computer Processing of Remotely-Sensed Images: An Introduction*, 3rd ed.; Wiley Online Library: Chichester, UK, 2004; p. 460.

12. Motoda, H.; Liu, L.H. *Feature Extraction, Construction and Selection: A Data Mining Perspective*; Springer: Norwell, MA, USA, 1998.
13. Pal, M.; Foody, G.M. Feature selection for classification of hyperspectral data by SVM. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2297–2307, DOI: 10.1109/TGRS.2009.2039484.
14. Schmidt, K.; Skidmore, A. Exploring spectral discrimination of grass species in African rangelands. *Int. J. Remote Sens.* **2001**, *22*, 3421–3434, DOI:10.1080/01431160152609245.
15. Thenkabail, P.S.; Enclona, E.A.; Ashton, M.S.; Legg, C.; De Dieu, M.J. Hyperion, IKONOS, ALI, and ETM plus sensors in the study of African rainforests. *Remote Sens. Environ.* **2004**, *90*, 23–43, DOI: 10.1016/j.rse.2003.11.018.
16. Liu, X.; Sun, Y.; Wu, Y. Reduction of hyperspectral dimensions and construction of discriminating models for identifying wetland plant species. *Spectroscopy and Spectral Analysis* **2012**, *32*, 459–464, DOI: 10.3964/j.issn.1000-0593(2012)02-0459-06.
17. Gausman, H.; Allen, W. Optical parameters of leaves of 30 plant species. *Plant Physiol.* **1973**, *52*, 57–62.
18. Asner, G.P.; Wessman, C.A.; Bateson, C.; Privette, J.L. Impact of tissue, canopy, and landscape factors on the hyperspectral reflectance variability of arid ecosystems. *Remote Sens. Environ.* **2000**, *74*, 69–84, DOI: 10.1016/S0034-4257(00)00124-3.
19. Turner, W.; Spector, S.; Gardiner, N.; Fladeland, M.; Sterling, E.; Steininger, M. Remote sensing for biodiversity science and conservation. *Trends Ecol. Evol.* **2003**, *18*, 306–314, DOI: 10.1016/S0169-5347(03)00070-3.
20. Asner, G.P. Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sens. Environ.* **1998**, *64*, 234–253, DOI: 10.1016/S0034-4257(98)00014-5.
21. Qi, J.; Moran, M.; Cabot, F.; Dedieu, G. Normalization of sun/view angle effects using spectral albedo-based vegetation indices. *Remote Sens. Environ.* **1995**, DOI: 10.1016/0034-4257(95)00034-X.
22. Clark, M.L.; Roberts, D.A. Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. *Remote Sens.* **2012**, *4*, 1820–1855, DOI: 10.3390/rs4061820.
23. Nagendra, H. Using remote sensing to assess biodiversity. *Int. J. Remote Sens.* **2001**, *22*, 2377–2400, DOI: 10.1080/01431160117096.
24. Milton, E.J.; Schaepman, M.E.; Anderson, K.; Kneubahler, M.; Fox, N. Progress in field spectroscopy. *Remote Sens. Environ.* **2009**, DOI: 10.1016/j.rse.2007.08.001.
25. RAMSAR. The list of wetlands of international importance. The secretariat of the convention on wetlands (Ramsar, Iran, 1971), 2012. Available online: <http://www.ramsar.org/pdf/sitelist.pdf> (accessed on 4 August 2014).
26. The Statistical Institute of Jamaica. *Population and Housing Census 2011 Jamaica*; General Report; The Statistical Institute of Jamaica: Jamaica, 2012, Volume 1.
27. Rivard, B.; Sanchez-Azofeifa, G.A.; Foley, S.; Calvo-Alvarado, J.C. Species classification of tropical tree leaf reflectance and dependence on selection of spectral bands. In *Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests*; CRC Press: Boca Raton, FL, USA, 2008; pp. 141–159, DOI: 10.1201/9781420053432.ch6.

28. Yu, B.; Ostland, M.; Gong, P.; Pu, R. Penalized discriminant analysis of *in situ* hyperspectral data for conifer species recognition. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 2569–2577, DOI: 10.1109/36.789651.
29. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.
30. Merton, R. Monitoring community hysteresis using spectral shift analysis and the red-edge vegetation stress index. In Proceedings of the Seventh Annual JPL Airborne Earth Science Workshop, Pasadena, CA, USA, 12–16 January 1998; pp. 12–16.
31. Gamon, J.A.; Field, C.B.; Goulden, M.L.; Griffin, K.L.; Hartley, A.E.; Joel, G.; Penuelas, J.; Valentini, R. Relationships between NDVI, canopy structure, and photosynthesis in 3 Californian vegetation types. *Ecol. Appl.* **1995**, *5*, 28–41, DOI: 10.2307/1942049.
32. Gitelson, A.A.; Merzlyak, M.N. Remote estimation of chlorophyll content in higher plant leaves. *Int. J. Remote Sens.* **1997**, *18*, 2691–2697, DOI: 10.1080/014311697217558.
33. Lichtenthaler, H.K.; Lang, M.; Sowinska, M.; Heisel, F.; Miede, J.A. Detection of vegetation stress via a new high resolution fluorescence imaging system. *J. Plant Physiol.* **1996**, *148*, 599–612, DOI: 10.1016/S0176-1617(96)80081-2.
34. Barnes, J.D.; Balaguer, L.; Manrique, E.; Elvira, S.; Davison, A.W. A reappraisal of the use of DMSO for the extraction and determination of Chlorophylls-a and Chlorophylls-b in lichens and higher-plants. *Environ. Exp. Bot.* **1992**, *32*, 85–100, DOI: 10.1016/0098-8472(92)90034-Y.
35. Peñuelas, J.; Gamon, J.A.; Fredeen, A.L.; Merino, J.; Field, C.B. Reflectance indexes associated with physiological-changes in nitrogen-limited and water-limited sunflower leaves. *Remote Sens. Environ.* **1994**, *48*, 135–146, DOI: 10.1016/0034-4257(94)90136-8.
36. Peñuelas, J.; Filella, I. Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends Plant Sci.* **1998**, *3*, 151–156, DOI: 10.1016/S1360-1385(98)01213-8.
37. Gong, P.; Pu, R.; Heald, R.C. Analysis of *in situ* hyperspectral data for nutrient estimation of giant Sequoia. *Int. J. Remote Sens.* **2002**, *23*, 1827–1850, DOI: 10.1080/01431160110075622.
38. Chen, J.M. Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Can. J. Remote Sens.* **1996**, *22*, 229–242.
39. Thomas, J.; Gausman, H. Leaf reflectance vs. Leaf chlorophyll and carotenoid concentrations for eight crops. *Agron. J.* **1977**, *69*, 799–802, DOI: 10.2134/agronj1977.00021962006900050017x.
40. Peñuelas, J.; Pinol, J.; Ogaya, R.; Filella, I. Estimation of plant water concentration by the reflectance water index WI (r900/r970). *Int. J. Remote Sens.* **1997**, *18*, 2869–2875, DOI: 10.1080/014311697217396.
41. Nagler, P.L.; Inoue, Y.; Glenn, E.P.; Russ, A.L.; Daughtry, C.S.T. Cellulose absorption index (CAI) to quantify mixed soil-plant litter scenes. *Remote Sens. Environ.* **2003**, *87*, 310–325, DOI: 10.1016/j.rse.2003.06.001.
42. Rock, B.; Vogelmann, J.; Williams, D.; Vogelmann, A.; Hoshizaki, T. Remote detection of forest damage. *BioScience* **1986**, *36*, 439–445.
43. Gao, B.C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **1996**, DOI: 10.1016/S0034-4257(96)00067-3.
44. Gálvao, L.S.; Formaggio, A.R.; Tisot, D.A. Discrimination of sugarcane varieties in southeastern Brazil with EO-1 Hyperion data. *Remote Sens. Environ.* **2005**, DOI: 10.1016/j.rse.2004.11.012.

45. Pu, R.; Ge, S.; Kelly, N.M.; Gong, P. Spectral absorption features as indicators of water status in coast live oak (*Quercus agrifolia*) leaves. *Int. J. Remote Sens.* **2003**, *24*, 1799–1810, DOI: 10.1080/01431160210155965.
46. Datt, B. Visible/near infrared reflectance and chlorophyll content in *Eucalyptus* leaves. *Int. J. Remote Sens.* **1999**, *20*, 2741–2759, DOI: 10.1080/014311699211778.
47. Sims, D.A.; Gamon, J.A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.* **2002**, *81*, 337–354, DOI: 10.1016/S0034-4257(02)00010-X.
48. Sims, D.A.; Gamon, J.A. Estimation of vegetation water content and photosynthetic tissue area from spectral reflectance: A comparison of indices based on liquid water and chlorophyll absorption features. *Remote Sens. Environ.* **2003**, *84*, DOI: 10.1016/S0034-4257(02)00151-7.
49. Pu, R.; Foschi, L.; Gong, P. Spectral feature analysis for assessment of water status and health level in coast live oak (*Quercus agrifolia*) leaves. *Int. J. Remote Sens.* **2004**, *25*, 4267–4286, DOI: 10.1080/01431160210155965.
50. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F. *E1071: Misc Functions of the Department of Statistics (e1071)*; TU Wien: Vienna, Austria, 2012.
51. Max, K. caret: Classification and regression training; R package version 5.17–7; 2013.
52. Liaw, A.; Wiener, M. Classification and regression by RandomForest. *R news* **2002**, *2*, 18–22.
53. Romanski, P. fselector: Selecting attributes; R package version 0.19; 2013.
54. Strobl, C.; Boulesteix, A.-L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional variable importance for random forests. *BMC Bioinf.* **2008**, *9*, 307, DOI: 10.1186/1471-2105-9-307.
55. Xie, Y.; Sha, Z.; Yu, M. Remote sensing imagery in vegetation mapping: A review. *J. Plant Ecol.* **2008**, *1*, 9–23, DOI: 10.1093/jpe/rtm005.
56. Lehmann, E. *Nonparametrics: Statistical Methods Based on Ranks (POD)*, 1st ed., Prentice-Hall: 1975/Revised ed., Springer: Berlin, Germany, 2006.
57. Wang, L.; Sousa, W.P. Distinguishing mangrove species with laboratory measurements of hyperspectral leaf reflectance. *Int. J. Remote Sens.* **2009**, DOI: 10.1080/01431160802474014.
58. Guang, Z.; Maclean, A.L. A comparison of canonical discriminant analysis and principal component analysis for spectral transformation. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 841–847.
59. Breiman, L. Random forests. *Machine Learn.* **2001**, *45*, 5–32, DOI: 10.1023/A:1010933404324.
60. Strobl, C.; Malley, J.; Tutz, G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* **2009**, *14*, 323, DOI: 10.1037/a0016973.
61. Kohavi, R.; John, G.H. Wrappers for feature subset selection. *Artif. Intell.* **1997**, *97*, 273–324, DOI: 10.1016/S0004-3702(97)00043-X.
62. Féret, J.-B.; Asner, G.P. Tree species discrimination in tropical forests using airborne imaging spectroscopy. *IEEE Trans. Geosci. Remote Sens.* **2012**, *51*, DOI: 10.1109/TGRS.2012.2199323.
63. Mas, J.; Flores, J. The application of artificial neural networks to the analysis of remotely sensed data. *Int. J. Remote Sens.* **2008**, *29*, 617–663, DOI: 10.1080/01431160701352154.
64. Bergmeir, C.; Benítez, J. Neural networks in R using the Stuttgart neural network simulator: RSNNs. *J. Statist. Softw.* **2012**, *46*, 1–26.

65. Zell, A.; Mache, N.; Hübner, R.; Mamier, G.; Vogt, M.; Schmalzl, M.; Herrmann, K.-U. SNNS (Stuttgart neural network simulator). In *Neural Network Simulation Environments*; Springer: New York, NY, USA, 1994; pp. 165–186, DOI: 10.1007/978-1-4615-2736-7_9.
66. Hastie, T.; Tibshirani, R. Generalized additive models. *Statistical Sci.* **1986**, *1*, 297–318.
67. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B (Statistical Methodology)* **2005**, *67*, 301–320, DOI: 10.1111/j.1467-9868.2005.00503.x.
68. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Statist. Softw.* **2010**, *33*, 1–22.
69. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Statist. Soc. B (Methodological)* **1996**, *58*, 267–288, DOI: 10.1111/j.1467-9868.2011.00771.x.
70. Bajwa, S.; Bajcsy, P.; Groves, P.; Tian, L. Hyperspectral image data mining for band selection in agricultural applications. *Trans. ASAE* **2004**, *47*, 895–908.
71. Cho, M.; Sobhan, I.; Skidmore, A.; de Leeuw, J. Discriminating species using hyperspectral indices at leaf and canopy scales. In Proceedings of the XXI Congress : Silk Road for Information from Imagery, Beijing, China, 3–11 July 2008; pp. 369–376.
72. Kalacska, M.; Bohman, S.; Sanchez-Azofeifa, G.A.; Castro-Esau, K.; Caelli, T. Hyperspectral discrimination of tropical dry forest lianas and trees: Comparative data reduction approaches at the leaf and canopy levels. *Remote Sens. Environ.* **2007**, *109*, DOI: 10.1016/j.rse.2007.01.012.
73. Civco, D.L. Artificial neural networks for land-cover classification and mapping. *Int. J. Geogr. Inf. Sci.* **1993**, *7*, 173–186, DOI: 10.1080/02693799308901949.
74. Jacquemoud, S.; Ustin, S.; Verdebout, J.; Schmuck, G.; Andreoli, G.; Hosgood, B. Estimating leaf biochemistry using the prospect leaf optical properties model. *Remote Sens. Environ.* **1996**, *56*, 194–202, DOI: 10.1016/0034-4257(95)00238-3.
75. Asner, G.P.; Martin, R.E. Canopy phylogenetic, chemical and spectral assembly in a lowland Amazonian forest. *New Phytol.* **2011**, *189*, 999–1012, DOI: 10.1111/j.1469-8137.2010.03549.x.
76. Zomer, R.; Trabucco, A.; Ustin, S. Building spectral libraries for wetlands land cover classification and hyperspectral remote sensing. *J. Environ. Manage.* **2009**, *90*, 2170–2177, DOI: 10.1016/j.jenvman.2007.06.028.
77. Lin, Y.; Lique, Z. Identification of the spectral characteristics of submerged plant *Vallisneria spiralis*. *Acta Ecol. Sinica* **2006**, *26*, 1005–1010, DOI: 10.1016/S1872-2032(06)60019-X.
78. Hestir, E.L.; Khanna, S.; Andrew, M.E.; Santos, M.J.; Viers, J.H.; Greenberg, J.A.; Rajapakse, S.S.; Ustin, S.L. Identification of invasive vegetation using hyperspectral remote sensing in the California Delta ecosystem. *Remote Sens. Environ.* **2008**, *112*, DOI: 10.1016/j.rse.2008.01.022.
79. Ishii, J.; Washitani, I. Early detection of the invasive alien plant *Solidago altissima* in moist tall grassland using hyperspectral imagery. *Int. J. Remote Sens.* **2013**, *34*, 5926–5936, DOI: 10.1080/01431161.2013.799790.
80. Pengra, B.W.; Johnston, C.A.; Loveland, T.R. Mapping an invasive plant, *Phragmites australis*, in coastal wetlands using the EO-1 Hyperion hyperspectral sensor. *Remote Sens. Environ.* **2007**, *108*, 74–81, DOI: 10.1016/j.rse.2006.11.002.
81. Zedler, J.B.; Kercher, S. Causes and consequences of invasive plants in wetlands: Opportunities, opportunists, and outcomes. *Critical Rev. Plant Sci.* **2004**, DOI: 10.1080/07352680490514673.

82. Somers, B.; Asner, G.P. Multi-temporal hyperspectral mixture analysis and feature selection for invasive species mapping in rainforests. *Remote Sens. Environ.* **2013**, *136*, 14–27, DOI: 10.1016/j.rse.2013.04.006.
83. Somers, B.; Asner, G.P. Hyperspectral time series analysis of native and invasive species in Hawaiian rainforests. *Remote Sens.* **2012**, *4*, 2510–2529, DOI: 10.3390/rs4092510.
84. Adam, E.; Mutanga, O.; Rugege, D. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review. *Wetlands Ecology and Management* **2010**, *18*, 281–296, DOI 10.1007/s11273-009-9169-z.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).