Remote Sensing of Environment 216 (2018) 385–400

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

# Characterization of indicator tree species in neotropical environments and implications for geological mapping

Cibele Hummel do Amaral[a,*], Teodoro Isnard Ribeiro de Almeida[b], Carlos Roberto de Souza Filho[c], Dar A. Roberts[d], Stephen James Fraser[e], Marcos Nopper Alves[f], Moreno Botelho[g]

[a] Departamento de Engenharia Florestal, Universidade Federal de Viçosa, Avenida Purdue s/n, Viçosa, MG 36570-900, Brazil
[b] Instituto de Geociências, Universidade de São Paulo, Rua do Lago 562, São Paulo, SP 05508-080, Brazil
[c] Instituto de Geociências, Universidade de Campinas, PO Box 6152, Campinas, SP 13083-870, Brazil
[d] Department of Geography, University of California Santa Barbara, 1832 Ellison Hall, Santa Barbara, CA 93106, United States
[e] Mineral Resources Flagship, Commonwealth Scientific and Industrial Research Organisation, PO Box 883, Kenmore, QLD 4069, Australia
[f] Centro Pluridisciplinar de Pesquisas Químicas, Biológicas e Agrícolas, Universidade de Campinas, Avenida Alexandre Cazelatto 999, Paulínia, SP 13148-218, Brazil
[g] MB Soluções Florestais, Rua Vereador Gilberto Valério Pinheiro 134/102, Viçosa, MG 36570-000, Brazil

ARTICLE INFO

*Keywords:*
Geobotany
Cerrado biome
Sedimentary formations
Multivariate analysis
Spectroscopy
Spectral Mixture Analysis

ABSTRACT

Geobotanical remote sensing (GbRS) in the strict sense is an indirect approach to obtain geological information in heavily vegetated areas for mineral prospecting and geological mapping. Using ultra- and hyperspectral technologies, the goals of this research comprise the definition and mapping of Neotropical tree species that are associated with geological facies (here called geo-environments) as well as their spectral discrimination at leaf and crown scales. This work also aims to investigate the possible relationship between leaf and crown spectral and chemical properties. The study was developed at the Mogi-Guaçu Ecological Station, in the Cerrado domain, southeastern Brazil. Data from 70 sample units, such as sediment texture and species from inventories, were first analyzed through vectorial quantization using Self-Organizing Maps (SOM). Principal Component Analysis and Spearman's ranked correlation coefficients were used to define geo-environments and target-species, respectively. Biochemical and visible to shortwave infrared (VSWIR) point spectral data (350–2500 nm) were collected from the leaves of the target-species, during both rainy and dry seasons. Spectral data from target-species crowns were obtained from hyperspectral images (530–2.532 nm, ProSpecTIR-VS sensor) with 1 m spatial resolution, and acquired in the beginning of the dry season. These spectra were classified using Multiple Endmember Spectral Mixture Analysis (MESMA) with two endmembers (EMs). Based on the MESMA results with two EMs, the best dataset per target-species was chosen for pixel-based image unmixing with three EMs (target-species, other vegetation types and shade). From 121 species sampled in the field, two proved to be associated with floodplains (Alluvial Deposits sequence), two with hills and plateaus of the Aquidauna Formation (Carboniferous sedimentary rocks, Paraná Basin), and two more with a specific facies of the Aquidauna Formation that has a distinctive presence of coarse and very coarse sand. Five target-species were well discriminated at the leaf scale, reaching 90.0% and 85.0% of global accuracies in the rainy season and in the dry season, respectively. Accurate spectral discrimination appears to be linked to the considerable biochemical variability of their leaves in both seasons. Three species were discriminated at the crown scale, with 70.6% of global accuracy. When eight other landscape scale vegetation classes were included in the analyses, only *Qualea grandiflora* Mart. produced a satisfactory accuracy (61.1% and 100% of producer and user's accuracies, respectively). The spatial distribution of its fraction in the unmixed image, particularly, matches with the geological facies to which it was associated in the field. Ecological requirements for successfully mapping indicator species include broad and random distribution of the target-species' population, and singular physiological, phenological (and spectral) behavior at the imagery acquisition date. Our study shows that, even in tropical conditions, it is possible to use plant species mapping to support geological delineation, where rock exposures are typically rare.

* Corresponding author,
*E-mail address:* chamaral@ufv.br (C.H.d. Amaral).

https://doi.org/10.1016/j.rse.2018.07.009
Received 26 January 2018; Received in revised form 5 July 2018; Accepted 8 July 2018
Available online 17 July 2018

## 1. Introduction

Geobotany is an old science, commonly understood as a technique for mineral prospecting in vegetated terrains (e.g., Cannon, 1960, 1971; Brooks, 1972, 1983; Prasad, 1987). In a broader sense, it comprises the study of the physical and biological processes that are of interest to Earth Science (Ustin et al., 1999). According to those authors, Geobotany has the potential of providing a better understanding of phenology, abundance and the distribution of plants through holistic ecophysiological models, connecting plant behavior to edaphic conditions and geology.

In western culture, recognition of an association between a plant species and a specific rock type dates back to the 4th century B·C, credited to Theophrastus (Kruckeberg, 2002). However, there are practical difficulties related to this science by requiring both geological and botanical knowledge. Since the advent of multispectral remote sensing, Geobotany has gained increasing importance (Lyon and Lee, 1970; Ustin et al., 1999) in its strict and broad senses (e.g., Carranza and Hale, 2002; Higgins et al., 2011; Lammoglia and de Souza Filho, 2013; Sanches et al., 2013; Sirén et al., 2013; Hede et al., 2015; Wang et al., 2018).

The first Geobotanical Remote Sensing (GbRS) studies focused essentially on mineral deposits, as shown by various NASA Technical Reports (e.g., Arden Jr and Westra, 1977). Within the Cerrado biome (Brazilian tropical savanna), GbRS studies seem to be also focused on the characterization of particular mineralizing processes (e.g. Almeida-Filho and Castelo Branco, 1992; Almeida-Filho et al., 1996; Almeida-Filho and Vitorello, 1997), and hydrocarbon microseepage (e.g., Souza Filho et al., 2008). These studies highlighted geobotanical anomalies due to chemical stresses. However, the application of the tool in regional geology over common rocks is still rare. Evidence of association between Neotropical plant communities' distributions and common local geology has been found using multispectral remote sensing in western Amazon for example (e.g., Higgins et al., 2011, 2012, 2014a).

As foreseen by Sabins (1999), the availability of hyperspectral sensors has also encouraged new GbRS investigations. Madritch et al. (2014) showed that the canopy visible to shortwave infrared (VSWIR) hyperspectral reflectance from genotypes of trembling aspen (*Populus tremuloides*) is directly correlated with soil traits and belowground processes in two ecoregions: the Great Lakes region and Western region of the United States. In Neotropical environments, Asner et al. (2005) found associations between canopy traits and the age of underlying substrate using hyperspectral vegetation indices from the Hyperion sensor. Vitousek et al. (2009) correlated variations in canopy nitrogen content and vegetation height to substrate age using airborne VSWIR hyperspectral and Light Detection And Ranging (LIDAR) data. Higgins et al. (2014b) observed clear relationships between vegetation species composition, VSWIR spectral response and LIDAR structural data on two different underlying sedimentary formations in a Central Panama broadleaf forest. However, the information that can be obtained through hyperspectral remote sensing goes beyond what has been shown so far in GbSR approaches at the Neotropical zone, which hosts hyper-diverse vegetation types.

Remote discrimination and mapping of plant species in hyper-diverse communities is not an easy task. The high inter-specific and intra-specific spectral variability of tropical formations is still poorly understood by the scientific community (Asner, 2008; Asner and Martin, 2008). The current consensus is that the success of species mapping in tropical environments is a result of the species-specific structural, biochemical, physiological and phenological characteristics, making it a distinct “optical type” in face of other neighbor species (Asner and Martin, 2008; Gamon, 2008; Ustin and Gamon, 2010). Spectral discrimination of Neotropical species at the individual tree crown level, which is possible only with sensors of fine spatial resolution, was achieved in 2005 by Clark et al. (2005) in Costa Rica. Later, Kalacska et al. (2007), Féret and Asner (2013), Somers and Asner (2014), Baldeck et al. (2015) and Ferreira et al. (2016) reported similar successes in tree species mapping.

However, understanding the characteristics related to a crown's reflectance and its spectral classification is a challenge that starts at the leaf level (Castro-Esau and Kalacska, 2008). Evidence is emerging that at the leaf level hyperspectral remote sensing detects not only species but functional, structural and phylogenetical properties, which may vary among species and functional groups (e.g., Clark et al., 2005; Gamon et al., 2005; Asner, 2008; Asner and Martin, 2009; Asner and Martin, 2016; McManus et al., 2016). Successful spectral discrimination of Neotropical species at leaf scales has been shown in the literature since the 2000's (e.g., Cochrane, 2000; Castro-Esau et al., 2004; Clark et al., 2005; Castro-Esau et al., 2006; Rivard et al., 2008; Féret and Asner, 2011; Hesketh and Sánchez-Azofeifa, 2012; Ferreira et al., 2013; Prospere et al., 2014).

Focusing on plant functional groups in Neotropical forests, recent advances comprise species discrimination at the leaf level from various habitats (e.g., Castro-Esau et al., 2004; Kalacska et al., 2007; Sánchez-Azofeifa et al., 2009) and successional stages (e.g., Alvarez-Añorve et al., 2012). Hesketh and Sánchez-Azofeifa (2012) studied leaf spectral variation within tree and liana species in the rainy and the dry seasons, in wet and dry forests of Central America, observing different seasonal responses and good discrimination in both seasons. Due to the high intensity of the species-specific spectral variations between seasons, the authors highlighted the necessity of leaf chemical analyses to strengthen the observed spectral variations.

A reasonable amount of studies on Neotropical species' spectral discrimination, both at leaf and crown scales, as well as on emergent trees mapping, can be found in the literature, as shown above. Nevertheless, studies focusing on geological indicator tree species are generally lacking.

Thus, the objective of this investigation is to characterize and map indicator tree species to support geological facies delineation in an ecological station in southeastern Brazil. Specifically, we aim (i) to discriminate, chemically and spectrally, the tree species at leaf scale in both rainy and dry seasons; (ii) to distinguish these species spectrally at the crown scale in the dry season; and (iii) to map the sub-pixel fraction of the species using high-spatial-resolution VSWIR hyperspectral images. In addition, we discuss the relationships between target-species leaf traits and leaf and crown spectral discrimination, the challenges on discriminating and mapping indicator tree species previously defined in the field, as well as the application of the method to geobotanical studies in hyper-biodiverse landscapes, as found in the Neotropics.

## 2. Materials and methods

### 2.1. Study area

The study area is the Mogi-Guaçu Ecological Station (MGES), located in the Mogi-Guaçu municipality, State of São Paulo, southeastern Brazil. The MGES is approximately 1230 ha (Fig. 1) consisting of large hills and a floodplain, with elevations between 560 and 700 m. Most of the soils are hydromorphic floodplain soils. The MGES includes approximately 17 km of the Mogi-Guaçu River. We chose this area because common sedimentary formations without geochemical anomalies compose the local geology. Thus, we could test our GbRS methodology, designed for local (and regional) geologic mapping. The MGES geology is composed of the Aquidauana Formation and alluvial deposits from the Mogi-Guaçu River (CPRM, 2006). The Aquidauana Formation (Carboniferous sedimentary rocks, Paraná Basin) is a fluviolacustrine stratum with reddish sandstones and siltstones and, in addition, conglomerates, diamictites, rhythmites and shales. The diamictites, which are also red in color, contain dispersed clasts, some with striations of glacial origin, indicating that they represent tillites. Despite this lithological diversity, Fúlfaro and Bjomberg (1993) observed that the soils developed on this formation are commonly sandy. The alluvial deposits

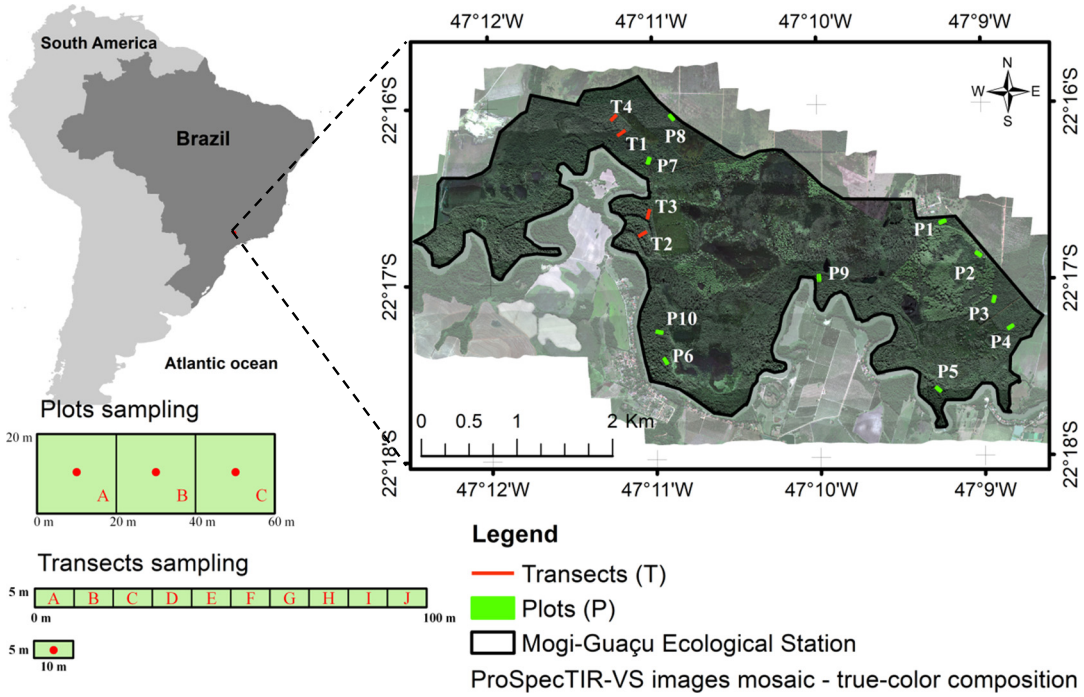


**Fig. 1.** Location of the study area and spatial distribution of the plots and transects whitin it. The ten plots and four transect subsamples composing the 70 studied sample units are illustrated on the bottom left side of the image.

(Quaternary sediments) – whose source area is that sedimentary geological unit – consist of sands, gravels, silts, clays and, locally, peat.

The MGES is in a southeastern marginal area of the Cerrado biome (tropical Brazilian savanna), exhibiting mosaics of vegetation characteristic of the Cerrado and Atlantic Forest biomes (Joly and Bicudo, 1999). The semi-deciduous seasonal forests, from the Atlantic Forest biome, occur as extra-Cerrado physiognomies due to their association, even remotely, with watercourses (Ribeiro and Walter, 2008). In the first floristic study performed in the area, Eiten (1963) noted that the occurrences of the woodland savanna (Cerradões) and gallery forests are associated with the highest and lowest portions of the relief, respectively, dominated by substrates composed of rocks from the Aquidauana Formation and Quaternary sediments.

### 2.2. Geological facies and definition of indicator species

First, an extensive survey of the vegetation strata (physiognomies) was performed in the field using the physiognomic classification of the Cerrado provided by Ribeiro and Walter (2008). This survey was aided by previous analysis and photo-interpretation of images with 1 m spatial resolution. A stratification of the woody typologies and a selection of sub-environments within each stratum were previously performed for the physiognomic sampling. Only forest physiognomies were considered in this study. Ten plots of 20 × 60 m and four transects of 5 × 100 m, perpendicular to the axis of oxbow lakes of the Mogi-Guaçu river, were defined (Fig. 1). The plots were subdivided into three contiguous subplots of 20 × 20 m, and transects were subdivided into 10 contiguous sampling units of 5 × 10 m. This process allowed for the recognition and sampling of subtle variations in the physical environment, which can influence the distribution of species – essential information for the geobotanical analysis.

All individuals with a DBH (diameter at breast height; 1.3 m above ground) ≥ 5 cm were sampled. We recorded the DBH, the estimated height (H), the geographical coordinates and the taxonomic identification (Angiosperm Phylogeny Group - APG classification IV) for all sampled woody individuals.

Sediments samples (weighing around 300 g) were collected at 2 m depth in the center of each sampling unit (n = 70) with the aid of an auger. Grain size distributions were determined for the sediment samples and classified into seven classes from lower than 0.002 mm to 2 mm: clay, silt, very fine sand, fine sand, medium sand, coarse sand and very coarse sand.

The Self-Organizing Map method (SOM; Kohonen, 1998) was applied to the field data using the SiroSOM software (CSIRO Mineral Resources Flagship, Pullenvale, QLD, AU). Initially, the distributions of the variables (grain size and presence/absence of species) were normalized using a logarithmic function. Seed-vectors were randomly distributed in the n-dimensional space defined by the input data, and then trained through vector quantization and measurements of vector similarity to represent the structure and patterns of the sampling units. This training was performed using a two-step iterative process: competition and cooperation. These steps were applied various times for each sampling unit until the winning seed-vector represented the input data in the best manner. The 2D output map has a key characteristic of preservation of the relative topological relationships between the node vectors. To assess the “effectiveness” of a SOM, two measures are calculated: the average quantization error (qe), which is the average distance between each data vector and its best matching units (BMU, nodes in the 2D maps); and the topographic error (te), the proportion of all vectors for which the first and second BMUs of a given input sample vector are not adjacent units on the map. These measures correspond to the map resolution (qe) and topology preservation (te) respectively.

Principal component analysis and Spearman's ranked correlation (ρ) were performed on the SOM BMU results, in order to analyze the relationships between samples and variables, respectively. ρ is considered significant when $\geq |0.04|$. Significant positive and negative values indicate the occurrence of directly and inversely proportional relationships between the variables, respectively. In order to validate those correlations, Student's *t*-test probability values were generated and analyzed. t-test values closer to zero indicate more symmetric distributions for a population. For the definition of the geo-environments based on differing textural compositions, the SOM's PCA used the grain size distributions in sediments from the 70 sampling units. The selection of the species sampled exclusively in the defined geo-environments and

their relationships with the textural sediment categories were achieved through the SOM and ρ analyses, which were trained using grain size distributions in sediments and data on the presence/absence of species in the 70 sampling units.

### 2.3. Target-species discrimination

#### 2.3.1. Leaf chemical and spectral data

Three individuals from each target species were selected for study, and branches with sun leaves were collected at the upper canopy. The spectral measurement, weighing and storage of the leaves occurred less than 3 h after their collection in order to conserve their biochemical characteristics. Five leaves were collected from the top to the bases of the branches to include leaves of various ages, sizes and colors in the analyses. Spectral measurements, fresh leaf weighing, and cold storage for further biochemical analyses were done at the MGES base. These procedures were carried out in the 2012 rainy and dry seasons.

A portion of the leaves was stored in paper bags and weighed in the field. These bags were transported to the *Pluridisciplinary Center for Chemical, Biological and Agricultural Research* (CPQBA) of the *University of Campinas* (UNICAMP), where they were oven dried (40 °C–72 h) and weighed again for water content measurement. Another portion was stored in plastic bags and transported in an ice chest (0 °C to −10 °C) to the *Laboratory of Plant Physiology* at UNICAMP. By using extraction methods based on Holm (1954) and Updegraff (1969), concentrations of chlorophylls a and b (μg/mg), carotenoids + xanthophylls (μg/mg), anthocyanin (Δ/mg); cellulose (mg/g), hemicellulose (mg/g) and insoluble lignin (%) were obtained.

Five leaves from each sampled individual had their reflectance spectra collected with a FieldSpec® 3 Hi-Res (Analytical Spectral Devices, Boulder, CO, US) radiometer. This spectrometer samples between 0.35 and 2.5 μm, with full width half maximum (FWHM) of 3 nm (0.35–0.7 μm), 8.5 nm (0.7–1.4 μm) and 6.5 nm (~2.1 μm), and original sampling interval of 1.4 (0.35–1.0 μm) and 2.0 nm (1.0–2.5 μm) resampled to 1.0 nm in software (Analytical Spectral Devices, Boulder, CO, US). The spectra were collected from the adaxial (upper) surface of the leaves, excluding the central veins, by using a plant probe and a leaf clip. Both instruments allow the collection of nadir measurements with identical illumination conditions, and no atmospheric contamination. A Spectralon panel (Labsphere Inc., Durham, NH, US) was used as a reflectance standard. The absolute reflectances were retrieved and the offsets between sensors were corrected through USGS-PRISM software routines (Kokaly, 2011). Only bands from 0.4 to 2.45 μm were retained for processing due to the low signal-noise ratio of the bands located in the beginning and in the end of the spectra (Fig. 2).

The VSWIR (0.4–2.45 μm) data were subdivided into spectral ranges: visible (VIS; 0.4–0.7 μm), near-infrared (NIR; 0.7–1.3 μm) and shortwave infrared (SWIR; 1.3–2.45 μm). These four sets of spectra had their continuum removed (CR), in order to normalize for variation in brightness and highlight the absorption features of each spectral interval (Clark and Roush, 1984). Sixteen spectral libraries were generated for processing, taking into consideration two seasons, four spectral subsets, and two different spectrum types: reflectance and normalized (CR). Those spectral libraries are called here as: VSWIR, VSWIR(CR), VIS, VIS(CR), NIR, NIR(CR), SWIR e SWIR(CR), from rainy (RS) and dry (DS) seasons.

#### 2.3.2. Crown spectral data

The ProSpecTIR – VS hyperspectral data (SpecTIR – Reno, NV, US/FotoTerra – São Paulo, SP, Brazil) were obtained on June 8th, 2010 by a two-sensor aerial system. The first sensor operates between 400 and 970 nm with spectral sampling interval of ~2.9 nm, and the second between 970 and 2450 nm (~8.5 nm). Seventeen east-west image strips were acquired at a nadir viewing, at an average height above the ground of ~1350 m. The strips are ~320 m wide with a spatial resolution of 1 m. Pre-processing steps included radiometric calibration, reflectance retrieval, spectral enhancement, georeferencing, mosaicking, bad bands exclusion and non-vegetation features masking were performed on the hyperspectral dataset, as described in Amaral et al. (2015).

The pre-processed VSWIR (216 channels; 530 nm-2352 nm) dataset had its CR and was subdivided into VIS-NIR1 (82 channels; 530–919 nm) and NIR2-SWIR (134 channels; 1141–2352 nm). Thus, six different datasets were tested in spectral discrimination at the crown scale: VSWIR and VSWIR (CR); VIS-NIR1 and VIS-NIR1(CR); NIR2-SWIR and NIR2-SWIR(CR).

Geographic coordinates were collected from target-species individuals using a differential GPS. A differential system was required due to the nature of the work: at the level of the individual tree crown – ITC (Clark et al., 2005) in dense hyper-diverse canopies. The distance between receivers (base and portable) was always lower than 4.5 km. The antenna of the portable receiver was coupled to the extensible rod, with a data transmission cable of 30 m; it was raised to the canopy with the aid of a tree-climber. The post-processed coordinates present precision higher than 1 m.

In addition to the target-species classes (Table 1 and Fig. 3), eight other classes of representative vegetation types on the landscape were used for the crown scale spectral unmixing, as well as for the pixel-based fraction mapping (Table 1). The reference data of the Cerradão and Riparian Forest species classes were collected on crowns from non-target species with the aid of plots and transects' shapefiles, where the targets do not occur. The other classes had their coordinates collected by a navigation GPS. Based on geographical location, image interpretation and spectral analyses, polygons of up to 9 $m^2$ were obtained from the crowns of interest making no distinction between sunlit and shade pixels. The mean spectra of each polygon were extracted from the six different images (VSWIR and VSWIR(CR); VIS-NIR1 and VIS-NIR1(CR); NIR2-SWIR and NIR2-SWIR(CR)).

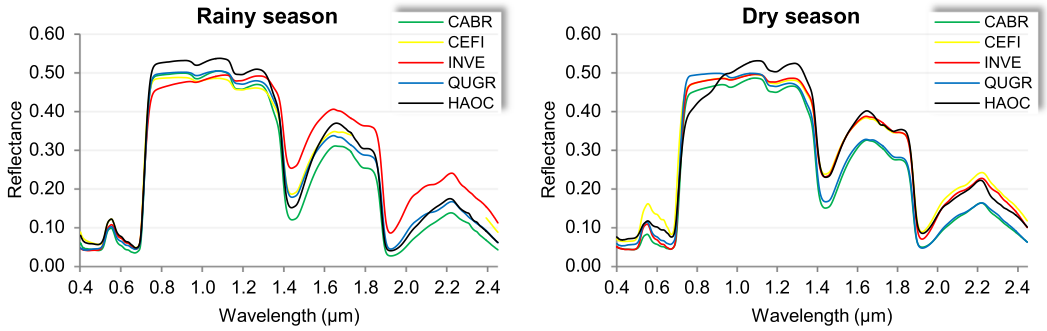


**Fig. 2.** Average leaf spectra from the target-species collected in both the rainy and dry season in the Mogi-Guçu Ecological Station, southeastern Brazil. CABR = *Calophyllum brasiliense* Cambess.; CEFI = *Cedrela fissilis* Vell.; INVE = *Inga vera* Willd. subsp. *affinis* (DC.) T.D. Penn.; QUGR = *Qualea grandiflora* Mart.; HAOC = *Handroanthus ochraceus* (Cham.) Mattos.

**Table 1**
Classes, initials and number of spectra extracted from each of the six images: reflectance and continuum-removed VSWIR (530–2352 nm), VIS-NIR1 (530–919 nm) and NIR2-SWIR (1141–2352 nm) datasets, for performing the different proposed analyses.

| Spectral analysis # 2 + pixel-based fraction mapping | | | |
|---|---|---|---|
| | Class | Initial | Number of spectra |
| Spectral analysis #1 | *Cedrela fissilis* | CEFI | 22 |
| | *Inga vera* subsp. *affinis* | INVE | 31 |
| | *Qualea grandiflora* | QUGR | 28 |
| | Global libraries | | 81 |
| | Cerradão species | CERspp | 18 |
| | *Citrus spp.* | CITspp | 20 |
| | *Dendrocalamus sp.* | DENDsp | 20 |
| | *Eucalyptus spp.* | EUCspp | 20 |
| | Riparian forest species | RIFOspp | 28 |
| | *Pinus spp.* | PINspp | 20 |
| | *Poaceae spp.* | POAspp | 20 |
| | *Sacharum spp.* | SACspp | 20 |
| | Global libraries | | 328 |

### 2.3.3. Endmember selection

Endmembers (EMs) are spectra that better represent the analyzed classes and are considered pure (Adams et al., 1993). For EM selection, first, leaf spectral libraries were randomly divided into training and validation libraries by considering a proportional limit of ~50% of spectra for each spectral library (Roth et al., 2012). A threshold between – 0.05 and 1.05 was defined for EMs fraction (as in Dennison and Roberts, 2003; Roth et al., 2012) and various Root Mean Square Error (RMSE) thresholds were empirically tested, in order to achieve the highest accuracy of spectral classification, on reflectance and CR data. The first threshold tests were based on the literature (Dennison and Roberts, 2003; Youngentob et al., 2011; Roth et al., 2012), being defined for the target-species leaf spectra classification: RMSE ≤ 3.5% for reflectance data and RMSE ≤ 7.5% for CR data.

Global crown spectral libraries were also subdivided into training and validation libraries. An absolute limit of ten spectra per class was defined for EM selection, in order to have a balanced training and to preserve a reasonable amount of spectra for validation (Roth et al., 2012). EM fraction threshold and maximum allowed RMSE were defined for the datasets as follow: EM fraction from −0.05 to 1.05 and RMSE ≤ 2.5% for reflectance spectra, and from −0.01 to 1.01 and RMSE ≤ 6.5% for CR spectra.

The EM selection metrics Endmember Average RMSE (EAR; Dennison and Roberts, 2003), Minimum Average Spectral Angle (MASA; Dennison et al., 2004) and Count-Based Index (CoBIndex; Roberts et al., 2003 and Clark, 2005) – jointly called EMC (Roberts et al., 2007) – and the algorithm Iterative Endmember Selection (IES; Schaaf et al., 2011) were tested on the training libraries. The class EMs selected through EMC where those ones with the lowest EAR, the lowest MASA, and the highest CoBIndex. Thus, one to three EM were selected by EMC for each class.

The IES endmember selection was performed by using an Interactive Data Language – IDL rotine (Schaaf et al., 2011; Roth et al., 2012). The EM set was automaticaly selected by adding and subtracting spectra from the EM set until the highest Kappa coefficient, during classifications with two EM, was achieved. When the procedure did not select at least one EM for a certain class, the EM selected by EMC for this class were inserted into the IES EM library, and it is called here IES* EM library.

### 2.3.4. Multiple endmember spectral mixture analysis

Spectral Mixture Analysis (SMA) models spectra from a library as a linear combination of pure spectra (EM), determining their relative proportion within a mixed spectrum (Adams et al., 1993). While traditional SMA techniques allow only one EM per class, Multiple Endmember Spectral Mixture Analysis (MESMA) allows several types and numbers of EMs per class, generating linear models from different sets of EMs (Roberts et al., 1998). MESMA can be used as a classifier where viable EM selections are assessed by whether they meet fraction constraints (typically −0.05 to 1.05) and fit constraints (RMSE ≤ 2.5%, for example). The technique is suitable for species-specific spectra classification, by surmounting within species (and crowns) spectral variability (Roth et al., 2012) as well as brightness variation between flight lines when working with aerial imaging spectroscopy (Amaral et al., 2015).

The validation libraries were classified by the EM spectral libraries, by using MESMA with two EM IDL codes (Roberts et al., 2012) and the thresholds defined for each data type (leaf and crown datasets). The technique was performed on the target-species leaf datasets, on the target-species crown datasets (Crown Spectral Analysis # 1), and on the eleven landscape vegetation – including target-species – crown datasets (Crown Spectral Analysis # 2).

Accuracies of the MESMA spectral modeling were analyzed through Overall accuracy and Kappa coefficient metrics, and class accuracies through Producer's and User's accuracies. Based on the assumption that a commission error (expressed by the user's accuracy) of a class is equally an omission error (producer's accuracy) of another class (Congalton, 1991; Congalton and Green, 1999), pairwise analyses of the spectral misclassification among species classes were performed.

### 2.3.5. Statistical analysis

In order to discriminate the target species chemically and to aid the analyses on spectral discrimination at leaf and crown scales, the significance of the variability in leaf traits was analyzed using one way

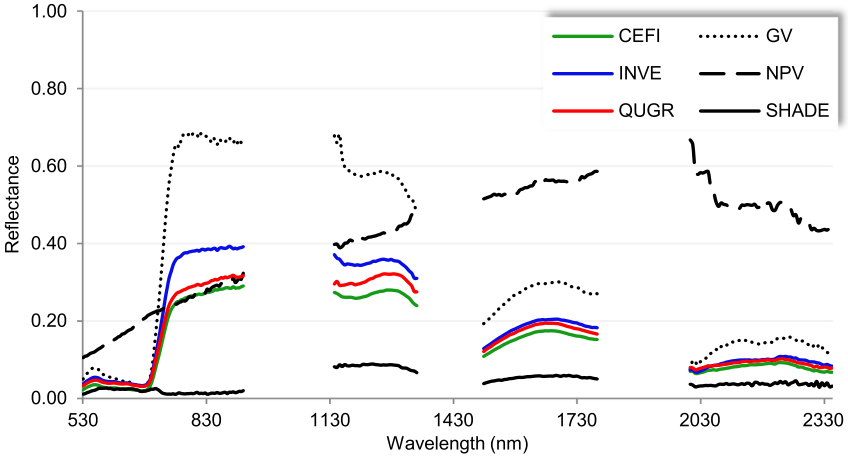


**Fig. 3.** Average crown spectra from target species, plus green vegetation (GV), non-photosynthetic vegetation (NPV) and shade reference endmembers extracted from the ProSpecTIR-VS image, in the Mogi-Guçu Ecological Station, southeastern Brazil. CEFI = *Cedrela fissilis* Vell.; INVE = *Inga vera* Willd. subsp. *affinis* (DC.) T.D. Penn.; QUGR = *Qualea grandiflora* Mart.

ANOVA (F test). We tested the null hypothesis that there were no significant differences in leaf traits among the studied species within a season and between seasons. We also tested whether the subdivision of the leaf and crown spectra datasets into spectral ranges (and seasons for leaf spectra), different data types (reflectance and CR spectra) and endmember selection techniques (EMC and IES) influenced the spectral misclassification among species and as a whole.

A Tukey test was performed on all datasets with statistically significant variations at 5% of significance in one way ANOVA, in order to compare the contrast among means of two different treatments. In this test, the null hypothesis states that the mean value of the treatment *i* is equal to the mean value of the treatment *i'*, with at least 5% of significance. The studentized range (q) distribution for a certain level of significance, the number of treatments, the degree of freedom of the residue, and the mean square of the residue were used for the Minimum Significant Difference (MSD) calculation.

### 2.4. Target-species fraction mapping

The pixel-based spectral modeling was based on the best results (the highest and most balanced user and producer's accuracies for each target species) obtained for each target-species class in the MESMA with two endmember, using the eleven vegetation classes from the landscape (Crown Spectral Analysis # 2). Each pixel was modeled by three endmembers (EMs): a target-species EM, other vegetation types EM, and photometric shade, using ViperTools (Roberts et al., 2007). The EMs selected for pixel-based modeling were those that correctly classified the spectra of their own classes in the previous step (targe-species discrimination at crown scale). Constraints of fraction threshold and maximum allowed RMSE, previously defined, were kept for image unmixing.

The resulting unmixed images were shade normalized (Smith et al., 1990; Dennison and Roberts, 2003) and evaluated at both crown (n = 247) and pixel (n = 1036) scales. Regions of interest defined for extracting spectra from the images and composing the libraries of the previous MESMA with two EM approach were used for that. When more than one dataset or EM library provided high accuracy results for a certain target-species, the best image and EM for its fraction mapping was defined, at this moment, by its highest sub-pixel fraction abundance within its individual's crowns. In order to optimize the target-species fraction mapping, those unmixed images were classified into two classes, by using as threshold the maximum sub-pixel abundance of a certain target-species within a crown of another vegetation type.

## 3. Results

### 3.1. Geo-environments of the study area and their indicator tree species

Sediments from all 70 sampling units were analyzed (and clustered) to define the geo-environments present in the study site. An average quantization error of 0.644 and a topographic error of 0.0 were achieved in the Self-Organizing Map, by setting a rough training with 20 interactions and a fine training with 400 iteractions. The component plots, which indicate the contributions of each variable to the defined nodes in the 2D space, revealed the spatial associations between the textural classes in the 70 sampling units. It is possible to see dissimilarities between nodes with high silt content and nodes with high medium sand content, as well as between those nodes with high very fine sand content and high coarse and very coarse sand contents (Fig. 4). The U-matrix indicated similarities between nodes represented by the Carboniferous units (at the bottom) and between Quaternary sample units (at the top), as clusters in lower temperature colors (blues-greens). These two geological sequences are clearly segregated by dissimilar nodes, in higher temperature colors (reds-oranges), in the U-matrix.

By applying PCA to the SOM component plots, the distribution of the scores in the 2D space (PC1 and PC2) shows that there are four well-defined geological facies, here called geo-environments (Fig. 5). Two of these environments are associated with Quaternary sediments – Alluvial Deposits (I and II), and the other two are associated with the Carboniferous rocks – Aquidauana Formation (III and IV), as follows: floodplain (including cut-off channels deposits), with negative scores in PC1 and positive scores in PC2 (I); (old) natural levee deposits and point bar deposits, with negative scores in PC1 and in PC2 (II); hills and plateaus, with positive scores in PC1 and PC2 (III); and hill edges (including (old) eroded natural levees), with positive scores in PC1 and negative scores in PC2 (IV). The scores of PC1, therefore, are negatively associated with the Quaternary sedimentary deposits and positively with the soils formed on the Aquidauana Formation.

These four main geo-environments are described below, according to their geological soil data:

Geo-environment I. This zone features hydromorphic soils, poorly drained with variable base saturation, consistent with different flooding levels. The sediments are mainly pelitic (clay and silt) and possess an insignificant to non-existent sand fraction. This geo-environment constitutes the ordinary floodplain of the meandering river, including the oxbow lakes that make up the lacustrine to fluvial systems during floods. It occurs up to an elevation of 570 m.

Geo-environment II. This geo-environment represents the deposits from old levees, point bars and edges of the ordinary floodplain, where the soils are better drained and the sediments consist mainly of silt, very fine sand and fine sand fractions. Its boundaries characterize the edges of the Quaternary sequence. It occurs up to an elevation of 573 m (in the east) and 570 m (in the west).

Geo-environment III. This zone features deep, well-drained, mostly dystrophic soils, with a texture dominated by fine, medium and coarse sand fractions. This geo-environment is distinguished from the others by the presence of fine and medium sand fractions and the low (to non-existent) presence of silt, very fine and very coarse sand fractions in its sediments. It occurs in the hills and plateaus on the Aquidauana Formation and in the interfluvial regions. According to the normal level of the riverbed (which runs from east to west in the area), this geo-environment can occur above elevations of 573 m (in the east) and 570 m (in the west).

Geo-environment IV. These soils are shallower compared to the other geo-environments, are well drained and, in addition to the clay fraction, feature distinct medium, coarse and very coarse sand fractions. It is distinguished from the other geo-environments by the significant presence of coarse and very coarse sand fractions and the insignificant to non-existent presence of the very fine sand and silt fractions. It occurs in the steeper sloped regions on the edges of the Aquidauana Formation, approximately between the elevations of 575 and 585 m, and includes the convex banks (and old banks) eroded via undermining by the Mogi-Guaçu river.

Fig. 6 summarizes schematic cartography of the relief and drainages (Fig. 6-A), the stratigraphy (Fig. 6-B), the vegetation physiognomies (Fig. 6-C), and the geo-environments distribution (Fig. 6-D). In this last figure, Geo-environment I is formed where the water flow rate is lowest, locally producing lacustrine environments and depositing sediments rich in clay and silt. Geo-environment II is characterized by point bars facies and deposits of levees and old levees where the water flow rate is higher and promotes the deposition of silt, very fine sand and fine sand fractions (i.e., coarser fractions than those observed in Geo-environment I). Due to the map spatial resolution, the area defined as "Geo-environment I and II" comprises mixed aspects of these environments.

Although four geo-environments (geological facies) were well defined, only three were observed having species exclusively associated with them. From the 114 species (1348 individuals) sampled across the ten plots and 52 species (193 individuals) sampled in the four transects, we found six species that had sampled individuals restricted to a certain geo-environment, as follow:

Geo-environment I-a) *Calophyllum brasiliense* Cambess.: this species

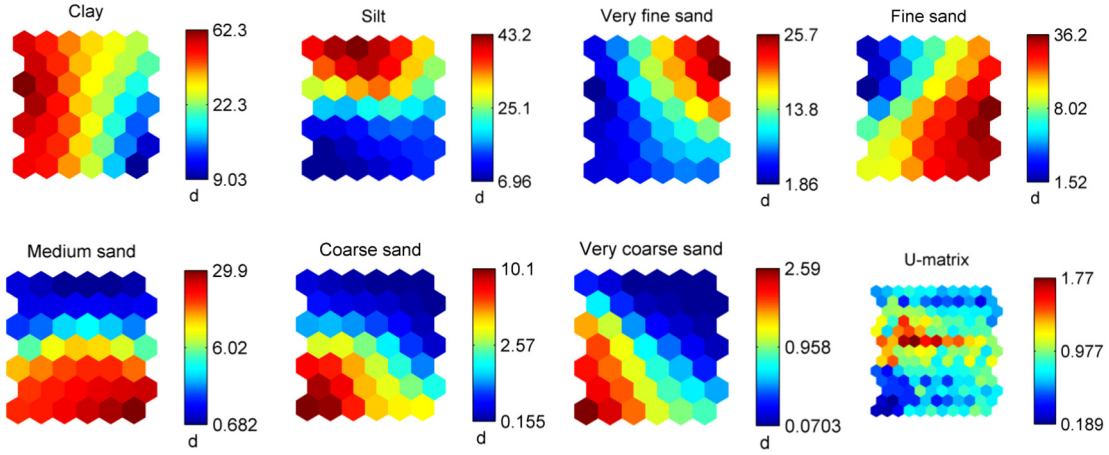


**Fig. 4.** Component plots of the contributions of the grain size fractions to the formation of the nodes in the Self-Organizing Map and U-matrix (unified distance matrix) representation. These results summarize the structure of the data in the 70 sampling units in 2D representations. The color-temperature scale for the seven components (variables) is scaled across the range of input values from low (blue) to high values (red) in percentage. The U-Matrix map shows the distance between neurons in the input space and is also coded using a color-temperature scale: lower temperature colors (blues-greens) indicate similarity between adjacent nodes (as clusters), and higher temperature colors (reds-oranges) indicate dissimilarity (clusters separators). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

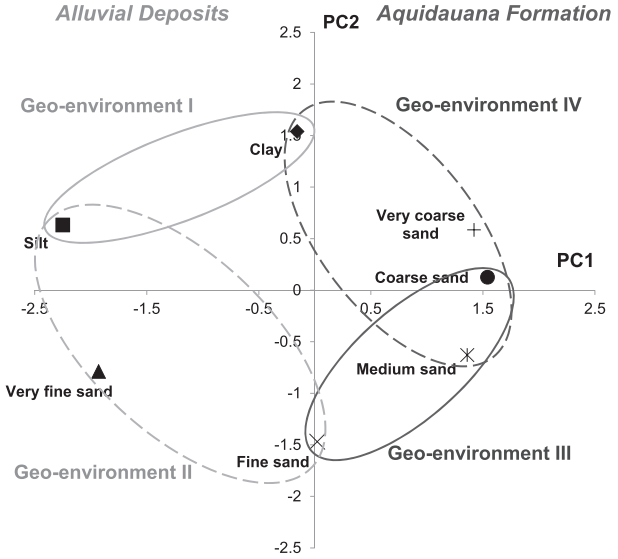


**Fig. 5.** Distribution of the scores of the grain size fractions of sediments among the principal components (PC1 and PC2), and the main geo-environments of the Mogi-Guaçu Ecological Station, Brazil.

was observed only in the floodable cut-off channels within the riparian forest. Sampled in high density in the sampling subunits P10-B, P10-C and T1-H, this species exhibits a positive correlation with the silt fraction of the sediments and a negative correlation with the fine sand, medium sand, coarse sand and very coarse sand fractions (Table 2).

Geo-environment I-b) *Inga vera* Willd. Subsp. *affinis* (DC.) T.D. Penn.: this species was observed in the sampling subunits P6-A, P6-B, P10-B, P10-C, T1-E, T1-F, T1-I and T2-G. It occurs in the riparian forest in seasonally flooded environments and is dense in certain paleochannels. This species exhibits a positive correlation with the silt fraction and a negative correlation with the fine sand, medium sand, coarse sand and very coarse sand fractions (Table 2).

Geo-environment III-a) *Qualea grandiflora* Mart.: this common species in the Cerrado was observed in the sampling subunits P1-A/B, P3-A/B/C and P8-A/B/C (woodland savanna). This species exhibits a positive correlation with the fine and medium sand fractions of the sediments and a negative correlation with the silt fraction in the study area (Table 2).

Geo-environment III-b) *Handroanthus ochraceus* (Cham.) Mattos: this species is present only in Cerrado areas and was observed in the sampling subunits P3-A and P8-A (woodland savanna), as well as in various Cerrado regeneration areas. It shows a positive correlation with the fine and medium sand fractions and a negative correlation with the silt fraction (Table 2).

Geo-enviroment IV-a) *Cedrela fissilis* Vell.: this species is present in the sampling subunits P7-C, T1-C, T1-D, T4-A, T4-B, T4-E, T4-J, as well as in other areas in the SSF continuum, outside of the floodplain. It exhibits a positive correlation with the very coarse, coarse, and medium sand fractions and a negative correlation with the silt and fine sand fractions of the sediments (Table 2).

Geo-enviroment IV-b) *Zeyheria tuberculosa* (Vell.) Bureau: this species is present only in the riparian forest without fluvial influence (upland) in the sampling subunits P7-A, P7-B, P7-C, T1-A and T4-C, as well as in other unsampled areas that are still in the continuum of this physiognomy. This species exhibits a positive correlation with the very coarse and coarse sand and silt fractions and a negative correlation with the silt and fine sand sediment fractions (Table 2).

Table 2 indicates that the species associated with geo-environment I (*C. brasiliense* and *I. vera* subsp. *affinis*) exhibit a negative correlation with the discriminating grain size fractions of geo-environments III and IV (Aquidauana Formation). The species is associated with geo-environments III (*Q. grandiflora* and *T. ochracea*) and IV (*C. fissilis* and *Z. tuberculosa*) and shows a negative correlation with the discriminating fractions of geo-environments I and II (silt for geo-environment I and II and very fine sand for geo-environment II). The near zero value of the Student's *t*-test indicates that the sampled population has a normal distribution.

### 3.2. Target-species discrimination

#### 3.2.1. Leaf traits variation

The null hypothesis, which states the absence of leaf trait variations between target-species was rejected for all leaf pigments and accepted for leaf structural compounds and water ($\alpha = 0.05$), with the exception of cellulose (rejected at 0.1% of significance), for the rainy season's dataset. However, when analyzed the leaf traits from the dry season collection, only carotenoids + xanthophyll content did not present variations ($\alpha = 0.05$) between target-species (Table 3).

The mean values of hemicellulose ($\alpha = 0.01$) and insoluble lignin and water ($\alpha = 0.05$) content, from each target-species, also showed

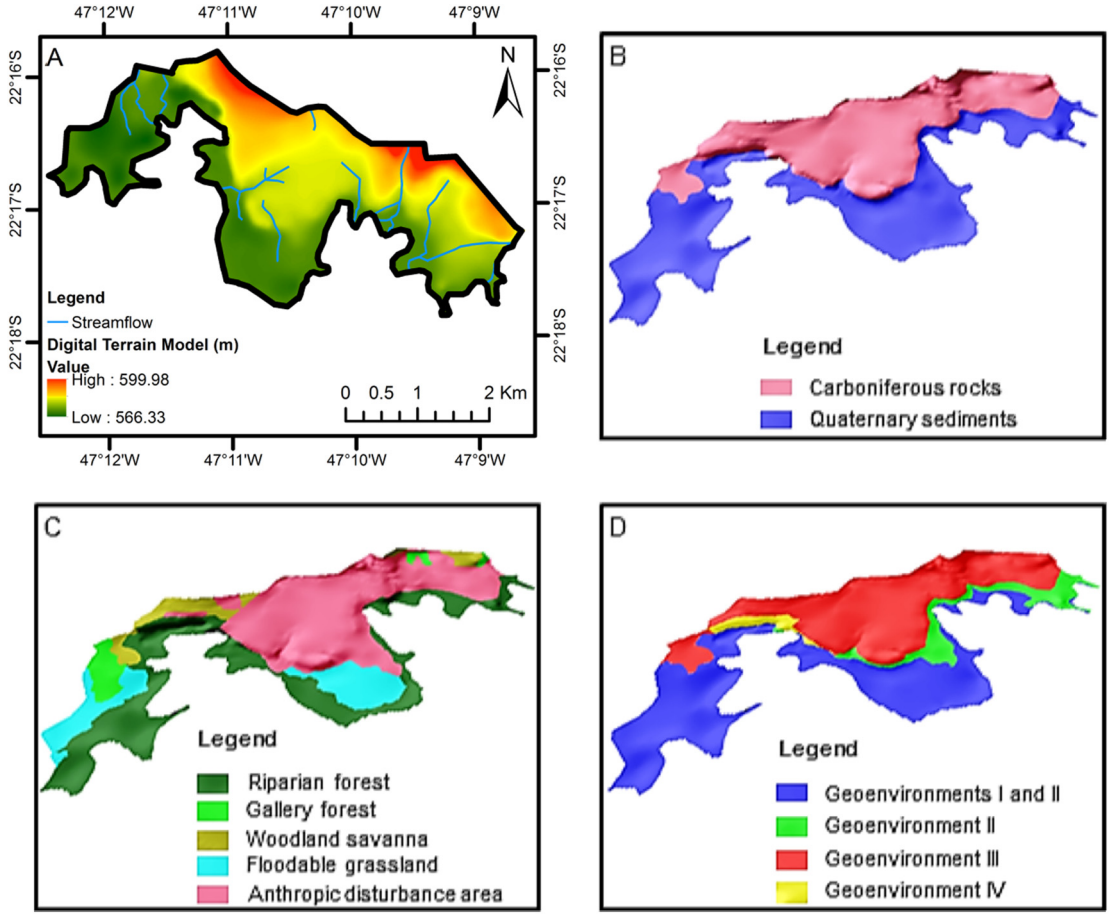


**Fig. 6.** Schematic 2D and 3D visualization of the spatial distribution of the elevation and drainage lines (A), local geology (B), vegetation physiognomies (C), and geo-environments (D) of the study area.

**Table 2**
Significant Spearman's coefficients (ρ) between species and sediments grain sizes.

| Positive correlation | | | Negative correlation | | |
|---|---|---|---|---|---|
| Parameter | ρ | *t*-Student | Parameter | ρ | *t*-Student |
| *Calophyllum brasiliense* | | | | | |
| Silt | 0.709 | 0.000 | Fine sand | −0.727 | 0.000 |
| | | | Medium sand | −0.624 | 0.000 |
| | | | Coarse sand | −0.583 | 0.000 |
| | | | Very coarse sand | −0.416 | 0.006 |
| *Inga vera* subsp. *Affinis* | | | | | |
| Silt | 0.688 | 0.000 | Fine sand | −0.656 | 0.000 |
| | | | Medium sand | −0.626 | 0.000 |
| | | | Coarse sand | −0.614 | 0.000 |
| | | | Very coarse sand | −0.491 | 0.001 |
| *Qualea grandiflora* | | | | | |
| Fine sand | 0.660 | 0.000 | Silt | −0.485 | 0.001 |
| Medium sand | 0.583 | 0.000 | | | |
| *Handroanthus ochraceus* | | | | | |
| Fine sand | 0.634 | 0.000 | Silt | −0.483 | 0.001 |
| Medium sand | 0.568 | 0.000 | | | |
| *Cedrela fissilis* | | | | | |
| Very coarse sand | 0.796 | 0.000 | Silt | −0.632 | 0.000 |
| Coarse sand | 0.737 | 0.000 | Very fine sand | −0.558 | 0.000 |
| Medium sand | 0.503 | 0.001 | | | |
| *Zeyheria tuberculosa* | | | | | |
| Very coarse sand | 0.917 | 0.000 | Very fine sand | −0.619 | 0.000 |
| Coarse sand | 0.720 | 0.000 | Silt | −0.547 | 0.000 |
| Clay | 0.576 | 0.000 | | | |

statistically significant variations between seasons. While insoluble lignin and water contents decreased from the rainy to the dry seasons (60.26% to 18.95% and 38.49% to 24.25%, respectively), hemicellulose contents increased significatively (from 4.42% to 7.83%).

Within all leaf traits, those that had variations with at least 0.1% of significance are highlighted in the biochemical discrimination of the target-species. HAOC had the highest average content of chlorophyll *a* in the rainy season (Fig. 7-A), which is statistically significantly ($\alpha = 0.01$) higher than those observed for all other target-species. For the other four species, the chlorophyll *a* content did not present significant contrasts. INVE leaves had anthocyanin content higher than the other four species ($\alpha = 0.05$ in comparison to QUGR, and $\alpha = 0.01$ to the other species), while the QUGR anthocyanin content was higher ($\alpha = 0.05$) in comparison to the values obtained from CABR, CEFI and HAOC leaves (Fig. 7-B). In the rainy season, HAOC and CABR had the highest cellulose content, being HAOC values higher ($\alpha = 0.05$) than CABR values (Fig. 7-C). TAOC cellulose content was higher ($\alpha = 0.01$) than the content in CEFI, INVE, and QUGR. Cellulose in CABR leaves was higher than in leaves from CEFI ($\alpha = 0.05$), and from QUGR and INVE ($\alpha = 0.01$), respectivetely.

When analyzing the target-species leaf traits from the dry season, variation in hemicellulose, insoluble lignin, and water content is highlighted. Hemicellulose content is lower ($\alpha = 0.01$) in INVE leaves than in the other four species' leaves. While QUGR and HAOC are considered indicators of the same geo-environment – interfluvial hills and plateaus of the Aquidauana Formation –, QUGR had lower hemicellulose content ($\alpha = 0.05$) than HAOC in the drought period (Fig. 7-D). A different behavior was observed among these species when analyzing the insoluble lignin content (Fig. 7-E). Both species had lower

**Table 3**
Null hypothesis tests (ANOVA/F test), where $H_0$ states the absence of leaf traits variation between target-species, for both rainy and dry seasons.

| Leaf trait | Rainy season | | Dry season | |
|---|---|---|---|---|
| | $H_0$ | *p*-Value | $H_0$ | *p*-Value |
| Chlorophyll *a* (μg/mg) | Rejected*** | 3.73e−04 | Rejected* | 1.46e−02 |
| Chlorophyll *b* (μg/mg) | Rejected* | 2.20e−02 | Rejected** | 8.01e−03 |
| Carotenoids + xanthophyll (μg/mg) | Rejected* | 3.07e−02 | Accepted* | 5.34e−02 |
| Anthocyanin (Δ/mg) | Rejected*** | 2.14e−04 | Rejected** | 2.06e−03 |
| Cellulose (mg/g) | Rejected*** | 9.05e−05 | Rejected** | 8.37e−03 |
| Hemicellulose (mg/g) | Accepted* | 1.24e−01 | Rejected*** | 3.81e−04 |
| Insoluble lignin (%) | Accepted* | 6.52e−02 | Rejected*** | 7.28e−06 |
| Water (%) | Accepted* | 1.92e−01 | Rejected*** | 8.71e−06 |

*** F significative at 0.1% of probability.
** F significative at 1% of probability.
* F significative at 5% of probability.

concentrations of structural compounds in comparison to the other target-species (α = 0.01), without statistically significant variation (α = 0.05) among themselves. Lignin content was also lower (α = 0.05) in CEFI than in CABR leaves. HAOC water content was statistically lower (α = 0.01) than the content in the QUGR, CABR, INVE and CEFI leaves, respectively, in the dry season (Fig. 7-F). In addition, CEFI presented lower water content than QUGR, CABR (α = 0.01) and INVE (α = 0.05); and INVE lower than QUGR (α = 0.05).

### 3.2.2. Spectral discrimination at leaf scale

The target-species classification using VSWIR reflectance spectra and EMs selected using IES had the highest overall accuracy (90.0%) and Kappa coefficient (0.875) for the rainy season. The worst results were the classifications based on CR VIS spectra, independent of the metric used for EM selection (50.0% e 0.375). For the dry season, the highest overall accuracy (85.0%) and Kappa coefficient (0.814) were obtained when classifying CR VSWIR data with EM selected via EMC. The target-species spectral discrimination in the dry season was worse when using CR VIS and NIR datasets with EM selected via EMC; both classifications presented the lowest accuracies (42.5% 0.294).

For both rainy and dry seasons, statistically significant differences (α = 0.05) were not observed between classification accuracies performed: on various spectral ranges (VIS, NIR, SWIR e VSWIR), of CR or reflectance spectra, and also by using EM selected by different metrics (i.e. EMC, IES (IES*)). While the highest accuracy was achieved for the dataset from the rainy season, the second and third ones were achieved for data from the dry season. Statistically significant (α = 0.05) variations between classification accuracies using datasets from both seasons (rainy vs. dry) were not observed, even when analyzing the different spectral ranges separately.

By considering the highest producer's (PA) and user's (UA) accuracies from each target-species class, CABR, CEFI and QUGR were better discriminated with datasets from the rainy season (SWIR/IES* for CABR and QUGR, and VSWIR/IES for CEFI). INVE and HAOC were better discriminated when using CR VSWIR data and EM selected via EMC, from the dry season (Table 4). However, only HAOC had statistically significant variability (α = 0.05) between class-specific accuracies from classifications performed on data of both seasons. While this class had an average user's accuracy of 93.75% in the dry season, in the rainy season it was of 72.98%. In the rainy season, more spectra from other target-species were erroneously assigned to the HAOC class. There are no statiscally significant (α = 0.05) variations between class-specific accuracies from different spectral ranges, at both seasons.

Spectral discrimination errors between pairs of target-species were significant (α = 0.05) only when analyzing spectra from the dry season. The highest misclassification was observed for CEFI & QUGR (14.10% of their spectra) followed by CABR & QUGR (12.53%). There were no observed misclassifications between INVE & HAOC and QUGR & HAOC. The misclassification rates of species pairs did not variate, significantly at α = 0.05, between classifications performed on datasets from the rainy and from the dry seasons.

### 3.2.3. Spectral discrimination at crown scale

The highest classification accuracy between the three target-species addressed at the crown scale –CEFI, INVE and QUGR; each one representative of a different geo-environment – was achieved by using CR VSWIR dataset and EMs selected using IES (overall accuracy = 70.59% and Kappa coefficient = 0.551). Although we observed differences between the overall accuracy averages achieved when classifying VSWIR (63.24, σ = 5.94), VIS-NIR1 (62.75, σ = 5.72), and NIR2-SWIR (50.00, σ = 8.55) datasets, their variations were not statistically significant at, at least, 5% of significance. This was also observed when considering the variations between classifications performed with: reflectance or CR spectra and EM selected using EMC or IES metrics.

*Cedrela fissilis* (CEFI) and *Inga vera* subsp. *affinis* (INVE) were better discriminated when classifying VIS-NIR1 reflectance spectra: the first one with an EM selected via EMC (producer's accuracy: PA = 66.67%; and user's accuracy: UA = 88.89%), and the second one with an EM selected via IES (PA = 80.95% and UA = 89.47%). *Qualea grandiflora* (QUGR) had the highest and most balanced class-specific accuracies (PA = 72.22% and UA = 68.42%) when discriminated with CR VSWIR dataset and EM selected via IES. Only the UA metric variated between target-species classes (α = 0.05), between INVE and QUGR particularly. While INVE showed average of UA equal to 69.18% (σ = 12.60%), QUGR showed 56.93% (σ = 10.02%). There were no statistically significant variations of PA and UA obtained by the target-species classes in classifications performed on various spectral ranges datasets, and with reflectance or CR spectra.

The spectral misclassifications were statistically different between the pairs of species: CEFI & INVE and CEFI & QUGR, in both VIS-NIR1 and VSWIR datasets classifications (Fig. 8). By using VIS-NIR1 dataset, an average of 7.58% (σ = 8.02%) CEFI & INVE validation spectra were erroneously classified between them, while an average of 24.17% (σ = 9.57%) CEFI & QUGR validation spectra were confused among themselves. In the VSWIR dataset classifications the misclassification rate averages for CEFI & INVE was 11.36% (σ = 3.81%), and for CEFI & QUGR was 20.83% (σ = 4.19%).

### 3.2.4. Spectral discrimination at crown scale including other vegetation classes

The most accurate classification of the eleven vegetation classes – which are representative of the landscape and include the target-species – was achieved with the CR VSWIR dataset and EMs selected using IES (overall accuracy = 45.26% and Kappa coefficient = 0.397). There were no significative differences (α = 0.05) between MESMA with two EM performance, either varying the spectral ranges, types of spectra, and EM selection metrics.

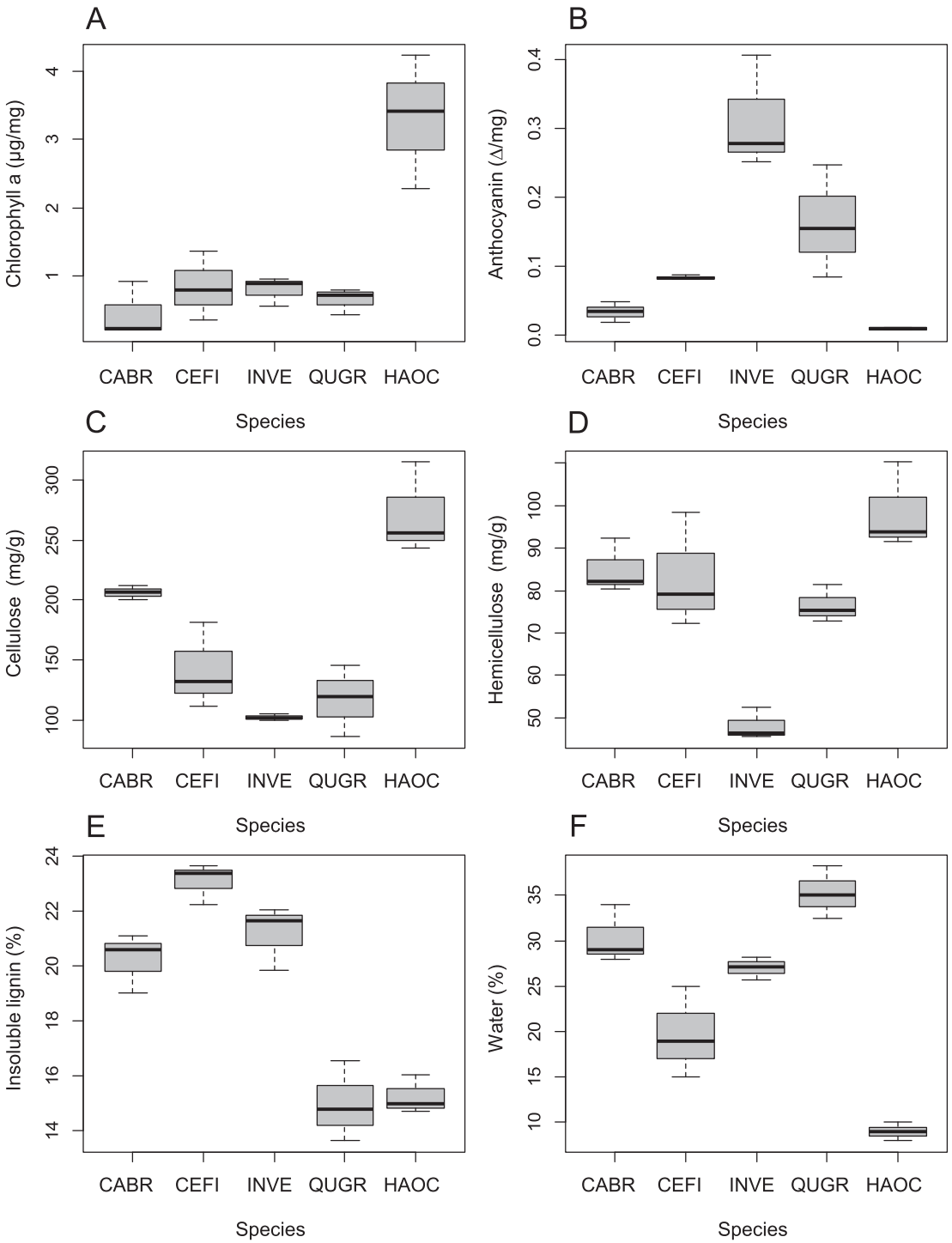


**Fig. 7.** Target-species leaf traits variations, at 0.1% of significance: chlorophyll *a* (µg/mg) (A), anthocyanin (Δ/mg) (B) and cellulose (mg/g) (C), from the rainy season; and hemicellulose (mg/g) (D); insoluble lignin (%) (E) and water (%) (F), from the dry season. The lower and upper edges of the boxes indicate the interval between 25 and 75% of the data distribution. The horizontal lines outside the boxes indicate the minimum and maximum values of the dataset that were not outliers.

PA was lower ($\alpha = 0.05$) for the CEFI class than for INVE and QUGR classes; such difference was not observed for the UA achieved by those classes. PA values from INVE class were different ($\alpha = 0.05$) troughout spectral ranges; this class achieved PA average of 57.14% ($\sigma = 12.29\%$) in VSWIR, 44.05% ($\sigma = 14.22\%$) in NIR2-SWIR, and 25.0% ($\sigma = 9.02\%$) in VIS-NIR1 datasets classifications. In contrast, CEFI class presented higher ($\alpha = 0.05$) PA values in classifications based on CR spectra (30.56, $\sigma = 16.39\%$) than in raw reflectance spectra (11.11%, $\sigma = 12.55\%$). The highest and most balanced PA and UA differed between the target-species classes, as shown in the Table 5.

### 3.3. Target-species fraction mapping

QUGR sub-pixel fraction was the most accurately modeled among the target-species. However, its average fraction abundance was lower within its validation pixels (0.53, $\sigma = 0.23$) and crowns (0.52, $\sigma = 0.24$) than the average fraction of CEFI within CEFI validation pixels and crowns (0.58, $\sigma = 0.23$, and 0.57, $\sigma = 0.21$) and than INVE

**Table 4**
Producer's (PA) and user's (UA) accuracies, in percentage, for each target-species class, reached in the most accurated classifications per spectral range (VSWIR, VIS, NIR, SWIR) and season (rainy and dry). CABR = *Calophyllum brasiliense*. CEFI = *Cedrela fissilis*. INVE = *Inga vera* subsp. *affinis*. QUGR = *Qualea grandiflora* HAOC = *Handroanthus ochraceus*. VSWIR = Visible to shortwave infrared. VIS = Visible. NIR = Near infrared. SWIR = Shortwave infrared. CR = Continuum-removed. IES = Interative Endmember Selection method. EMC = EAR-MASA-CoB metrics for endmember selection.

| Dataset/metric | CABR | | CEFI | | INVE | | QUGR | | HAOC | |
|---|---|---|---|---|---|---|---|---|---|---|
| | PA | UA | PA | UA | PA | UA | PA | UA | PA | UA |
| Rainy season | | | | | | | | | | |
| VSWIR/IES | 87.5 | 87.5 | 87.5 | 87.5 | 100.0 | 88.9 | 75.0 | 100.0 | 100.0 | 88.9 |
| SWIR/IES* | 87.5 | 100.0 | 12.5 | 50.0 | 100.0 | 66.7 | 100.0 | 100.0 | 100.0 | 72.7 |
| NIR/IES | 62.5 | 83.3 | 100.0 | 66.7 | 75.0 | 85.7 | 37.5 | 75.0 | 87.5 | 63.6 |
| VIS/IES | 87.5 | 87.5 | 62.5 | 62.5 | 100.0 | 72.7 | 50.0 | 80.0 | 50.0 | 66.7 |
| Dry season | | | | | | | | | | |
| VSWIR(CR)/EMC | 87.5 | 70.0 | 75.0 | 85.7 | 100.0 | 100.0 | 62.5 | 83.3 | 100.0 | 100.0 |
| SWIR(CR)/IES | 87.5 | 87.5 | 62.5 | 100.0 | 100.0 | 88.9 | 87.5 | 70.0 | 87.5 | 87.5 |
| NIR(CR)/IES | 75.0 | 85.7 | 62.5 | 55.6 | 75.0 | 66.7 | 62.5 | 55.6 | 75.0 | 100.0 |
| VIS/IES | 62.5 | 100.0 | 75.0 | 50.0 | 50.0 | 80.0 | 62.5 | 55.6 | 87.5 | 87.5 |

average fraction within its validation pixels and crowns (0.62, σ = 0.22, and 0.62, σ = 0.19). In contrast, excluding outliers, the maximum QUGR sub-pixel fraction observed within pixels and crowns from other vegetation types' classes was always lower than 0.49 and 0.45, respectively. This corroborates the absence of commission errors for QUGR class observed in the previous section. It was not observed for CEFI and INVE classes, which presented high abundance (≥0.50) of their fractions within other vegetation types' crowns as at pixel scale (for six and seven classes respectively) as at crown scale (for two and four classes respectively) (Table 6).

To avoid any erroneous target-species mapping (i.e. to exclude possibilities of commission errors), we used the maximum (outliers)

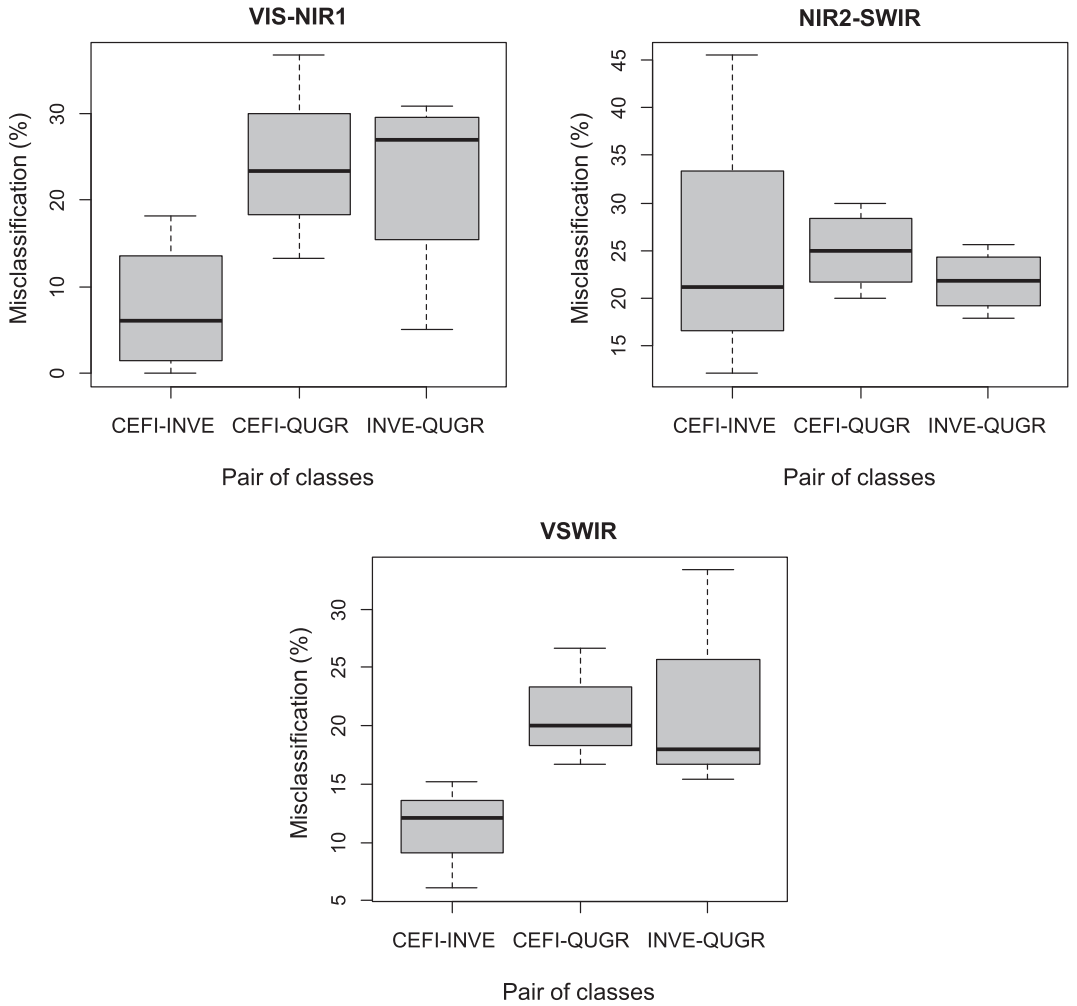


**Fig. 8.** Misclassification rates between pairs of target-species: *Cedrela fissilis* (CEFI), *Inga vera* subsp. *affinis* (INVE) and *Qualea grandiflora* (QUGR), observed in crown spectra classifications of VIS-NIR1, NIR2-SWIR e VSWIR datasets. The lower and upper edges of the boxes indicate the interval between 25 and 75% of the data distribution. The horizontal lines outside the boxes indicate the minimum and maximum values of the dataset that were not outliers.

**Table 5**
Best producer and user's accuracies for the target-species classes (CEFI, INVE and QUGR) achieved in the Multiple Endmember Spectral Mixture Analyses with eleven vegetation classes. CEFI = *Cedrela fissilis*. INVE = *Inga vera* subsp. *affinis*. QUGR = *Qualea grandiflora*. VSWIR = Visible to shortwave infrared. CR = Continuum-removed. IES = Interative Endmember Selection method. EMC = EAR-MASA-CoB metrics for endmember selection.

| Class | Dataset/metric | Producer's accuracy (%) | User's accuracy (%) |
|---|---|---|---|
| CEFI | VSWIR(CR)/IES and EMC | 41.67 | 45.45 |
| INVE | VSWIR/IES and EMC | 71.43 and 61.90 | 48.39 and 59.09 |
| QUGR | VSWIR(CR)/EMC | 61.11 | 100.00 |

target-species fraction abundances observed within a crown of other vegetation type to slice and classify the target-species fraction images into target-species pixels and other vegetation type pixels. The maximum outliers were 0.79 of CEFI fraction within a QUGR crown, 0.82 of INVE fraction in POAspp canopy, and 0.62 of QUGR fraction within a EUCspp crown. Thus, pixels classified as CEFI are those with > 80% of its fraction, as INVE those pixels with at least 85% of its fraction, and as QUGR those ones with > 65% of its fraction (Fig. 9). It is important to highlight the spatial distribution of QUGR fraction in its majority on the Aquidauana Formation sequence, enhancing the geological contact (zoom (inset) in Fig. 9).

## 4. Discussion

Discussion regarding target species spectral discrimination at different levels, such as leaf, crown and pixel, are presented in Sections 4.1, 4.2 and 4.3, respectively. In Section 4.3, we discuss our pixel-based fraction mapping of tree species that we previously selected in the field as geological indicators. Geobotany research perspectives for future Earth Observing Missions are highlighted as well.

### 4.1. Leaf level analyses

Classifications accuracies within several spectral ranges were similar between rainy and dry seasons. More accurate classifications using VIS, NIR and VSWIR datatasets classifications were obtained with spectra from the rainy season. Superior classification results using SWIR datasets were obtained with spectra from the dry season. These observations match with significant variation in leaf pigments (excluding chlorophyll *b*) and cellulose, and the absence of hemicellulose, insoluble lignin and water content differences between species in the rainy season. Those compounds (hemicellulose, insoluble lignin and water), which interact with the infrared energy, only varied between species in the dry season ($\alpha = 0.001$). They were the unique leaf traits that variated significantly between seasons.

The set of target-species was well discriminated in both seasons. Considering the best classification results per spectral range, user and producer's accuracies were higher than 87.5%. This fact seems to be related to the high leaf chemical variation among species, indicating different physiological behaviors between them during the wet and drought periods. Castro-Esau et al. (2006) and Hesketh and Sánchez-Azofeifa (2012) have observed similar results between species with several habitats in dry and rainy Neotropical forests in North and Central America.

Distinct spectral and chemical patterns were not found among indicator species of the same geological facies. Thus, even if the pairs of species CABR & INVE, and QUGR & HAOC, which are associated to the same edaphic conditions, are considerated groups of distinct functional types (Lavorel et al., 1997; Reich et al., 2003); they do not represent different groups chemically and spectrally. An exception was observed for insoluble lignin content in the dry season. It was statistically lower in leaves from both QUGR and HAOC, which originated from Cerrado and are considered indicators of the hills and plateaus of the Aquidauana Formation, than in leaves from CABR, CEFI and INVE, which originated from the Seasonal Semi-deciduous Forest and occur in the Riparian Forests (upland and lowland). This result contrasts with observations by Asner (1998), who found various lignin and cellulose contents within Cerrado species in central Brazil. Asner et al. (2009) studied chemical and spectral properties of Australian tropical tree species along an altitudinal gradient with different substrates. They observed tendencies of higher specific leaf area, water, chlorophyll *a* and carotenoids at the lower elevation sites. Leaf chemistry was related to leaf spectra, and the high spectral variability was found to be driven at the species level.

The rates of misclassification between pairs of species were significant only in the dry season. QUGR & CEFI had the highest misclassification (14.10% of their spectra), followed by QUGR & CABR (12.53%). Note that each one is representative of a different geo-enviroment. Based on overall results from the leaf scale analyses, we could conclude that pairs of indicator species, considerated as distinct functional types, do not constitute distinct optical types. This fact has been discussed in the literature, indicating that the energy usage is more likely related to species' strategies than to functional groups (e.g., Roth

**Table 6**
Percentage of modeled pixels (%MP) and crowns (%MC) in the MESMA images unmixing and mean values ($\overline{X}$) (standard deviation (σ)) of the target-species fractions within class validation samples. VSWIR = Visible to shortwave infrared image. CR = Continuum-removed image. IES = Interative Endmember Selection method. EMC = EAR-MASA-CoB metrics for endmember selection. CEFI = *Cedrela fissilis*. INVE = *Inga vera* subsp. *affinis*. QUGR = *Qualea grandiflora*.

| Class | VSWIR(CR)/IES | | | | VSWIR/IES | | | | VSWIR(CR)/EMC | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | %MP | %MC | CEFI fraction | | %MP | %MC | INVE fraction | | %MP | %MC | QUGR fraction | |
| | | | Pixel | Crown | | | Pixel | Crown | | | Pixel | Crown |
| | | | $\overline{X}$ (σ) | $\overline{X}$ (σ) | | | $\overline{X}$ (σ) | $\overline{X}$ (σ) | | | $\overline{X}$ (σ) | $\overline{X}$ (σ) |
| CEFI | 100.0 | 100.0 | 0.58 (0.23) | 0.57 (0.21) | 100.0 | 100.0 | 0.38 (0.19) | 0.37 (0.18) | 96.8 | 100.0 | 0.22 (0.15) | 0.24 (0.14) |
| CERspp | 96.3 | 100.0 | 0.39 (0.19) | 0.38 (0.16) | 94.4 | 95.0 | 0.35 (0.22) | 0.34 (0.19) | 81.5 | 70.0 | 0.22 (0.18) | 0.22 (0.14) |
| CITspp | 87.0 | 100.0 | 0.27 (0.18) | 0.25 (0.14) | 89.0 | 100.0 | 0.28 (0.24) | 0.27 (0.20) | 49.0 | 100.0 | 0.21 (0.17) | 0.18 (0.12) |
| DENDsp | 100.0 | 100.0 | 0.36 (0.18) | 0.36 (0.13) | 96.3 | 100.0 | 0.47 (0.16) | 0.47 (0.08) | 99.1 | 100.0 | 0.19 (0.12) | 0.19 (0.06) |
| EUCspp | 100.0 | 100.0 | 0.21 (0.14) | 0.21 (0.09) | 100.0 | 100.0 | 0.25 (0.17) | 0.26 (0.13) | 99.0 | 100.0 | 0.29 (0.20) | 0.29 (0.16) |
| INVE | 100.0 | 100.0 | 0.35 (0.18) | 0.35 (0.14) | 100.0 | 100.0 | 0.62 (0.22) | 0.63 (0.19) | 100.0 | 100.0 | 0.24 (0.16) | 0.25 (0.14) |
| RIFOspp | 91.3 | 95.0 | 0.32 (0.21) | 0.31 (0.17) | 95.3 | 95.0 | 0.29 (0.20) | 0.29 (0.17) | 66.1 | 60.0 | 0.18 (0.14) | 0.17 (0.11) |
| PINspp | 78.2 | 100.0 | 0.16 (0.10) | 0.16 (0.07) | 92.0 | 100.0 | 0.27 (0.21) | 0.27 (0.18) | 44.8 | 96.4 | 0.17 (0.13) | 0.16 (0.08) |
| POAspp | 100.0 | 96.4 | 0.31 (0.21) | 0.31 (0.16) | 93.1 | 100.0 | 0.37 (0.27) | 0.36 (0.27) | 90.1 | 78.6 | 0.27 (0.14) | 0.27 (0.12) |
| QUGR | 100.0 | 100.0 | 0.43 (0.18) | 0.43 (0.16) | 98.7 | 100.0 | 0.33 (0.21) | 0.33 (0.17) | 96.2 | 95.0 | 0.53 (0.23) | 0.52 (0.24) |
| SACspp | 100.0 | 100.0 | 0.22 (0.16) | 0.22 (0.11) | 100.0 | 100.0 | 0.31 (0.19) | 0.31 (0.15) | 90.6 | 100.0 | 0.20 (0.13) | 0.19 (0.09) |

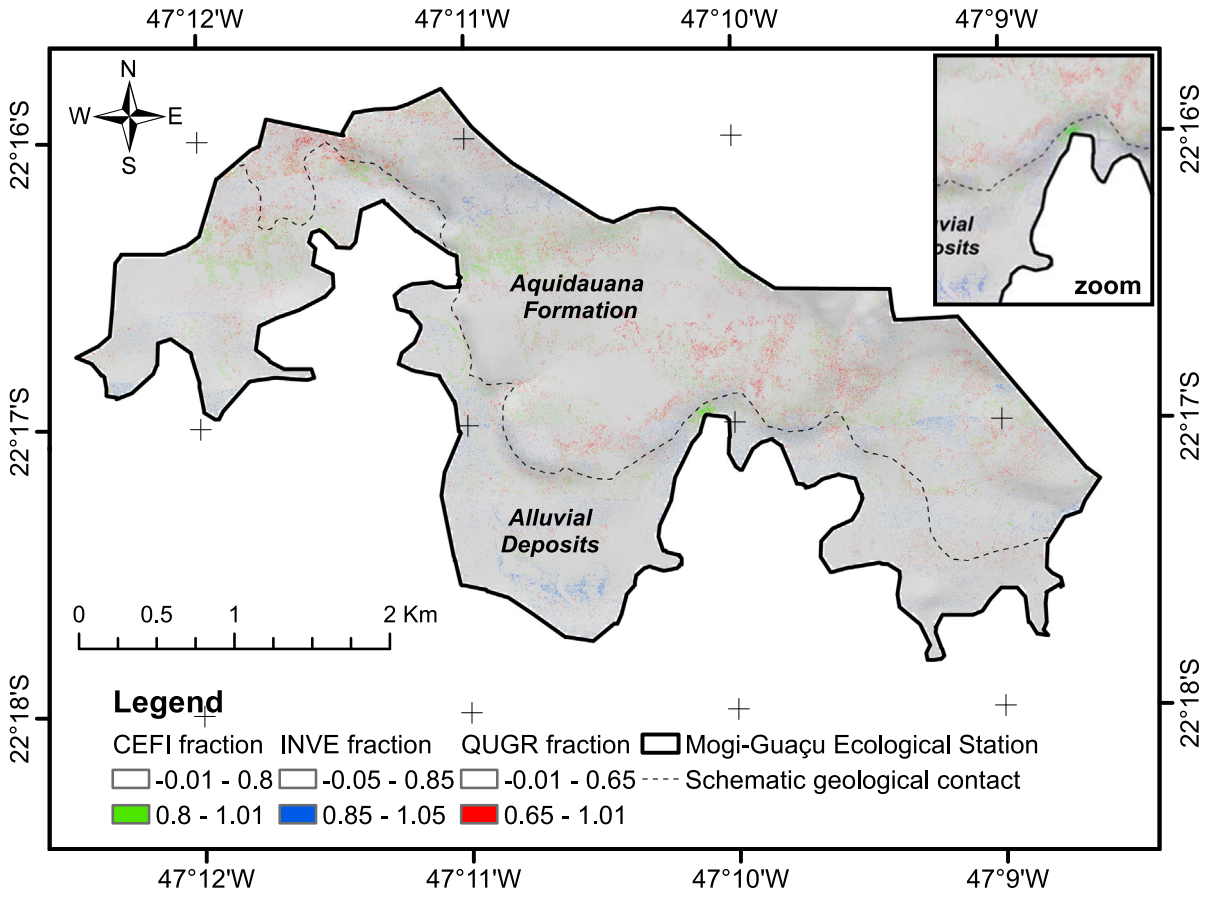


**Fig. 9.** Spectral fractions of *Cedrela fissilis* (CEFI – in green, ≥0.80), *Inga vera* subsp. *affinis* (INVE – in blue, ≥0.85) and *Qualea grandiflora* (QUGR – in red, ≥0.65), at the Mogi-Guaçu Ecological Station, São Paulo State, Brazil. Schematic geological contact, delineated at a scale of 1:10.000. Upright picture is a zoom where a geological contact can be delineated with the aid of the indicator species' fraction distribution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

et al., 2016). Thus, when a geobotany map at individual tree crown level is desired, investigators must take into account the fact that indicator species from various geological facies might present higher misclassification rates than those observed between species associated to the same geological environment.

### 4.2. Crown level analyses

Spectral discrimination at crown scale between three target-species – CEFI, INVE and QUGR, each one an indicator of a different geological facies – was found to be dependent on the inter-specific interactions with the incident light at the VIS-NIR1 (530–919 nm) spectral region. Chlorophylls and other pigments absorption features, as well as leaf (external and mesophyll structures) and crown (leaf area index and leaf angle distribution) scattering are responsible for spectral variations in visible-near infrared regions. Thus, phenological changes related to leaf senescence and abcision are found to be critical for the target-species discrimination at the imagery date, in the beginning of the dry season. However, while CEFI and INVE were better discriminated when using reflectance VIS-NIR1 spectra, QUGR was better discriminated by CR VSWIR spectra.

As observed at leaf scale, CEFI & QUGR had the highest misclassification rates at the crown scale, in both VIS-NIR1 and VSWIR (530–2352 nm) datasets spectral mixture analyses. While INVE (associated to quaternary sediments) is evergreen, CEFI and QUGR (associated to different facies of the Aquidauana Formation) are deciduous species. In addition, INVE leaves had higher chlorophyll *a* and *b* concentrations during the dry season in comparison to CEFI (ratios of 2.18 for chloro *a* and 1.98 for chloro *b*) and QUGR leaves (2.59 and 2.75, respectively); and its compound leaves have a winged petiole – which provides a greater photosynthetic area within its crowns. CEFI, which is deciduous, display crowns of various shapes with discontinuities. Its leaf fall is concentrated between May and July (Santos and Takaki, 2005), when the images were acquired. Thus the spectral difference among these species at the crown scale is contrasting: INVE spectra stand out for high green vegetation fraction (GV) and low non-photosynthetic vegetation (NPV) and shade while CEFI crowns had higher NPV and shade fractions (see Fig. 3). QUGR, also deciduous, usually experiences leaf turnover at the end of the dry season (e.g., Lenza and Klink, 2006; Silvério and Lenza, 2010). The highest water content, as well as the lowest insoluble lignin content in its leaves in the dry season might explain its better discrimination when using VSWIR than only VIS-NIR1 datasets. The water content ratios between QUGR and INVE and between QUGR and CEFI are of 1.31 and 1.79, and the lignin content ratios are of 0.71 and 0.65, respectively.

As observed in other investigations on species classification (e.g., Castro-Esau et al., 2006; Hesketh and Sánchez-Azofeifa, 2012), the overall accuracy from MESMA decreased (25.33%) when including more vegetation classes in the analyses. The inclusion of classes with several species (non-specific classes), such as native and non-native grasses (POAspp), other species from Cerradão (woodland savanna (CERspp)) and Riparian Forest (RIFOspp), also contributed for the increase of within-class spectral variability increases, decreasing the overall and class-specific accuracies. Only QUGR showed satisfactory class-specific accuracies when the additional eight classes of vegetation were included in the analyses, indicating that in the landscape there are

other vegetation types more spectrally similar to INVE and CEFI than to QUGR. It is important to highlight the absence of user's error for that last species class when MESMA was applied on CR VSWIR data using EM selected via EMC metrics.

The importance of the phenological stage in species discrimination – also supported by Hill et al. (2010) in a temperate deciduous forest and by Madonsela et al. (2017) in an African tropical savanna – indicates that each target-species will be better recognized in a particular season, and using various spectral ranges. Multitemporal approaches are recommended in further GbRS studies at individual tree crown level.

### 4.3. Indicator species mapping

The highest misclassification observed for CEFI & QUGR, previously found at both leaf and crown scales, was also observed at the pixel scale (after image unmixing). QUGR was the class that had the highest mean value of CEFI fraction sub-pixel abundance (0.43, in both pixel and crown validation samples), excluding the CEFI class. DENDspp was the class with highest INVE sub-pixel fraction (0.47, in both pixel and crown validation samples), excluding the INVE class; as INVE, DENDspp also stands out for high GV and low NPV and shade fractions (Amaral et al., 2015). The mean value of QUGR fraction sub-pixel abundance within pixels and crowns of other classes was always lower than 0.29.

CEFI is rare and had its fraction mapping influenced by brightness differences between flight lines. By contrast, INVE is dominant in the landscape. However, it might occur outside of the Alluvial Deposits sequence, where localized variations in micro-topography can lead to pelitic sediments deposition. These aspects, which were identified at this stage only, after fraction mapping, hindered satisfactory geobotanical relationships when analyzing CEFI and INVE fraction maps. There is a virtual absence of pixels classified as QUGR in the lowland dominated by quaternary sediments, except an important concentration at the westernmost region of the study area. This exception is likely more related to imperfections of the schematic geological contact delineation than to the hypothesis that assumes QUGR is an indicator of the hills and plateaus of the Aquidauana Formation, and to the image unmixing as well. We were not able to visit that region in the field, and so the contact has been delineated with the aid of a topographic chart (1:10.000). In constrast, the QUGR distribution was able to highlight some portions of the Aquidauana Formation edges, which might support local geological mapping. *Qualea grandiflora* (QUGR) is therefore considerated an efficient geobotanical indicator species of the hills and plateaus of the Aquidauana Formation in the study area – geological facies with distinct fractions of fine and medium sand in its sediments.

The pixel-based fraction mapping allowed us to conclude that only species with random and broad spatial distribution, but restricted to one geological facies, and with singular spectral behavior are able to be mapped using high spatial resolution hyperspectral images. Although intra-specific spectral variations constitute an important technique for geobotanic prospection in vegetated areas, when working with indicator species – which are adapted to their habitat and presented various phenological strategies – intra-specific spectral homogeneity and inter-specific spectral distinction (Castro-Esau et al., 2006) are key for geobotanical mapping in hyperdiverse forest formations, as the Neotropical ones.

It is worth mentioning that other classification techniques, as well as other remote sensing products, must be tested in further GbRS studies at individual tree crown level. In addition to the use of a multi-temporal approach (e.g., Somers and Asner, 2013) or techniques that have proven more accurate than MESMA (e.g., Somers and Asner, 2014), prior object segmentation has resulted in increased accuracy in automated species mapping in Neotropical hyperdiverse canopies (eg, Clark et al., 2005; Ferreira et al., 2016). Hyperspectral and LiDAR data fusion has also shown notable results on tropical species mapping (e.g., Colgan et al., 2012; Baldeck et al., 2014; Baldeck et al., 2015). As an example, Baldeck et al. (2014) mapped > 500,000 tree and shrub crowns in a South African savanna with an accuracy of 76%, by using Hi-FIS (384.8–1054.3 nm) and LiDAR data. The authors observed that nine species, including the seven most common in the landscape, are associated to topographic variables. They also clustered the species into communities, and found that 21.2% of the variations in species composition can be explained by topography. Nevertheless, the authors concluded that other soil-related factors might control the distribution of those species communities.

At the community level, Neotropical phytophysiognomies' distribution may also express the underlying geologic variation, and this poses a new research perspective for future Earth Observing Missions' applications. In addition to the HyspIRI (Hyperspectral Infrared Imager) and EnMap (Environmental Mapping and Analysis Program) hyperspectral missions, GEDI (Global Ecosystem Dynamics Investigation Lidar), ECOSTRESS (ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station), and FLEX (Fluorescence Explorer) missions must be explored in GbRS studies.

Future research efforts should also focus on how transferable are the results found here between areas, and how this GbRS methodology is suitable for geological delineation in larger terrains.

## 5. Conclusions

Our work shows the possibility to aid geological mapping through field selection, spectral discrimination and hyperspectral pixel-based mapping of indicator tree species in a heavily vegetated Neotropical landscape.

At the study site in southeastern Brazil, the set of target-species was well discriminated at leaf level in both rainy and dry seasons (90.0% and 85.0% of global accuracy, respectively). User and producer's accuracies were higher than 87.5% for all target-species. This fact seems to be related to the high leaf chemical variation among species, indicating different physiological strategies between them. Pairs of species associated to the same edaphic conditions, in general, do not constitute unlike chemical and spectral groups. An exception is observed when leaf insoluble lignin content is analyzed; it distinguishes the indicator species of the hills and plateaus of the Aquidauana Formation from the others in the dry season. We observed that classifications with higher accuracy based on VIS, NIR and VSWIR datatasets were obtained with leaf spectra from the rainy season. Using SWIR datasets exclusively, the classification results with higher accuracy are from data collected in the dry season. Such observations match with significant variation in leaf pigments (excluding chlorophyll *b*) and cellulose in the rainy season and with hemicellulose, insoluble lignin and water (which interact with the infrared energy) in the dry season.

The crown scale spectral discrimination between three target-species, each one an indicator of a different geological facies, was found to be commonly dependent on the inter-specific interactions with the incident light at the VIS-NIR1 (530–919 nm) spectral region (reaching an overall accuracy of 70.6%). While two of the species (*Cedrela fissilis* and *Inga vera* subsp. *affinis*) were better discriminated with VIS-NIR1 data, the other (*Qualea grandiflora*) was better discriminated with VSWIR data. In addition to the variations in deciduousness, the contrast in chlorophylls *a* and *b* and in water and insoluble lignin between species supports those species-specific spectral discriminations in the dry season.

When other vegetation types spectra are inserted into the MESMA, the success of the target-species discrimination at the crown scale changes. Only one of the species selected at the field as a geological indicator species proved to have particular ecological traits required to be successfully mapped on the mosaic of the airborne hyperspectral images. Those requirements are found to be the random and broad spatial distribution on the terrain of the target geological facies, as well as particular physiological, phenological and spectral behavior at the imagery acquisition date. Ultimately, we show that the spatial

distribution of *Qualea grandiflora* spectral fraction concurs with hills and plateaus of the Aquidauana Formation in the study area, highlighting the importance of tree species indicators as an auxiliary means for geologic mapping in highly vegetated Neotropical regions, where rock exposures are typically limited.

## Acknowledgment

The authors thank the São Paulo State Research Foundation (FAPESP) for supporting this research (grant # 2010/51758-2), the FOTOTERRA/SpecTIR team, particularly Guilherme Pinho and Conrad Wright, for providing the hyperspectral images, and the São Paulo State Forest Institute (IF-SP), for supporting field campaigns. The authors also appreciate the collaborations of Dirceu de Souza, Carlos Carrasco Arbieto, Sidney Schaberle Goveia, Lênio Soares Galvão, João Roberto dos Santos, Laerte Guimarães Ferreira Júnior, Cleyton de Carvalho Carneiro, Waldir Mantovani, Arnaldo Pereira (in memoriam), Omar Jorge Di Dio, Giovani Germano, Lucíola Magalhães, Giuliana Quitério, Flávia Mendes, Natasha Penatti, Daniela Gamito, Vitor Ribeiro, Francisco Cavallaro, Keely Roth and Seth Peterson. C.H.A thanks FAPESP for the doctoral scholarship (grant # 2010/51718-0). C.R.S.F and T.I.R.A. thank the National Council for Scientific and Technological Development (CNPq) for their research grants (grants # 309712/2017-3 and # 302925/2015-5, respectively).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2018.07.009.

## References

Adams, J.B., Smith, M.O., Gillespie, A.R., 1993. Imaging spectroscopy: Interpretation based on spectral mixture analysis. In: Pieters, C.M., Englert, P.A.J. (Eds.), Remote geochemical analysis: Elemental and Mineralogical Composition. Press Syndicate of the University of Cambridge, New York.

Almeida-Filho, R., Castelo Branco, R.M.G., 1992. Location of kimberlites using Landsat Thematic Mapper images and aerial photographs: the Redondão diatreme, Brazil. Int. J. Remote Sens. 13, 1449–1457.

Almeida-Filho, R., Vitorello, I., 1997. Remote sensing data integration in the definition of hydrotermally altered areas in vegetated terrain, central Brazil. Int. J. Remote Sens. 18 (8), 1835–1842.

Almeida-Filho, R., Vitorello, I., Correia, V.R., 1996. Use of Landsat Thematic Mapper imagery as mineral prospecting tool in the tin province of Goias, Brazil. Geocarto Int. 11 (1), 61–69.

Alvarez-Añorve, M.Y., Quesada, M., Sánchez-Azofeifa, G.A., Avila-Cabadilla, L.D., Gamon, J.A., 2012. Functional regeneration and spectral reflectance of trees during succession in a highly diverse tropical dry forest ecosystem. Am. J. Bot. 99 (5), 816–826.

Amaral, C.H., Roberts, D.A., Almeida, T.I., Souza Filho, C.R., 2015. Mapping invasive species and spectral mixture relationships with neotropical woody formations in southeastern Brazil. ISPRS J. Photogramm. Remote Sens. 108, 80–93.

Arden Jr., D.D., Westra, R.N., 1977. Remote sensing of geobotanical relations in Georgia. In: Technical Report (NASA-CR-150709), (103p).

Asner, G.P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. Remote Sens. Environ. 64 (3), 234–253.

Asner, G.P., 2008. Hyperspectral remote sensing of canopy chemistry, physiology, and biodiversity in tropical rainforests. In: Kalacska, M., Sanchez-Azofeifa, G.A. (Eds.), Hyperspectral Remote Sensing of Tropical and Subtropical Forests. CRC Press Taylor & Francis Group, Boca Raton, pp. 261–296.

Asner, G.P., Martin, R.E., 2008. Spectral and chemical analysis of tropical forests: scaling from leaf to canopy levels. Remote Sens. Environ. 112, 3958–3970.

Asner, G.P., Martin, R.E., 2009. Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests. Front. Ecol. Environ. 7 (5), 269–276.

Asner, G.P., Martin, R.E., 2016. Spectranomics: emerging science and conservation opportunities at the interface of biodiversity and remote sensing. Glob. Ecol. Conserv. 8, 212–219.

Asner, G.P., Carlson, K.M., Martin, R.E., 2005. Substrate age and precipitation effects on Hawaiian forest canopies from spaceborne imaging spectroscopy. Remote Sens. Environ. 98 (4), 457–467.

Asner, G.P., Martin, R.E., Ford, A.J., Metcalfe, D.J., Liddell, M.J., 2009. Leaf chemical and spectral diversity in Australian tropical forests. Ecol. Appl. 19 (1), 236–253.

Baldeck, C.A., Colgan, M.S., Féret, J.B., Levick, S.R., Martin, R.E., Asner, G.P., 2014. Landscape-scale variation in plant community composition of an African savanna from airborne species mapping. Ecol. Appl. 24 (1), 84–93.

Baldeck, C.A., Asner, G.P., Martin, R.E., Anderson, C.B., Knapp, D.E., Kellner, J.R., Wright, S.J., 2015. Operational tree species mapping in a diverse tropical forest with airborne imaging spectroscopy. PLoS One 10 (7), e0118403.

Brooks, R.R., 1972. Geobotany and Biogeochemistry in Mineral Exploration. Harper & Row, New York (290p).

Brooks, R.R., 1983. Biological Methods of Prospecting Minerals, 2nd ed. John Wiley & Sons, New York (322p).

Cannon, H.L., 1960. Botanical prospecting for ore deposits. Science 132 (3427), 591–598.

Cannon, H.L., 1971. The use of plant indicators in ground water surveys, geologic mapping, and mineral prospecting. Taxon 227–256.

Carranza, E.J.M., Hale, M., 2002. Mineral imaging with Landsat Thematic Mapper data for hydrothermal alteration mapping in heavily vegetated terrane. Int. J. Remote Sens. 23 (22), 4827–4852.

Castro-Esau, K.L., Kalacska, M., 2008. Tropical dry forest phenology and discrimination of tropical tree species using hyperspectral data. In: Kalacska, M., Sanchez-Azofeifa, G.A. (Eds.), Hyperspectral Remote Sensing of Tropical and Subtropical Forests. Taylor & Francis Group, New York, pp. 01–25 (Orgs.).

Castro-Esau, K., Sánchez-Azofeifa, G.A., Caelli, T., 2004. Discrimination of lianas and trees with leaf-level hyperspectral data. Remote Sens. Environ. 90 (3), 353–372.

Castro-Esau, K.L., Sánchez-Azofeifa, G.A., Rivard, B., Wright, S.J., Quesada, M., 2006. Variability in leaf optical properties of Mesoamerican trees and the potential for species classification. Am. J. Bot. 93 (4), 517–530.

Clark, M.L., 2005. An assessment of Hyperspectral and Lidar Remote Sensing for the Monitoring of Tropical Rain Forest Trees (PhD. Thesis). Department of Geography, University of California, Santa Barbara (UMI No. 3186829).

Clark, R.N., Roush, T.L., 1984. Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. J. Geophys. Res. 89, 6329–6340.

Clark, M.L., Roberts, D.A., Clark, D.B., 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. Remote Sens. Environ. 96, 375–398.

Cochrane, M.A., 2000. Using vegetation reflectance variability for species level classification of hyperspectral data. Int. J. Remote Sens. 21 (10), 2075–2087.

Colgan, M.S., Baldeck, C.A., Féret, J.B., Asner, G.P., 2012. Mapping savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data. Remote Sens. 4 (11), 3462–3480.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37 (1), 35–46.

Congalton, R.G., Green, K., 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, 1st ed. CRC press, Boca Raton.

CPRM - Brazilian Geological Service, 2006. Geologia e recursos minerais do estado de São Paulo: Sistema de Informações Geográficas (SIG) (Geology and mineral resources of the São Paulo state: Geographical Information System (GIS)). State Geological Maps. Scale 1:750.000. Programa Geologia do Brasil [Geology of Brazil Program] CPRM, São Paulo 1 CD-ROM.

Dennison, P.E., Roberts, D.A., 2003. Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. Remote Sens. Environ. 87, 123–135.

Dennison, P.E., Halligan, K.Q., Roberts, D.A., 2004. A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. Remote Sens. Environ. 93, 359–367.

Eiten, G., 1963. Habitat flora of Fazenda Campininha. In: Proceedings of the 1st Symposium on Cerrado. USP, São Paulo, pp. 189-23.

Féret, J.B., Asner, G.P., 2011. Spectroscopic classification of tropical forest species using radiative transfer modeling. Remote Sens. Environ. 115, 2415–2422.

Féret, J., Asner, G.P., 2013. Tree species discrimination in tropical forests usin g airborne imaging spectroscopy. IEEE Trans. Geosci. Remote Sens. 51 (1), 73–84.

Ferreira, M.P., Grondona, A.E.B., Rolim, S.B.A., Shimabukuro, Y.E., 2013. Analyzing the spectral variability of tropical tree species using hyperspectral feature selection and leaf optical modeling. J. Appl. Remote. Sens. 7 (1) (073502-073502).

Ferreira, M.P., Zortea, M., Zanotta, D.C., Shimabukuro, Y.E., de Souza Filho, C.R., 2016. Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. Remote Sens. Environ. 179, 66–78.

Fúlfaro, J.V., Bjomberg, A.J.S., 1993. Geologia [Geology]. In: Cintra, J.C.A., Albiero, J.H. (Eds.), Solos do interior de São Paulo [Soils of São Paulo State countryside]. Associação Brasileira de Mecânica dos Solos e Engenharia Geotécnica/Universidade de São Paulo, São Carlos, pp. 01–42.

Gamon, J.A., 2008. Tropical remote sensing – opportunities and challenges. In: Kalacska, M., Sanchez-Azofeifa, G.A. (Eds.), Hyperspectral remote sensing of tropical and subtropical forests. Taylor & Francis Group, New York, pp. 297–304 (Orgs.).

Gamon, J.A., Kitajima, K., Mulkey, S.S., Serrano, L., Wright, S.J., 2005. Diverse optical and photosynthetic properties in a neotropical dry Forest during the dry season: implications for remote estimation of photosynthesis. Biotropica 37 (4), 547–560.

Hede, A.N.H., Kashiwaya, K., Koike, K., Sakurai, S., 2015. A new vegetation index for detecting vegetation anomalies due to mineral deposits with application to a tropical forest area. Remote Sens. Environ. 171, 83–97.

Hesketh, M., Sánchez-Azofeifa, G.A., 2012. The effect of seasonal spectral variation on species classification in the Panamanian tropical forest. Remote Sens. Environ. 118, 73–82.

Higgins, M.A., Ruokolainen, K., Tuomisto, H., Llerena, N., Cardenas, G., Phillips, O.L., Vásquez, R., Räsänen, M., 2011. Geological control of floristic composition in Amazonian forests. J. Biogeogr. 38 (11), 2136–2149.

Higgins, M.A., Asner, G.P., Perez, E., Elespuru, N., Tuomisto, H., Ruokolainen, K., Alonso, A., 2012. Use of Landsat and SRTM data to detect broad-scale biodiversity patterns in Northwestern Amazonia. Remote Sens. 4 (8), 2401–2418.

Higgins, M.A., Asner, G.P., Perez, E., Elespuru, N., Alonso, A., 2014a. Variation in photosynthetic and nonphotosynthetic vegetation along edaphic and compositional gradients in northwestern Amazonia. Biogeosciences 11 (3), 3505–3513.

Higgins, M.A., Asner, G.P., Martin, R.E., Knapp, D.E., Anderson, C., Kennedy-Bowdoin, T.,

Saenz, R., Aguilar, A., Wright, S.J., 2014b. Linking imaging spectroscopy and LiDAR with floristic composition and forest structure in Panama. Remote Sens. Environ. 154, 358–367.
Hill, R.A., Wilson, A.K., George, M., Hinsley, S.A., 2010. Mapping tree species in temperate deciduous woodland using time-series multi-spectral data. Appl. Veg. Sci. 13 (1), 86–99.
Holm, G., 1954. Chlorophyll mutations in barley. Acta Agric. Scand. 4, 457–461.
Joly, C.A., Bicudo, C.E.M., 1999. Biodiversidade Do Estado de São Paulo, Brasil: síntese Do Conhecimento Ao Final Do século XX – Vol. 7: Infra-Estrutura Para conservação da Biodiversidade [Biodiversity of the State of São Paulo, Brazil: Summary of the Knowledge at the End of the 20th Century – Vol 7: Infrastructure for Biodiversity Conservation], 1st ed. FAPESP, São Paulo.
Kalacska, M., Bohlman, S., Sanchez-Azofeifa, G.A., Castro-Esau, K., Caelli, T., 2007. Hyperspectral discrimination of tropical dry forest lianas and trees: comparative data reduction approaches at the leaf and canopy levels. Remote Sens. Environ. 109, 406–415.
Kohonen, T., 1998. The self-organizing map. Neurocomputing 21, 01–06.
Kokaly, R.F., 2011. PRISM: Processing routines in IDL for spectroscopic measurements. Retrieved from. http://pubs.usgs.gov/of/2011/1155/.
Kruckeberg, A.R., 2002. Geology and Plant Life: The Effects of Landforms and Rock Types on Plants. University of Washington Press, Seatle (360p).
Lammoglia, T., de Souza Filho, C.R., 2013. Unraveling hydrocarbon microseepages in onshore basins using spectral–spatial processing of advanced spaceborne thermal emission and reflection radiometer (ASTER) data. Surv. Geophys. 34 (3), 349–373.
Lavorel, S., McIntyre, S., Landsberg, J., Forbes, T.D.A., 1997. Plant functional classifications: from general groups to specific groups based on response to disturbance. Trends Ecol. Evol. 12 (12), 474–478.
Lenza, E., Klink, C.A., 2006. Comportamento fenológico de espécies lenhosas em um cerrado sentido restrito de Brasília, DF [Phenological behavior of woody species in a *stricto sensu* cerrado of Brasília, DF]. Rev. Bras. Bot. 29 (4), 627–638.
Lyon, R.J.P., Lee, K., 1970. Remote sensing in exploration for mineral deposits. Econ. Geol. 65 (7), 785–800.
Madonsela, S., Cho, M.A., Mathieu, R., Mutanga, O., Ramoelo, A., Kaszta, Ż., Kerchove, R.V.D., Wolff, E., 2017. Multi-phenology WorldView-2 imagery improves remote sensing of savannah tree species. Int. J. Appl. Earth Obs. Geoinf. 58, 65–73.
Madritch, M.D., Kingdon, C.C., Singh, A., Mock, K.E., Lindroth, R.L., Townsend, P.A., 2014. Imaging spectroscopy links aspen genotype with below-ground processes at landscape scales. Philos. Trans. R. Soc. Lond. B 369, 20130194.
McManus, K.M., Asner, G.P., Martin, R.E., Dexter, K.G., Kress, W.J., Field, C.B., 2016. Phylogenetic structure of foliar spectral traits in tropical forest canopies. Remote Sens. 8 (3), 196.
Prasad, E.A.V., 1987. Geobotany and biogeochemistry in mineral exploration in the tropics. J. Geochem. Explor. 29, 427–428.
Prospere, K., McLaren, K., Wilson, B., 2014. Plant species discrimination in a tropical wetland using in situ hyperspectral data. Remote Sens. 6 (9), 8494–8523.
Reich, P.B., Wright, I.J., Cavender-Bares, J., Craine, J.M., Oleksyn, J., Westoby, M., Walters, M.B., 2003. The evolution of plant functional variation: traits, spectra, and strategies. Int. J. Plant Sci. 164 (S3), S143–S164.
Ribeiro, J.F., Walter, B.M.T., 2008. As principais fitofisionomias do Bioma Cerrado [The main plant physiognomies of the Cerrado Biome]. In: Sano, S.M., Almeida, S.P., Ribeiro, J.F. (Eds.), Cerrado: ecologia e flora [Cerrado: Ecology and Flora]. EMBRAPA-CPAC, Planaltina, pp. 153–212.
Rivard, B., Sánchez-Azofeifa, G.A., Foley, S., Calvo-Alvarado, J.C., 2008. Species classification of Tropical tree leaf reflectance and dependence on selection of spectral bands. In: Kalacska, M., Sanchez-Azofeifa, G.A. (Eds.), Hyperspectral Remote Sensing of Tropical and Subtropical Forests. Taylor & Francis Group, New York, pp. 141–159 (Orgs.).
Roberts, D.A., Gardner, M., Church, R., Ustin, S., Scheer, G., Green, R.O., 1998. Mapping Chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. Remote Sens. Environ. 65, 267–279.
Roberts, D.A., Dennison, P.E., Gardner, M., Hetzel, Y.L., Ustin, S.L., Lee, C., 2003. Evaluation of the potential of hyperion for fire danger assessment by comparison to the airborne visible infrared imaging spectrometer. IEEE Trans. Geosci. Remote Sens. 41, 1297–1310.
Roberts, D.A., Halligan, K.Q., Dennison, P.E., 2007. ViperTools. Retrieved from. http://www.vipertools.org/.
Roberts, D.A., Quattrochi, D.A., Hulley, G.C., Hook, S.J., Green, R.O., 2012. Synergies between VSWIR and TIR data for the urban environment: an evaluation of the potential for the Hyperspectral Infrared Imager (HyspIRI) Decadal Survey mission. Remote Sens. Environ. 117, 83–101.
Roth, K.L., Dennison, P.E., Roberts, D.A., 2012. Comparing endmember selection techniques for accurate mapping of plant species and land cover using imaging spectrometer data. Remote Sens. Environ. 127, 139–152.
Roth, K.L., Casas, A., Huesca, M., Ustin, S.L., Alsina, M.M., Mathews, S.A., Whiting, M.L., 2016. Leaf spectral clusters as potential optical leaf functional types within California ecosystems. Remote Sens. Environ. 184, 229–246.
Sabins, F.F., 1999. Remote sensing for mineral exploration. Ore Geol. Rev. 14 (3), 157–183.
Sanches, I.D., Souza Filho, C.R., Magalhães, L.A., Quitério, G.C.M., Alves, M.N., Oliveira, W.J., 2013. Assessing the impact of hydrocarbon leakages on vegetation using reflectance spectroscopy. ISPRS J. Photogramm. Remote Sens. 78, 85–101.
Sánchez-Azofeifa, G.A., Castro, K., Wright, S.J., Gamon, J., Kalacska, M., Rivard, B., et al., 2009. Differences in leaf traits, leaf internal structure, and spectral reflectance between two communities of lianas and trees: implications for remote sensing in tropical environments. Remote Sens. Environ. 113 (10), 2076–2088.
Santos, D.L.D., Takaki, M., 2005. Fenologia de *Cedrela fissilis* Vell.(Meliaceae) na região rural de Itirapina, SP, Brasil [Phenology of *Cedrela fissilis* Vell. (Meliaceae) in the rural region of Itirapina, SP, Brazil]. Acta Bot. Bras. 19 (3), 625–632.
Schaaf, A.N., Dennison, P.E., Fryer, G.K., Roth, K.L., Roberts, D.A., 2011. Mapping plant functional types at three spatial resolutions using multiple endmember spectral mixture analysis. GISci. Remote Sens. 48, 324–344.
Silvério, D.V., Lenza, E., 2010. Fenologia de espécies lenhosas em um cerrado típico no Parque Municipal do Bacaba, Nova Xavantina, Mato Grosso, Brasil [Phenology of woody species in a typical cerrado in Parque Municipal do Bacaba, Nova Xavantina, Mato Grosso, Brazil]. Biota Neotropica 10 (3), 205–216.
Sirén, A., Tuomisto, H., Navarrete, H., 2013. Mapping environmental variation in lowland Amazonian rainforests using remote sensing and floristic data. Int. J. Remote Sens. 34 (5), 1561–1575.
Smith, M.O., Ustin, S.L., Adams, J.B., Gillespie, A.R., 1990. Vegetation in deserts: I. A regional measure of abundance from multispectral images. Remote Sens. Environ. 31 (1), 1–26.
Somers, B., Asner, G.P., 2013. Multi-temporal hyperspectral mixture analysis and feature selection for invasive species mapping in rainforests. Remote Sens. Environ. 136, 14–27.
Somers, B., Asner, G.P., 2014. Tree species mapping in tropical forests using multi-temporal imaging spectroscopy: wavelength adaptive spectral mixture analysis. Int. J. Appl. Earth Obs. Geoinf. 31, 57–66.
Souza Filho, C.R., Augusto, V., Oliveira, W.J., Lammoglia, T., 2008. Detecção de exsudações de hidrocarbonetos por geobotânica e sensoriamento remoto multi-temporal: estudo de caso no Remanso do Fogo (MG) [Detection of hydrocarbon exudates by geobotany and multi-temporal remote sensing: a study case in Remanso do Fogo (MG)]. Rev. Bras. Geocienc. 38 (2), 228–243.
Updegraff, D.M., 1969. Semimicro determination of cellulose in biological materials. Anal. Biochem. 32 (3), 420–424.
Ustin, S.L., Gamon, J.A., 2010. Remote sensing of plant functional types. New Phytol. 186 (4), 795–816.
Ustin, S.L., Smith, M.O., Jacquemoud, S., Verstraete, M.M., Govaerts, Y., 1999. Geobotany: vegetation mapping for Earth Sciences. In: Rencz, A.N. (Ed.), Remote sensing for the Earth Sciences, 3rd ed. John Wiley and sons, New York, pp. 189–233.
Vitousek, P., Asner, G.P., Chadwick, O.A., Hotchkiss, S., 2009. Landscape-level variation in forest structure and biogeochemistry across a substrate age gradient in Hawaii. Ecology 90 (11), 3074–3086.
Wang, F., Gao, J., Zha, Y., 2018. Hyperspectral sensing of heavy metals in soil and vegetation: feasibility and challenges. ISPRS J. Photogramm. Remote Sens. 136, 73–84.
Youngentob, K.N., Roberts, D.A., Held, A.A., Dennison, P.E., Jia, X., Lindenmayer, D.B., 2011. Mapping two *Eucalyptus* subgenera using multiple endmember spectral mixture analysis and continuum-removed imaging spectrometry data. Remote Sens. Environ. 115 (5), 1115–1128.