ORIGINAL

CrossMark

Potential field-deployable NIRS identification of seven *Dalbergia* species listed by CITES

Filipe A. Snel[1,2] · Jez W. B. Braga[1] · Diego da Silva[1,2] · Alex C. Wiedenhoeft[3,4,5,6] · Adriana Costa[3,4] · Richard Soares[3] · Vera T. R. Coradin[2] · Tereza C. M. Pastore[2]

Abstract

Near-infrared spectroscopy (NIRS) is a potential, field-portable wood identification tool. NIRS has been studied as tool to identify some woods but has not been tested for *Dalbergia*. This study explored the efficacy of hand-held NIRS technology to discriminate, using multivariate analysis, the spectra of some high-value *Dalbergia* wood species: *D. decipularis*, *D. sissoo*, *D. stevensonii*, *D. latifolia*, *D. retusa,* all of which are listed in CITES Appendix II, and *D. nigra*, which is listed in CITES Appendix I. Identification models developed using partial least squares discriminant analysis (PLS-DA) and soft independent modeling by class analogy (SIMCA) were compared regarding their ability to answer two sets of identification questions. The first is the identification of each *Dalbergia* species among the group of the six above, and the second is the separation of *D. nigra* from a single group comprising the other species, grouping all *Dalbergia* as one class. For this latter study, spectra of *D. cearensis* and *D. tucurensis* were added to the broader *Dalbergia* class. These spectra were not included in the first set because the number of specimens was not enough to create an exclusive class for them. PLS-DA presented efficiency rates of over 90% in both situations, while SIMCA presented 52% efficiency at species-level separation and 85% efficiency separating *D. nigra* from other *Dalbergia*. It was shown that PLS-DA approaches are far better suited than SIMCA for generating a field-deployable NIRS model for discriminating these *Dalbergia*.

✉ Tereza C. M. Pastore
  tereza.pastore@florestal.gov.br

Extended author information available on the last page of the article

🖄 Springer

Introduction

Brazil is home to one of the most diverse and most threatened biomes in the world, the Atlantic Forest (Stehmann et al. 2009; Myers et al. 2000), and the rare *Dalbergia nigra* (Vell.) Benth. is the flagship threatened species of this forest, now only rarely found across its former range from southern Bahia to northern São Paulo. The wood of *D. nigra* and other 303 *Dalbergia* species is distributed worldwide, 39 species in Brazil, including trees, shrubs and vines (The Plant List 2013; Carvalho 1997). *D. nigra* is also known as Brazilian rosewood, jacarandá-da-Bahia, caviúna, palisandre and jaracandá-preto (Carvalho 1997) and is widely used for furniture and musical instruments (Lorenzi 1992), but it has been overexploited, leading to listing in Appendix I of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). *D. nigra* was the first tree species ever included in this appendix, though other *Dalbergia* species were listed in Appendix II and III (CITES 1992). In an effort to improve global protection for the genus and recognition of the difficulty in identifying *Dalbergia* woods at the species level, all species of *Dalbergia* were included in CITES Appendix II except for *D. nigra*, which remained in Appendix I (CITES 2017).

Seized logs normally do not retain characteristic morphological features anymore such as leaves, seeds and flowers making it almost impossible to identify the species and the geographic origin (Espinoza et al. 2015). The difficulty in identifying specimens at checkpoints facilitates the illegal timber trade, expands deforested areas and destroys local communities (UNODC 2016).

Although many species are threatened with extinction and their exploitation is prohibited, it is still possible to find them in the illegal market. One of the causes of the overexploitation is government inability to control wood trade. The most demanding problem is definitive identification of species at checkpoints (UNODC 2016). Several methodologies are being studied to find the most suitable tool capable of identifying specimens at species level quickly and in situ.

Species-level separation within *Dalbergia* has long been desirable, but with a few exceptions it is not possible by wood anatomy alone (Gasson et al. 2010; Wiemann and Ruffinatto 2012). Among the Latin-American species of *Dalbergia*, anatomically, *D. cearensis* Ducke is easily identifiable because it is the only one having a vessel frequency of over ten vessels mm^{-2} (Miller and Wiemann 2006; Gasson et al. 2010).

In 2006, Miller and Wiemann suggested that it is possible to distinguish *D. nigra* and *D. spruceana* Benth. using water and ethanol fluorescence of heartwood extracts, although the same method could not differentiate *D. nigra*, *D. granadillo* Pittier and *D. stevensonii* Standl. due to the similarity in the density and fluorescence characteristics of these species (Guzman et al. 2008). Since *Dalbergia* species contain a variety of phenolic compounds, Kite et al. (2010) suggested that the concentration of these compounds may vary among the species and provide a way to distinguish between them.

Lancaster and Espinoza (2012) reported on the discrimination of 13 species, including 10 of *Dalbergia*, using direct analysis in real time coupled with

time-of-flight (DART-TOF) mass spectrometer. In this work, the identity of 15 *D. nigra* specimens was confirmed. The methodology was able to differentiate specimens of *D. nigra* and of *D. spruceana*, but the unique compound present in *D. nigra*, dalnigrin, could not be distinguished from its isomer, kuhlmannin, present in *D. spruceana* (Kite et al. 2010). In 2015, Espinoza et al. reported that *D. nigra* could be distinguished from look-alike species such as *D. spruceana, D. stevensonii* and *D. tucurensis* Donn. Sm, also using DART-TOF.

Near-infrared spectroscopy (NIRS) when associated with multivariate analyses is an analytical methodology that brings information on high-weight chemical groups present in cellulose, hemicellulose, lignin and extractives molecules (resin, gums, etc.) in wood. However, due to the overlapped nature of the NIR spectra, this methodology does not allow to specify individual bands related to specific chemical bonds or compounds in complex samples (i.e., wood) (Schwanninger et al. 2011), which justify its association with chemometric methods. The main NIRS advantages are: it requires minimum sample preparation and has fast response as diffuse reflectance is used (Pasquini 2003; Burns and Ciurczak 2007). NIRS combined with multivariate analysis is a powerful tool for classifying and quantifying a wide variety of materials such as identification of plant part composition of forest logging and discrimination of *Fagales* wood samples from *Malpighiales* (Acquah et al. 2016; Carballo-Meilan et al. 2014). Several NIRS models have been created and improved to identify wood species. Adedipe et al. (2008) used NIRS and soft independent modeling of class analogies (SIMCA) to classify red oak and white oak species. Pastore et al. (2011) and Braga et al. (2011) have proposed partial least squares discriminant analysis (PLS-DA) identification methods to discriminate *Swietenia macrophylla* King (mahogany or bigleaf mahogany) from three look-alike species: *Carapa guianensis* Aubl., *Cedrela odorata* L. and *Micropholis melinoniana* Pierre, and more recent works added additional species (Soares et al. 2017) and mahogany's geographic origin (Bergo et al. 2016). Lazarescu et al. (2016) could distinguish western hemlock (*Tsuga heterophylla* (Raf.) Sarg. from amabilis fir (*Abies amabilis* (Douglas ex Loudon) J. Forbes) using near-infrared scans and analyzing the data with PLS-DA and artificial neural networks (ANN). Despite the richness of this previous literature, *Dalbergia* has not been studied using NIRS for wood identification.

The principal aim of this work is to demonstrate the possibility of the identification of *Dalbergia* species based on NIRS spectra obtained with a hand-held spectrometer. The spectrometer weighs 1.2 kg, and it is simple to manipulate. Enforcement agents can be quickly trained to acquire the spectra from the suspicious sample and read the results at checkpoints. In addition, two multivariate methods, PLS-DA and SIMCA, were evaluated to develop identification models and two approaches will be compared: (1) separate the wood of six *Dalbergia* at the species level and (2) separate the CITES Appendix I *D. nigra* from a combined class of the other CITES Appendix II *Dalbergia,* without attempting species-level separation within the Appendix II group. Providing NIRS-based identification models deployable on a hand-held spectrophotometer for *Dalbergia* in general and *D. nigra* specifically improves global capacity to enforce CITES regulations and can empower law enforcement to make field-level forensic evaluations in real time.

Materials and methods

Sample preparation

In this work, eight species of *Dalbergia* were studied. Spectra were obtained from specimens at the Forest Products Laboratory, Madison, USA, and the Forest Products Laboratory, Brasilia, Brazil (Table 1).

For each sample, a randomly selected surface of the heartwood with no apparent defect was sanded with grit 80. As the diffused light varies with surface roughness, this procedure homogenizes the surface. Spectra were taken only after carefully removing sanding dust from the surface to be measured. This procedure embeds wood surface information (the face variable, see Braga et al. 2011) in the species discrimination model so that any face can be tested in the field. A sample surface area has to be at least 1.2 cm×1.2 cm, the same area from which the diffuse reflectance light comes.

Near-Infrared Spectroscopy measurements

The near-infrared spectra were collected with the microPHAZIR™ RX Analyzer, a hand-held spectrometer, over the range of 1595–2396 nm with a resolution of 8 nm. For each one of the 249 samples, three spectra in different points were collected, totalizing 747 spectra not averaged. The *D. cearensis* and *D. tucurensis* specimens were not used in the species-level discrimination study because the number of specimens was not enough to build a robust model for them (Pastore et al. 2011; Braga et al. 2011), but they were used to build the general model for *Dalbergia nigra* discrimination.

To enable the appropriate data analysis, the spectra data were transferred from the device to a laptop, and data analysis was carried out in MATLAB R2011 and PLS_Toolbox 7.03. After model optimization, the conditions established in the PLS_Toolbox were reproduced in the equipment's software (Thermo microP-HAZIR Admin, version 4.0219) and the resulted model was transferred back to the spectrometer.

Table 1 Data of *Dalbergia* specimens according to species

Species	No. of specimens/spectra	Origin
D. cearensis Ducke	6/18	Brazil
D. tucurensis Donn. Sm.	9/27	Guatemala
D. decipularis Rizzini & A. Mattos	16/48	Brazil
D. sissoo DC.	19/57	India
D. stevensonii Standl.	14/42	Belize
D. latifolia Roxb.	19/57	Indonesia
D. retusa Hemsl.	22/66	Panama
D. nigra (Vell.) Benth	144/432	Brazil

Data preprocessing

Prior to the development of the discrimination or classification models, it is important to preprocess the data because the near-infrared spectrum contains redundant information that pollutes the usable chemical information. Through preprocessing, variations regarding baseline shifts due to additive and multiplicative scattering can be eliminated or at least minimized (Wise et al. 2006), keeping the pertinent chemical information for species identification. According to the previous works of this research group, all the data were preprocessed by the combination of first derivative using Savitzky–Golay smoothing (using an eleven-point filter width and second-order polynomial), standard normal variate (SNV) and mean centering.

Chemometric modeling

The identification of the *Dalbergia* species was evaluated by two distinct chemometric approaches. The first one was based on SIMCA, which is one of the most applied classification models. The second approach was a discrimination model by PLS-DA algorithm. Both models have already been described in detail (Soares et al. 2017; Martins et al. 2017; De Maesschalck et al. 1999; Flåten et al. 2004; Branden and Hubert 2005). Therefore, in this work just a brief description of their properties and differences will be presented. However, as outlier detection plays an important role in model development and validation, it is discussed in more detail in the Electronic Supplementary Material.

Soft independent modeling by class analogy (SIMCA)

SIMCA is a classification method that is based on the class modeling approach using principal components analysis (PCA). Initially, a PCA model is developed for each class separately to define the confidence regions in the score space of PCA, considering the variability of the samples in the training set (Branden and Hubert 2005; De Maesschalck et al. 1999). The number of principal components of the PCA model of each class is usually determined by cross-validation and chosen as the one that minimizes the classification errors in the training set. During the validation, an unknown specimen is projected in the PCA model of each class and the distance (d) is determined by Eq. 1. The sample is assigned to the class that presents the lower d. Otherwise, if the sample exceeds the critical value for d in all classes, it is classified as an outlier or not belonging to any predefined class. The distance (d) in relation to the PCA models of the predefined classes, wherein $d_{i,w}$ is related to a sample i and class w, is determined as:

$$d_{i,w} = \sqrt{\left(\frac{T_{i,w}^2}{T_{\text{critical},w}^2}\right)^2 + \left(\frac{Q_{i,w}}{Q_{\text{critical},w}}\right)^2} \qquad (1)$$

where $T_{i,w}^2$ is the Hotelling statistics, $Q_{i,w}$ is sum of the squares of the PCA residuals, $T_{\text{critical},w}^2$ and $Q_{\text{critical},w}$ are the corresponding 99.9% confidence level thresholds of T^2

and Q for the class w, respectively. More details of the determination of the T^2 and Q and their critical values can be found in the Electronic Supplementary Material (Ballabio and Todeschini 2009).

Partial least squares for discrimination analysis (PLS-DA)

In contrast to the SIMCA model, in PLS-DA the NIR spectra of all the *Dalbergia* species (classes) in the training set are organized in a single $\mathbf{X}$(I,M) matrix, where I represents the total number of spectra in $\mathbf{X}$ and M accounts for the number of wavelengths of the NIR spectrum. The discrimination model is developed by the correlation of this spectral matrix $\mathbf{X}$ with a $\mathbf{y}$(I,1) vector that indicates if the sample belongs to the species to be discriminated—class value of 1—or the other species—class value of 0. As a result, the PLS-DA models perform a binary discrimination, so that a total of six models were built to discriminate the studied species. In each model, two-thirds of samples were randomly selected to be part of the training group. The remaining samples were used as the validation group; these are the samples that will test the discrimination model created according to the training samples. The three spectra of each sample are only in one of the groups, either calibration or validation, so the spectra of the same sample are not validated against themselves. The number of latent variables A was determined through cross-validation according to Soares et al. (2017), who suggested the use of the root mean square error of cross-validation (RMSECV) for the minimization of the errors in the class value estimate and which provide a larger separation between the estimated values for the classes in the training (Soares et al. 2017; Martins et al. 2017).

PLS-DA identification of outliers was performed based on the Hotelling T^2, Q residuals and the confidence limits for the estimation of the estimated class values, as described by Soares et al. (2017) and detailed in the Electronic Supplementary Material.

After the model optimization, a discrimination threshold (DT) was estimated based on the dispersion of the estimated class values obtained for the training set in order to minimize the occurrence of false positive and false negative errors (Botelho et al. 2015; Soares et al. 2017). A detailed description of the determination of the DT can be found in Soares et al. (2017). It is important to note that when the number of specimens/spectra is significantly different between the species/classes, the probabilities of occurrence will be different in the training set (a priori probabilities), which can introduce a bias in the discrimination threshold and influence the discrimination. Given that the differences in the number of specimens observed in Table 1 express a limitation of obtaining, approximately, the same number of specimens for all species and not a result of their probabilities of occurrence, the determination of the discrimination threshold was performed using all classes with the same probabilities of occurrence, as detailed by Soares et al. (2017).

The identification of the class of a validation/unknown sample is determined only after the analysis in all PLS-DA models. Therefore, a hypothetic sample i is identified as belonging to a class 1 only if its estimated class value was larger than the DT when the sample is analyzed in the PLS-DA model for class 1 and simultaneously present class values lower than the DT for the models of the other classes.

Model performance metrics

To check the model performance, the samples from the validation set were used for the determination of the figures of merit of the models. Considering a hypothetic class *w*, these figures of merit can be resumed in five different parameters:

(FNR) false negative rate, which is the percentage of the samples of the class *w* identified as belonging to other classes;
(FPR) false positive rate, the percentage of the samples of other classes identified as belonging to the class *w*;
(STR) sensitivity rate, the percentage of samples belonging to *w* correctly identified as *w*;
(SPR) specificity rate, the percentage of samples belonging to the other classes correctly identified the other classes; and
(EFR) efficiency rate, estimated as 100% minus FNR and FPR.

Details of these calculations can be found in previous papers of this research group (Soares et al. 2017; Martins et al. 2017).

Results and discussion

Figure 1 presents the NIR spectra of the training samples before and after the preprocessing by the first derivative and SNV. In the raw spectra (Fig. 1-a), one can easily see the significant variation of the baseline, caused by the presence of additive and multiplicative scattering. This variation is mostly a result of physical effects due to the difference in the sample surface. After the preprocessing, the different levels of scattering were corrected or at least minimized, leading to a sharper line (Fig. 1-b). Visible differences between the samples of preprocessed spectra can be observed in the following regions of NIRS: 1640–1760, 1870–1910 and 2080–2160 nm. Most of these variations would be a result of the differences in the chemical composition between the species. However, due to the overlap between the bands in the NIR spectra, it is very difficult to infer about the chemical compounds responsible for these differences and assimilate these variations in NIR spectra to specific species, justifying the necessity of the multivariate analysis.

Species-level discrimination

It was sought to separate *D. decipularis, D. sissoo, D. stevensonii, D. latifolia, D. retusa* and *D. nigra* at the species level using PLS-DA and SIMCA models. As previously mentioned, due to the small number of specimens the *D. cearensis* and *D. tucurensis* species were not included in this analysis.
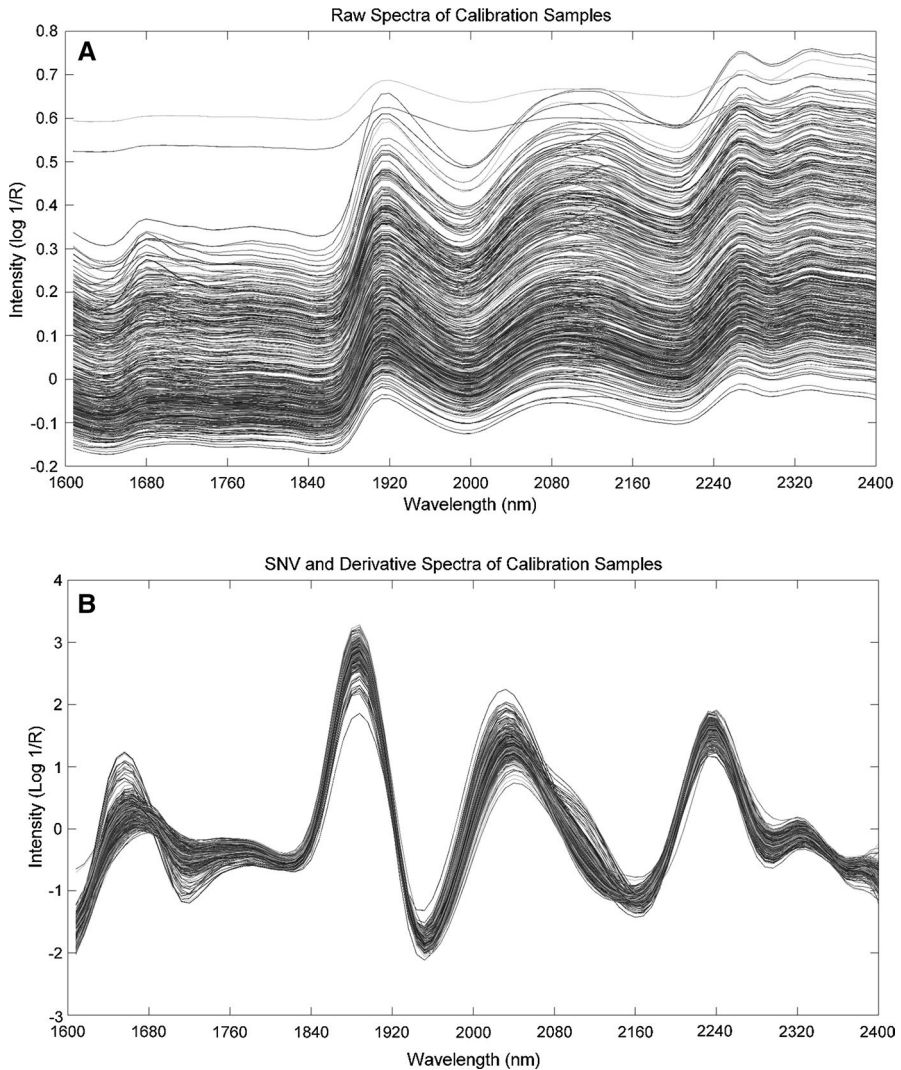
Raw Spectra of Calibration Samples

SNV and Derivative Spectra of Calibration Samples

Fig. 1 Spectral data of **a** raw spectra and **b** SNV and derivative spectra

Partial least square discrimination analysis: PLS-DA

In Table 2, the number of latent variables necessary to discriminate each species in relation to the others is presented. The explained variance in the $\mathbf{X}$ matrix and the $\mathbf{y}$ vector with these numbers of latent variables were 99.88%—*D. retusa*, 99.02%—*D. stevensonii*, 99.92—*D. sissoo*, 98.83%—*D. nigra*, 99.87%—*D. latifolia* and 99.92%—*D. decipularis*. In general, except for *D nigra* and *D stevensonii*, it was observed that a high number of latent variables (> 18) were required for four species, which might be related to the spectral variation indicating that samples of these

Table 2 Results for the number of latent variables and outliers identified in each of the PLS-DA models for six *Dalbergia* species

Species	*No.* of latent variables	Outliers excluded in the training set	Outliers excluded in validation set
D. nigra	12	7	5
D. retusa	19	8	4
D. sissoo	21	5	4
D. latifolia	19	5	5
D. decipularis	20	22	13
D. stevensonii	13	10	6

species are more similar among them and/or high variation in spectral information within the same species, justifying the use of more latent variables. Information about the amount of explained variance in the first 10 latent variables is shown in Table S1 in the Electronic Supplementary Material. In addition, Table 2 shows the number of spectra excluded as outliers in classes 1 or 0 in both training and validation phases. The percentage of outlier identification varies from 1.0 to 5.3%, being that the model with the largest occurrence of outliers was the discrimination of *D. decipularis*, even though this percentage of outliers can be considered acceptable. In addition, the identification of a sample as an outlier does not mean a wrong result; it represents an inconclusive identification, showing that this specimen presents some characteristics that make it distinct from the ones included in the training set. Therefore, an alternative analysis should be performed to confirm the identification of the species of this specimen.

Figure 2 shows the discrimination between the target species (class value = 1) from the other species (class value = 0) in each discrimination model. In Fig. 2, the validation outliers are represented but they were excluded from the figures of merit determination, which are presented in Table 3. Figure 2a shows the discrimination model of *D. nigra*, and Fig. 2b shows the discrimination model of *D. latifolia.* In Fig. 2 each point represents one spectrum.

The *D. nigra* model (Fig. 2-a) shows the highest efficiency rate (EFR = 98.6%). Each inverted triangle represents one spectra of *D. nigra*, and the inverted triangle cluster in the top left represents the *D. nigra* specimens used to calibrate the model. The inverted triangle cluster in the right is the spectra of *D. nigra* used in the validation. The top and bottom lines represent the limit of y value. Specimens with y value above the top line and below the bottom line are considered outliers. The estimated class value near 0.5 line represents the discrimination limit. According to the model, the specimens above the discrimination limit are *D. nigra* and the others are not *D. nigra*. The two inverted triangles below the discrimination limit in the validation samples represent the two false negative specimens, specimens that are *D. nigra* but the model had classified as not. The *D. nigra* model does not show false positives and correctly discriminated 221 specimens.

Even the models with the lowest efficiency rates—*D. latifolia* and *D. retusa*—show a high correct classification rate—certainly high enough to provide meaningful and reliable real-time feedback to law enforcement in a field-screening context.

Fig. 2 Discrimination models of *Dalbergia* specimens.—**a**: *D. nigra*.—**b**: *D. latifolia*. Upside down triangle: *D. nigra,* asterisk: *D. retusa*, square: *D. sissoo*, empty circle: *D. latifolia*, lozenge: *D. decipularis*, filled circle: *D. stevensonii*
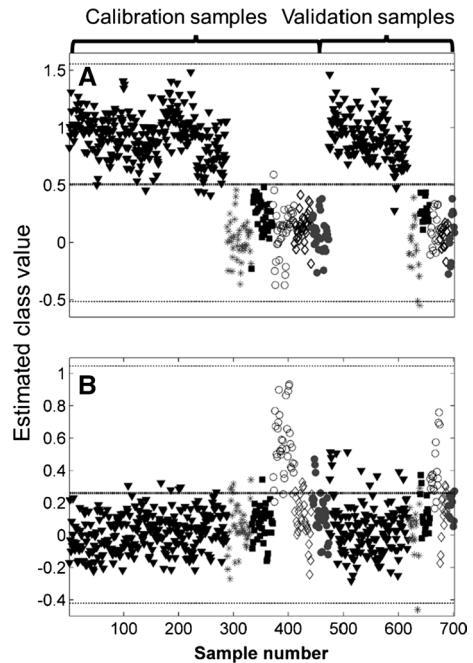

Table 3 Figures of merit of the six *Dalbergia* PLS-DA models

Figures of merit	Species					
	D. nigra	D. retusa	D. sissoo	D. latifolia	D. decipularis	D. stevensonii
FP[1]	0	0	5	19	5	7
FN[2]	2	3	0	2	0	0
TP[3]	142	14	18	16	14	12
TN[4]	79	207	201	186	196	203
FPR[5] (%)	0.0	0.0	2.4	9.3	2.5	3.3
FNR[6] (%)	1.4	17.6	0.0	11.1	0.0	0.0
SPR[7] (%)	100	100	97.6	90.7	97.5	96.7
STR[8] (%)	98.6	82.4	100.0	88.9	100.0	100.0
EFR[9] (%)	98.6	82.3	97.6	79.6	97.5	96.7

[1]*FP* number of false positives, [2]*FN* number of false negatives, [3]*TP* number of true positives, [4]*TN* number of true negatives, [5]*FNR* false negative rate, [6]*FPR* false positive rate, [7]*SPR* specificity rate, [8]*STR* selectivity rate, [9]*EFR* efficiency rate

D. latifolia model presents the lowest efficiency rate, which has only two false negative but 19 false positives, leading to specificity rate of 90.7%, sensitivity rate of 88.9% and efficiency rate of 79.6%. These results suggest that the variation in the chemical characteristics within *D. latifolia* is greater than in any other species studied in this work. *Dalbergia latifolia* grows across a wide geographic range, from India to Southeast Asia and Melanesia. The authors intend to, in the near future,

collect more *Dalbergia* specimen spectra, including *D. latifolia*, to build a more robust model.

Soft independent model of class analogy: SIMCA

All models were developed with the entire spectral range, 19, 16, 14, 12, 11 and 11 principal components for the modeling of the species *D. nigra*, *D. retusa*, *D. sissoo*, *D. latifolia*, *D. decipularis* and *D. stevensonii*, respectively.

Table 4 indicates the results of the validation set using SIMCA. This table indicates all the results obtained by the SIMCA model. For instance, looking in the *D. nigra* column it can be observed how the samples of this species were classified, which gives us the information of how many samples of that species were correctly classified and the number of false negative results. Among all 141 spectra from *D. nigra* specimens, 140 were correctly classified, leaving just one false negative error, which indicates that this model presents good results in terms of high selectivity rate (STR) and low false negative rate (FNR). On the other hand, taking the first line of the table, it can be observed that 28 spectra from other species were wrongly classified as *D. nigra*, representing false positive errors (FPR) that contribute to high FPR and low SPR.

Through the confusion matrix, it is possible to calculate the figures of merit of each species, shown in Table 5. According to these results, it is possible to check which species are most similar according to each model.

Observing Tables 4 and 5, it is possible to verify which species are most similar among them. *D. retusa* and *D. stevensonii* are the species most confused with *D. nigra*. Out of 21 *D. retusa* spectra, 10 were classified as *D. nigra*, and from the 12 *D. stevensonii* spectra, 6 were also classified as *D. nigra*. This result endorses a study on anatomical features that indicates 33 qualitative similar features among those three species indicating high similarity among them (Gasson et al. 2011). In addition, a study considering quantitative wood anatomy and PCA could not distinguish a group of *Dalbergia* that includes *D. nigra* and *D. stevensonii* (Gasson et al. 2010). In addition, *D. nigra* and *D. stevensonii* have similar fluorescence characteristics (Camargos et al. 2001).

Table 4 Results for the classification of the species-level *Dalbergia* by the SIMCA model

	Actual class					
	D. nigra	*D. retusa*	*D. latifolia*	*D. sissoo*	*D. decipularis*	*D. stevensonii*
Classified as *D. nigra*	140	10	7	4	1	6
Classified as *D. retusa*	0	9	0	0	0	1
Classified as *D. latifolia*	1	0	8	1	3	2
Classified as *D. sissoo*	0	2	1	13	0	0
Classified as *D. decipularis*	0	0	2	0	11	0
Classified as *D. stevensonii*	0	0	0	0	0	3

Table 5 Figures of merit of species-level *Dalbergia* SIMCA analysis

Figures of merit	Species					
	D. nigra	*D. retusa*	*D. latifolia*	*D. sissoo*	*D. decipularis*	*D. stevensonii*
FP	28	1	7	3	2	0
FN	1	12	10	5	4	9
TP	140	9	8	13	11	3
TN	56	203	200	204	208	213
FPR (%)	33.3	0.5	3.4	1.5	1.0	0.0
FNR (%)	0.7	57.1	55.5	27.8	26.7	75.0
SPR (%)	66.7	99.5	96.6	98.5	99.0	100.0
STR (%)	99.3	42.9	44.5	72.2	73.3	25.0
EFR (%)	66.0	42.4	41.1	70.7	72.3	25.0

When using PLS-DA, the *D. nigra* model was the most reliable model, but when using SIMCA, it is only the second best with an efficiency rate of 66%, a considerable drop in efficiency when compared to PLS-DA (98.6%). Only one *D. nigra* specimen was misidentified, being classified as *D. latifolia*. However, 28 samples that are not *D. nigra* were predicted as being this species, leading to a low EFR.

All SIMCA model EFRs were worse than PLS-DA EFRs, and even the highest EFR from the SIMCA models (*D. decipularis,* 72.3%) is 7.3% lower than the worst model from PLS-DA (*D. latifolia*, 79.6%). These results suggest that the clusters of the species are overlapped due to the high similarity, the considerable variation in the specimens of the same species or both factors.

Discrimination of *Dalbergia nigra* from *Dalbergia* spp.

In this approach focus was put on discriminating *D. nigra* from seven other *Dalbergia* species grouped regardless of species, using PLS-DA and SIMCA. Although all *Dalbergia* species are listed in the Appendix of CITES, *D. nigra* is listed in Appendix I, which is the most restrictive and prohibits the international trade in specimens of these species except when the purpose of the import is not commercial (CITES 2017). Thus, in this model, only two classes were created: *D. nigra* that was considered the target class with a class value of 1 and *Dalbergia* spp. with a class value of 0. *Dalbergia* spp. includes *D. cearensis*, *D. tucurensis*, *D. decipularis*, *D. sissoo*, *D. stevensonii*, *D. latifolia* and *D. retusa*.

Partial least square discrimination analysis

In this case, the PLS-DA model required 16 latent variables, which is consistent with a great variability present in the samples, especially the class *Dalbergia* spp. that congregates seven distinct *Dalbergia* species. The distribution of the estimated class values is presented in Fig. 3. This distribution shows that, except for some clear outliers that present abnormally negative class values (around -1), the dispersions in the

Fig. 3 Discrimination between *D. nigra* and *D.* spp. for calibration and validation set. Upside down triangle: *D. nigra;* asterisk: *D.* spp
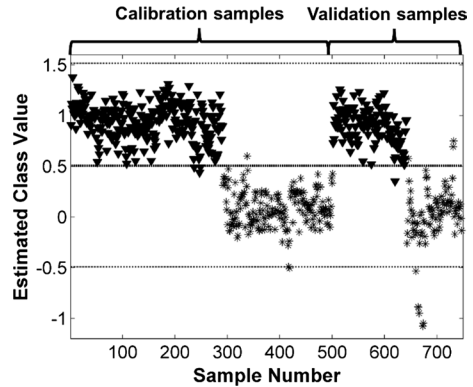
training and validation sets look similar, revealing that there is no overfitting in the model. In Fig. 3, each inverted triangle (*D. nigra*) and each asterisk (*D.* spp.) represent one spectrum of the respective specimen. The top and bottom lines represent the limit of y value. Specimens with y value above the top line and below the bottom line are considered outliers. The estimated class value near 0.5 line represents the discrimination limit. According to the model, specimens above the discrimination limit are discriminated as *D. nigra* and those below the discrimination limit as *D.* spp.

Nine and 13 outlier spectra were observed in the training and validation set, respectively, which represents the exclusion of only 1.8 and 5.2% of the spectra in each set. Considering the heterogeneity present in the specimens, this percentage of exclusion can be considered acceptable for practical purposes.

Figure 3 and Table 6 indicate the reliability of the discrimination of *D. nigra* from other *Dalbergia* species. Only three misclassifications—2 FP and 1 FN—out of 233 spectra (78 specimens) were in validation set. The two FP are the two asterisks above the discrimination limit, which are not *D. nigra* but the model discriminated them as *D. nigra*. The FP is the inverted triangle below the discrimination

Table 6 Figures of merit for PLS-DA discrimination between *D. nigra* versus *Dalbergia* spp.

Figures of merit	*D. nigra* × *Dalbergia* spp.
FP	2
FN	1
TP	140
TN	90
FPR (%)	2.2
FNR (%)	0.7
SPR (%)	97.8
STR (%)	99.3
EFR (%)	97.1

Table 7 Classification results for *D. nigra* versus *Dalbergia* spp. according to SIMCA model

Predicted	Actual class	
	D. nigra class	*Dalbergia* spp. class
Predicted as *D. nigra*	134	10
Predicted as *Dalbergia* spp.	7	95

Table 8 Figures of merit of SIMCA model for *D. nigra* versus *Dalbergia* spp.

Figures of merit	Species	
	D. nigra.	*Dalbergia* spp.
FP	10	7
FN	7	10
TP	134	95
TN	95	134
FPR (%)	9.5	5.0
FNR (%)	5.0	9.5
SPR (%)	94.3	98.3
STR (%)	98.3	94.3
EFR (%)	85.5	85.5

limit; this inverted triangle represents one *D. nigra* specimen discriminated as not *D. nigra*. Thus, the FPR and FNR are relatively small, leading to a high efficiency rate (97.1%).

Soft independent modeling of class analogy

Tables 7 and 8 indicate how the SIMCA model responded to the validation set. Out of 141 *D. nigra* specimens, 134 were correctly classified, but 7 were not—TP: 134 and FN: 7. Out of 105 *Dalbergia* spp. specimens, 95 were correctly classified as not *D. nigra*, but 10 specimens of *Dalbergia* spp. were classified as *D. nigra*—TN: 95 and FP: 10—leading to an efficiency rate of 85.5%. This result represents an increase in efficiency of approximately 33% in relation to the SIMCA species-level model.

As with the species-level models, PLS-DA showed better EFR, STR, and SPR than SIMCA in separating *D. nigra* from *Dalbergia* spp., presenting 23% more EFR.

Spectrometer use

The use of a hand-held spectrometer is very simple, and result interpretation is simple as well. After specimen preparation (see above), law enforcement agent must point the spectrometer toward the heartwood surface and wait for approximately 15 s, and then the result will show on the spectrometer screen. Agents will not have to study the spectra or to develop the model. The model will already be at the

portable spectrometer, and it will automatically indicate to which species the specimen belongs to. The result interpretation is as simple as it can be; on the spectrometer screen it will be written: "It is *D. nigra*" or "It is *D. retusa*".

If a suspicious specimen is not *Dalbergia*, the result is "not identified".

Conclusion

Six different visually confusable species of *Dalbergia* from various countries were separated at the species level using the simple, rapid and reliable NIRS method. Evaluation of PLS-DA and SIMCA modeling demonstrated higher efficiency of PLS-DA models. Specimens of seven *Dalbergia* species treated as a single *Dalbergia* spp. group could be discriminated from *D. nigra* wood with high efficiency rate (97%). Similarly, the efficiency rate for the species-level *D. nigra* model was also high (98%). Although these models present similar efficiencies, the species level includes a lower number of species due to the necessity of more specimens in order to develop individual discrimination models for each species. On the other hand, the model for discrimination of *D. nigra* × *D.* spp. focuses on the species included in the Appendix I of CITES. The present results, obtained with a hand-held device, clearly show that this method is ready to be tested by law enforcement in the field for the identification of *D. nigra* at checkpoints. Expanding both sample breadth (number of taxa) and sample depth (number of specimens per taxon) is emphasis of future work and will permit the development of more broadly applicable groups of discrimination models. Broad development and deployment of hand-held wood identification technologies is one tangible piece of a global cooperative effort to apply forensic wood science in the fight against illegal logging. This work corroborates previous results found by the team for different wood families that NIRS associated with PLS-DA is a tool for timber trade control and species conservation.

References

Acquah GE, Via BK, Billor N, Fasina OO, Eckhardt LG (2016) Identifying plant part composition of forest logging residue using infrared spectral data and linear discriminant analysis. Sensors 16:1375–1390. https://doi.org/10.3390/s16091375

Adedipe OE, Dawson-Andoh B, Slahor J, Osborn L (2008) Classification of red oak (*Quercus rubra*) and white oak (*Quercus alba*) wood using a near infrared spectrometer and soft independent modelling of class analogies. J Near Infrared Spectrosc 16(1):49–57. https://doi.org/10.1255/jnirs.760

Ballabio D, Todeschini R (2009) Multivariate classification for qualitative analysis. In: Da-Wen S (ed) Infrared spectroscopy for food quality analysis and control, 1st edn. Academic Press, London, pp 83–104

Bergo MCJ, Pastore TCM, Coradin VTR, Wiedenhoeft AC, Braga JWB (2016) NIRS identification of *Swietenia macrophylla* is robust across specimens from 27 countries. IAWA J. 37(3):420–430. https://doi.org/10.1163/22941932-20160144

Botelho BG, Reis N, Oliveira LS, Sena MM (2015) Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. Food Chem 181:31–37. https://doi.org/10.1016/j.foodchem.2015.02.077

Braga JWB, Pastore TCM, Coradin VTR, Camargos JAA, Silva AR (2011) The use of near infrared spectroscopy to identify solid wood specimens of *Swietenia macrophylla*. IAWA J 32:285–296. https://doi.org/10.1163/22941932-90000058

Branden KV, Hubert M (2005) Robust classification in high dimensions based on the SIMCA method. Chemometr Intell Lab Syst 79:10–21. https://doi.org/10.1016/j.chemolab.2005.03.002

Burns D, Ciurczak EW (2007) Handbook of Near-Infrared Analysis. CRC Press, Boca Raton

Camargos JAA, Coradin VTR, Czarneski MC, de Oliveira D, Meguerditchian I (2001) Trees of Brazil catalog. IBAMA, Brasilia

Carballo-Meilan A, Goodman AM, Baron MG, Gonzalez-Rodriguez J (2014) A specific case in the classification of woods by FTIR and chemometric: discrimination of Fagales from Malpighiales. Cellulose 21:261–273. https://doi.org/10.1007/s10570-013-0093-2

Carvalho AM (1997) A synopsis of the genus *Dalbergia* (Fabaceae: Dalbergieae) in Brazil. Brittonia 49:97–109. https://doi.org/10.2307/2807701

CITES (1992) Check list of CITES species. CITES. http://checklist.cites.org/#/en/search/output_layout=alphabetical&level_of_listing=0&show_synonyms=1&show_author=1&show_english=1&show_spanish=1&show_french=1&scientific_name=Dalbergia+nigra&page=1&per_page=20. Accessed 10 March 2017

CITES (2017) CITES: Appendices I, II and III to the Convention on International Trade in Endangered Species of Wild Fauna and Flora. CITES. https://www.cites.org/eng/app/appendices.php. Accessed 10 March 2017

De Maesschalck R, Candolfi A, Massart DL, Heuerding S (1999) Decision criteria for soft independent modelling of class analogy applied to near infrared data. Chemometr Intell Lab Syst 47:65–77. https://doi.org/10.1016/S0169-7439(98)00159-2

Espinoza EO, Wiemann MC, Barajas-Morales J, Chavarria GD, McClure PJ (2015) Forensic analysis of CITES-protected *Dalbergia* timber from the Americas. IAWA J. 36(3):311–325. https://doi.org/10.1163/22941932-20150102

Flåten GR, Grung B, Kvalheim OM (2004) A method for validation of reference sets in SIMCA modelling. Chemom Intell Lab Syst 72(1):101–109. https://doi.org/10.1016/j.chemolab.2004.03.003

Gasson P, Miller R, Stekel DJ, Whinder F, Ziemińska K (2010) Wood identification of *Dalbergia nigra* (CITES Appendix I) using quantitative wood anatomy, principal components analysis and naïve Bayes classification. Ann Bot 105(1):45–56. https://doi.org/10.1093/aob/mcp270

Gasson P, Baas P, Wheeler E (2011) Wood anatomy of cited-listed tree species. IAWA J 32:155–198. https://doi.org/10.1163/22941932-90000050

Guzman JAS, Richter HG, Rodriguez AR, Fuentes TFJ (2008) Wood fluorescence of commercial timbers marketed in Mexico. IAWA J 29:311–322. https://doi.org/10.1163/22941932-90000189

Kite GC, Green PWC, Veitch NC, Groves MC, Gasson PE, Simmonds MSJ (2010) Dalnigrin, a neoflavonoid marker for the identification of Brazilian rosewood (*Dalbergia nigra*) in CITES enforcement. Phytochemistry 71(10):1122–1131. https://doi.org/10.1016/j.phytochem.2010.04.011

Lancaster C, Espinoza E (2012) Analysis of select *Dalbergia* and trade timber using direct analysis in real time and time-of-flight mass spectrometry for CITES enforcement. Rapid Commun Mass Spectrom 26(9):1147–1156. https://doi.org/10.1002/rcm.6215

Lazarescu C, Hart F, Pirouz Z, Panagiotidis K, Mansfield SD, Barret JD, Avramidis S (2016) Wood species identification by near-infrared spectroscopy. International Wood Products Journal 8:32–35. https://doi.org/10.1080/20426445.2016.1242270

Lorenzi H (1992) Brazilian trees: manual of identification and cultivation of arboreal plants from Brazil. Plantarum, Nova Odessa

Martins AR, Talhavini M, Vieira ML, Zacca JJ, Braga JWB (2017) Discrimination of whisky brands and counterfeit identification by UV–Vis spectroscopy and multivariate data analysis. Food Chem 229:142–151. https://doi.org/10.1016/j.foodchem.2017.02.024

Miller RB, Wiemann MC (2006) Separation of *Dalbergia nigra* from *Dalbergia spruceana*. U.S. Department of Agriculture. https://www.fpl.fs.fed.us/documnts/fplrp/fpl_rp632.pdf. Accessed 10 Feb 2017

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403:853–858. https://doi.org/10.1038/35002501

Pasquini C (2003) Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. J Braz Chem Soc 14(2):198–219. https://doi.org/10.1590/s0103-50532003000200006

Pastore TCM, Braga JWB, Coradin VTR, Magalhaes WLE, Okino EYA, Camargos JAA, Muniz GIB, Bressan OA, Davrieux F (2011) Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: discrimination of true mahogany, cedar, andiroba, and curupixa. Holzforschung 65:73–80. https://doi.org/10.1515/HF.2011.010

Schwanninger M, Rodrigues JC, Fackler K (2011) A review of band assignments in near infrared spectra of wood and wood components. J Near Infrared Spectrosc 19(5):287–308. https://doi.org/10.1255/jnirs.955

Soares LF, Silva DC, Bergo MCJ, Coradin VTR, Braga JW, Pastore TCM (2017) Evaluation of a NIR handheld device and PLS-DA for discrimination of six similar Amazonian wood species. Quim Nova 40:418–426. https://doi.org/10.21577/0100-4042.20170014

Stehmann JR, Forzza RC, Salino A, Sobral M, Costa DP, Kamino LHY (2009) Plants of Atlantic Forest. Jardim Botânico do Rio de Janeiro, Rio de Janeiro

The Plant List (2013) The plant List 2013 Version 1.1. The plant List. http://www.theplantlist.org. Accessed 7 Apr 2017

UNODC—United Nations Office on Drugs and Crime (2016) Best practice guide for forensic timber identification. UNODC. https://www.unodc.org/documents/Wildlife/Guide_Timber.pdf. Accessed 18 Mar 2017

Wiemann MC, Ruffinatto F (2012) Separation of *Dalbergia stevensonii* from *Dalbergia tucurensis*. USDA Forest Service. https://www.fpl.fs.fed.us/documnts/fplrp/fpl_rp665.pdf. Accessed 09 May 2017

Wise BM, Shaver JM, Gallagher NB, Windig W, Bro R, Koch RS (2006) PLS Toolbox Version 4.0 for use with Matlab™. Eigenvector Research Inc. http://mitr.p.lodz.pl/raman/jsurmacki/pliki/zajecia/LMDiT/cw3/LMDiT_PLS_Manual_4.pdf. Accessed 13 March 2017

Affiliations

Filipe A. Snel[1,2] · Jez W. B. Braga[1] · Diego da Silva[1,2] · Alex C. Wiedenhoeft[3,4,5,6] · Adriana Costa[3,4] · Richard Soares[3] · Vera T. R. Coradin[2] · Tereza C. M. Pastore[2]

[1] Chemistry Institute, University of Brasília, Brasília, DF 70910-000, Brazil

[2] Forest Products Laboratory, Brazilian Forest Service, Brasília, DF 70818-970, Brazil

[3] Forest Products Laboratory, USDA Forest Service, Madison, WI 53726, USA

[4] Department of Botany, University of Wisconsin, Madison, WI 53706, USA

[5] Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907, USA

[6] Ciências Biológicas (Botânica), Universidade Estadual Paulista – Botucatu, São Paulo, Brazil