Analyst

Assessing Utility of Handheld Laser Induced Breakdown Spectroscopy as a Means of Dalbergia Speciation

Journal:	*Analyst*
Manuscript ID	AN-ART-05-2019-000984.R1
Article Type:	Paper
Date Submitted by the Author:	28-Jun-2019

Complete List of Authors:	Celani, Caelin; University of Delaware, Department of Chemistry & Biochemistry Lancaster, Cady; US Fish and Wildlife National Forensics Laboratory Jordan, James; USGS Espinoza, Ed; National Fish & Wildlife Forensic Lab Booksh, Karl; University of Delaware, Department of Chemistry and Biochemistry

Assessing Utility of Handheld Laser Induced Breakdown Spectroscopy as a Means of *Dalbergia* Speciation

Caelin Celani[1], Cady A. Lancaster[2], James A. Jordan[3], Edgard O. Espinoza[4], Karl S. Booksh[1]

1. *Department of Chemistry and Biochemistry, University of Delaware*
2. *United States Forest Service International Programs*
3. *Water Cycle Branch, United States Geological Survey*
4. *National Fish and Wildlife Forensics Lab, United States Fish and Wildlife Service*

Abstract

Many species of *Dalbergia* are prized hardwoods, generally referred to as 'Rosewood,' and used in high-end products due to their distinctive hue and scent. Despite more than 58 species of *Dalbergia* being listed as endangered in Appendix 1 of The Convention on International Trade in Endangered Species of Fauna and Flora (CITES), the illegal logging and trade of this timber is ongoing. In this work, a handheld laser induced breakdown spectrometer (LIBS) was used to analyze seven *Dalbergia* species and two other exotic hardwood species to evaluate the ability of handheld LIBS for rapid classification of *Dalbergia* in the field. The KNN model of the classification presented 80% to 90% sensitivity for discriminating between *Dalbergia* species in the training set. PLS-DA models were based on a binary decision tree structure. Cumulatively, the PLS-DA decision tree model showed greater than 97% sensitivity and 99% selectivity for prediction of *Dalbergia* species included in the training set. The data presented in the following study are promising for the use of handheld LIBS devices and both KNN and PLS-DA models for applications in customs screenings at the port of entry of hard woods, among others.

Introduction

Illegal logging is one of the leading causes of deforestation today. The global trade in timber is estimated to be worth between $30-100 Billion USD annually with up to 40% of the tropical timber traded each year entering the market illegally.[1] The impact on the United States in 2006 from imported products manufactured from illegally harvested timber was estimated at $3.6 billion USD.[2]

The genus *Dalbergia* (Leguminosae), which contains approximately 250 tree species distributed worldwide around the tropics possess properties (i.e. appearance, aroma, etc.) desirable for use in the manufacture of furniture and cabinetry, is perhaps one of the most widely recognized endangered timber species due to its over-exploitation. The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) lists 58 species of protected *Dalbergia,* of which 51 are endemic to Madagascar, one is from Indochina and the remaining six are from Central and South America. In 1992, in response to the threat of logging, Brazilian rosewood (*Dalbergia nigra*) became the first tree species to be listed as prohibited for international trade. Despite its inclusion on Appendix 1 of CITES, *D. nigra* continues to be logged illegally and enters the international market through illegal channels as was recently illustrated in a dispute between U.S. federal law enforcement and certain guitar manufacturers.[3]

For the international community to deter such unlawful practices by presenting a risk of prosecution, methods are needed that can provide rapid determination of CITES status within the supply chain of the protected species' import/export. For example, the ability to make a species determination of a wood sample may be contingent upon identifying the geographic origin. In typical mass spectrometry laboratory analyses, isotopic distribution can be used to identify the region of origin for vegetation.[4–6] Thus, due to well-defined habitats for *Dalbergia* species and methods for determining isotopic distribution, *Dalbergia* provenance may be determined if the species can be identified with a spectroscopic field measurement.[7]

To date, only laboratory-based methods are available for origin identification. Subject matter experts can often discern species of exotic woods though the examination of organic compounds, isotopic composition of the bulk wood, analysis of the trace elements within the wood, or a combination of methods. At ports of entry, a subject matter expert may identify timber products by their morphological features, either on-site or off-site. However, those distinctions are typically limited to the species level and not every species or sub-species identification can be readily made. Aside from macroscopic wood anatomical identification, the most common identification procedure includes the use of 'detector dogs' trained to recognize scents of certain wood species. As of 2016, no analytical technique is readily available for rapid-field identification of wood species, although near- and mid-infrared spectroscopy (NIR, IR) and automated macroscopic wood anatomical identification are being developed.[8]

NIR and IR, and to a lesser extent, X-ray fluorescence spectroscopy (XRF) and laser induced breakdown spectroscopy (LIBS), are being investigated as a field-portable method for identification of wood species. Recent NIR and IR research on classification of wood samples and wood oils produce low error rates with chemometric models, however the vast majority still require a multi-day atmospheric equilibration time, grinding prior to analysis, and use benchtop instrumentation, making these methods insufficient for rapid on-site analyses.[9–14] While the use of handheld NIR & IR spectrometers is documented, the research in this area is limited to only a few groups with a wide range of classification success rates.[15,16] While XRF is successfully used for elemental analysis, application of XRF to wood has been limited to examining the elemental composition of wood wastes.[17] From a practical standpoint, the reliance on a radioactive source would raise add an extra layers of safety and training concerns with placing portable XRF instrumentation at customs entry points for screening hardwood shipments.

LIBS is an attractive option for screening of exotic hardwood samples in the field. LIBS is an adaptation of atomic emission spectroscopy that is capable of providing rapid, multi-element analysis of many materials in any physical state, similar to XRF.[18] LIBS technology has attributes that make it particularly useful for solving difficult and exotic problems such as the detection of Chemical, Biological, Radiological, Nuclear, and Explosive (CBRNE) threats and exploring the solar system.[19–21] LIBS has received attention because of its ability to perform analyses at a distance and *in situ*. LIBS instruments can be configured for handheld,[22–29] or standoff[30–33] detection applications with multiple hand-held LIBS instruments on the market.[34–37]

Spectroscopic analyses for wood classification are highly reliant upon chemometric methods such as k-Nearest Neighbors (KNN) and Partial Least Squares Discriminate Analysis (PLS-DA), among others.[9–16,18] The need of chemometrics in classification problems such as determination of *Dalbergia* species lie in the highly convolved, multivariate nature of spectra collected across many classes of wood samples. While success has been demonstrated with PLS-DA models for obtaining sensitivities and selectivities of nearly 100% for wood oils of the same tree,[13] speciation was not a goal of the study. The application of PLS-DA and NIR to both speciation and determination of provenance showed high sensitivity and selectivity for speciation (100%) but low sensitivity and selectivity (64%-99%) for provenance classification.[38]

In this paper classification among seven species of *Dalbergia* and two other exotic hardwoods by a hand-held LIBS instrument supported with multivariate methods common in chemometrics and machine learning is reported. The ability to rapidly and reliably identify hardwood species at ports of entry would provide a boon for enforcement agencies and minimize the negative impact of false positives on legitimate exotic wood traders.

Experimental

Spectra Collection.

Wood samples of seven *Dalbergia* and two exotic hardwoods were obtained from commercial sources (**Table 1**). The non-*Dalbergia* hardwoods were chosen based on geographical region to assess if there was a geographical signal that could be detected between *Dalbergia* and non-*Dalbergia* species grown in similar locations. Samples were pre-cut, non-sanded, flat, board-like wood sections with approximate dimensions of 0.5x2x3 in. and consisted only of heartwood, as heartwood is the purpose of the illegal timber trade. All samples were validated using traditional wood anatomy and mass spectrometry.[39] Prior to spectra collection, no sample preparation or instrument optimization was performed. Classification was based on the putative species provided by the vender. No further validation was performed. For each of the 9 classes, approximately 10 distinct hardwood samples, each from a different tree, were measured, producing 90 recorded spectra in total. Laser induced breakdown spectroscopy (LIBS) spectra were acquired using the handheld SciAps Z-200C LIBS Analyzer (Boston, MA), containing a 1064 nm source operating at 5-6 mJ and a 50 Hz repetition rate. For each collected spectrum, the laser was rastered over 6 discrete spots, in a 3x2 grid pattern, where each spot was exposed to 10 cleaning laser shots followed by 16 data collection shots. Through this treatment, each collected spectrum was an average of 96 individual spectra. Automated calibration and wavelength scale validation was accomplished using an internal reference of 316 stainless steel. Example LIBS spectra are shown in **Figure 1A**.

INSERT TABLE 1 ABOUT HERE

Analysis.
All computations were performed in the Matlab 2018b (Matlab Natick, MA) operating environment augmented by the PLS-Toolbox (Eigenvector, Chelan WA). Spectral baseline removal by an automatic Whittaker filter[40] and Savitzky-Golay[41] first derivative were performed by algorithms included in the PLS-Toolbox. Following preprocessing of the data, decluttering by External Parameter Orthogonalization (EPO)[42] and analyses by either uniform (unweighted) KNN or PLS-DA was also performed in the PLS-TOOLBOX. Decision trees were generated by 'in house' software.

The Whittaker filter is a means of estimating baseline signal and deemphasizing any baseline humps that may unequally augment signal intensity to a subset of the variables in the spectra. Mathematically, spectral contributions are determined to belong to either baseline or signal by iteratively fitting a low order polynomial to the baseline in a piecewise manner. Whittaker parameters were set to $\lambda = 100$ and $P = 0.001$ for the least squares estimation of the baselines. These parameters set the allowable amount of curvature (l) and negativity (P) in the estimated baseline.

The Savitzky-Golay method applied the first derivative of a local 2nd order polynomial to simultaneously remove any residual baseline off-set and smooth high frequency random errors in each spectrum. The filter width was set to 15 points, approximately half the width of the LIBS peaks.

External Parameter Orthogonalization (EPO) was applied with a 3 principal component model to estimate the uninformative variance within each class (class). The portions of the whole data set that is correlated to this variance is removed. The EPO algorithm is a "hard" orthogonalization method, i.e. it all variance is totally removed. Consequently, care must be exhibited to not over-fit the data with a large number of PCs in the EPO method and begin to remove usable signal.

For each analysis, the 90-sample data set was divided into a 72-sample training set and a 18 sample validation set. For the validation set, 2 samples were randomly removed from each class. When PLS-DA was applied 3 times on different training and validation sets, a truly unique validation set was randomly determined such that there were always 2 samples removed from each class and no sample was assigned to a training set twice. For cross-validation of the KNN and PLS-DA models, a further 1 sample (10% of total data) is temporarily removed, thus only 7 samples per class are employed to generate the KNN or PLS-DA models during cross-validation.

Results & Discussion

To test the efficacy of hand-held LIBS instrumentation for differentiating among multiple *Dalbergia* species, two classification methods common in chemometric and machine learning applications were investigated. KNN is a non-parametric, unguided classification technique in that it does not use prior knowledge of class identities to assemble sample groupings based on observational (in this case LIBS spectra) similarities.[43,44] By contrast, PLS-DA is a guided, parametric model that relies on linear regression to differentiate among two (or more) putative classes of samples.[45–47] *A priori* class information is enlisted to divide a high dimensional space defined by multivariate observations (again, LIBS spectra here).

KNN and PLS-DA with decision trees each offer a particular set of advantages and disadvantages for classification problems. The advantage of KNN is, thus, robustness against misidentified sample classes in the training set and resistance to overfitting the model by relying on chance correlations of random errors within the data set. The disadvantage of KNN is the lost opportunity of building a more sensitive or parsimonious model by enlisting all *a priori* knowledge about the training set in the modelling process. With PLS-DA, the use of *a priori* class knowledge to drive the PLS model mitigates the obfuscating effect of random errors and other sources of variance within the training set. With the LIBS spectra of wood samples these variances can be large relative to the spectral differences unique to differentiating among *Dalbergia* species. However, the linear PLS models are often inefficient in simultaneously differentiating among three or more classes. Consequently, binary decision trees,[48,49] such as those used in Classification and Regression Trees (CART),[50] are employed to simplify the differentiation among multiple classes (*Dalbergia* species) into a series of simple, often binary, choices.

Preprocessing of LIBS spectra.

The raw LIBS spectra (**Figure 1A**) exhibited a significant baseline with broad, variable features around 390 nm, 460 nm, and 510 nm. A few variables, notably the most intense features at 395 nm, 398 nm, 425 nm and 195 nm presented large sample-to-sample variability within each class. Additionally, LIBS analyses of the wood samples returns sparse spectra; approximately 90% of the variables are below the 3 standard deviations of instrumental noise. Applying a KNN model, with n = 3, to the raw data yielded correct classification of only 35 of 90 hardwood samples in the calibration model (**Figure 2A**). Perhaps not surprisingly, the two distinct non-*Dalbergia* classes (8 & 9) are accurately classified relative to the seven *Dalbergia* species (**Figure 2A**).

INSERT FIGURE 1 ABOUT HERE

Preprocessing of the LIBS spectra was essential for developing sensitive and selective models for classification of the 9 distinct types of hardwood sampled. Application of a Whittaker filter followed by determination of the spectral 1st derivative significantly decreased the baseline variance (**Figure 1B**). The error structure of each spectrum was normalized by taking the square root of the absolute value of the intensity for each baseline corrected spectrum (**Figure 1C**). For all analyses, variables representing the uninformative spectral baseline were removed by applying a cut-off filter set to either 0.5 units or 1.0 units for the mean normalized spectrum. The cut-off filters reduced the number of points per sample from 17431 to 1849 or 489, respectively. After normalization and variable selection, the spectra were autoscaled ($\mu = 0, \sigma = 1$), decluttered by EPO, then autoscaled again. The benefit of some variable selection to ignore uninformative baseline variables is evident when comparing the KNN (n = 3) models with no variable removal (**Figure 2B**, 55/90 correct) to the models applied to the 0.5 cut-off filter (**Figure 2C** 81/90 correct) and the 1.0 cut-off filter (**Figure 2D** 83/90 correct). While not autoscaling before or after applying EPO occasionally presented a more sensitive or selective model, the technique of pre- and post-autoscaling was found to be the most robust across preliminary models and was employed for all data presented here. Similarly, preliminary analyses could not support a greater merit to either cut-off filter values, so both the 0.5 filter and 1.0 filter are considered for future analyses.

INSERT FIGURE 2 ABOUT HERE

Classification by KNN

Predictive KNN (n = 3) models determined from a 72-sample training set and applied to an 18-sample test set yielded a 94% success rate for classification of the test set with the 0.5 cut-off filter and a 100% success rate with the 1.0 cut-off filter. Only 1 *Dalbergia* sample was misclassified in the test set. Cross-validated classification errors for the two 72-sample training sets were appreciably greater than the two 18 sample test (prediction) sets. However, the cross-validation errors are effectively based on only 6 samples per class, while prediction errors were based on 7 training samples per class and the preliminary analyses above employed 10 samples per class. In spite of the functionally sparser data clouds for each class, cross-validation was still 90% correct for the 0.5 cut-off (**Figure 3A**) and 79% correct for the 1.0 cut-off (**Figure 3B**).

Closer examination of results yields a few observations about KNN model performance in this data set. For cross-validation, no class of *Dalbergia* samples has both a sensitivity and selectivity of 1.00; all *Dalbergia* classes either have samples misclassified as a different *Dalbergia* species or have samples from a different *Dalbergia* misappropriated into that class. The KNN model works better with the data from the 0.5 cut-off filter (7 miss-classifications) than with the 1.0 cut-off filter (15 miss-classifications). The KNN model has the hardest time classifying *Dalbergia melanoxylon* (class 3) with a sensitivity of 0.62 for the 0.5 cut-off data and only 0.25 for the 1.0 cut-off data. However, the KNN method deftly classifies the more distinct non-*Dalbergia* hardwoods with a sensitivity of 1.0 for the data with the applied 0.5 cut-off filter.

INSERT FIGURE 3 ABOUT HERE

To further validate that the KNN model was classifying based on real spectral features and not forcing classification through chance correlations within each class, classification was repeated with randomized class assignments. The KNN model (n = 3) returned only 4 of 90 correct classifications (4.4%), less than would be expected by random chance.

Classification by Partial Least Squares – Discriminant Analyses (PLS-DA)

The sensitivity and selectivity of PLS-DA for determining class identity of the 18-sample validation set was 0.97 and 0.99, respectively, for the 0.5 intensity cutoff filter and 0.98 and 0.99, respectively, for the 1.0 intensity cut-off filter. The difference between the 0.97 and 0.98 sensitivity is one sample misclassification across 3 realizations of an 18-sample validation set (54 samples total) (**Table 2**). The confusion matrix for 0.5 intensity cut-off data shows that one out of six test samples from class 2 was predicted to belong to class 9 and one of six test samples from class 3 was predicted to belong to class 7 (**Table 3**), while for the 1.0 intensity cut-off only one sample from class 2 was miss-allocated (**Table 4**).

INSERT TABLE 2 ABOUT HERE

The 0.92 sensitivity determined by cross-validation (**Table 2**) is impressive considering only 7 samples per class are employed to construct each PLS-DA model. For the 0.5 intensity cut-off data, the majority of the miss-classifications come from classes 1 and 2. An average of two class 1 samples were predicted as belonging to class 6 and from the thee resampled analyses, a class 1 sample was predicted to belong to class 4 on 1 of the 3 replicates (**Table 3**). At the same time, on average 1 sample from class 2 was miss-classified as class 7 and another as class 9 during cross validation. The 1.0 intensity cut-off data exhibited a more uniform distribution of miss-classifications during cross-validation (**Table 4**). Only class 7 had 1.00 sensitivity and only class 4 averaged more than one miss-classification.

INSERT TABLE 3 ABOUT HERE

INSERT TABLE 4 ABOUT HERE

The key to success of the PLS-DA models for classification of *Dalbergia* species is the simplification of the classification process by employing decision trees to transform the classification model to a series of binary decisions. The two decision trees independently determined for data sets from each intensity cut-off are very similar. They differ only in class grouping of the penultimate decision of each branch and the number of factors employed in the PLS-DA model for each decision (**Scheme 1**). For example, the first decision of each data set is whether a sample belongs to the group of classes 2, 7, 8, and 9 or the group of classes 1, 3, 4, 5, and 6. With the 0.5 intensity cut-off data, a 1 factor PLS-DA model was employed and a 3 factor PLS-DA model was employed for the 1.0 intensity cut-off data. At the ends of the decision trees, differences can be seen in how the model distinguishes among the final three classes. For example, at Level 3 the 0.5 intensity cut-off tree differentiates class 9 from classes 2 and 7 before distinguishing among classes 2 and 7 at Level 4. While the decision tree for the 1.0 cut-off data differentiates class 7 from classes 2 and 9 at Level 3 and then resolves class 2 from class 9 at Level 4. The differences in structure at the tips of the trees are understandable because the final differentiations are generally the hardest to resolve.

Assessment of the classification models

Classification of *Dalbergia* species by LIBS was expected to be a challenging application. Consequently, the 0.97 sensitivity and 0.99 selectivity for these samples by a PLS-DA decision tree should be viewed skeptically. One concern is that the high sensitivity for prediction could be the result of cherry picking the test set to lie in the center of each class. To guard against that, the test set was resampled 3 times without replacement such that 60% of the data was included in the test set at one time or another. All three models performed equivalently (**Supplemental Information**).

A second concern is that the models are forcing classification based on random signals. Were this the case, it is doubtful that the sensitivity for either cross-validation or prediction would above 0.9. However, to further validate that the models were not seizing upon spurious correlations, the classification models were repeated with randomized class affiliations. With randomized classes the KNN model (n = 3) miss-classified 84 out of 90 samples. The PLS-DA model at the Level 1 decision exhibited a cross-validation sensitivity of 0.44 and cross-validation selectivity of 0.61 even though 9 samples per class were employed in construction of each model.

It is worth noting that the wood samples were acquired from two different vendors. A modicum of confidence in the models is added because the initial classification is not a binary decision based on vendor. Classes 1 and 6, the two classes purchased from Gilmer Wood, do fall on the same of the first binary decision, but with three other classes of *Dalbergia*. While there might be a systematic difference in the wood spectra based on vendor, any difference here is not so great as to induce those samples from forming a distinct and separate branch of the decision tree.

INSERT SCHEME 1 ABOUT HERE

Conclusions

This preliminary study presents encouraging evidence that *Dalbergia* species can be rapidly identified by a handheld LIBS device. With appropriate data preprocessing both KNN and PLS-DA driven models can accurately classify *Dalbergia* species. The worst case observed 0.92 sensitivity is sufficient for customs screening of imported hardwoods at ports of entry. In such an application, the goal is not to definitively identify the species of an unknown *Dalbergia* sample. Rather, the rapid LIBS field test would screen the sample for a mismatch to the species reported on the manifest. If these two identities do not match, the shipment would be flagged for further investigation or laboratory analyses. As identified by Dormontt *et al.*, there is a growing need for both tools and methods that allow for rapid screening of potentially illegally sourced wood at ports of entry.[51] Consequently, a high sensitivity is desired to minimize impact on legitimate trade. A high selectivity is important so protected or trade-regulated species are not mistakenly identified as legal, tradeable woods.

The next step of this project is to test the models on a larger number of samples acquired from multiple sources. Simultaneously, the data collection parameters should be optimized to determine the minimum signal-to-noise needed to realize reliable classification. Lower signal-to-noise ratios can be more rapidly be acquired by handheld instrumentation. An additional exploratory task is to determine whether fusion of the LIBS spectra with complimentary information, such has from a handheld XRF or handheld IR reflectance, would significantly improve the models.

Acknowledgements

The authors thank NSF CHE**1506853** for support of the modeling component of this project. The authors also thank SciAps for loan of the Z-200 LIBS Analyzer.

References

(1) UNODC. The Globalisation of Crime: A Transnational Organized Crime Threat Assessment. *Glob. Crime A Transnatl. Organ. Crime Threat Assess.* **2013**, No. April, 314.

(2) China illegal tropical timber import export USA President Initiative Against Illegal Logging rubber wood furniture statistics pulp paper Brazil Cameroon Congo Equatorial Guinea Gabon Liberia Russia Malaysia Indonesia PNG Papua New Guinea Solomon Islands B http://www.globaltimber.org.uk/ChinaIllegalImpExp.htm (accessed Apr 16, 2019).

(3) U.S. Fish & Wildlife Service. *International Trade in Plants and Wildlife Information for Musicians and Manufacturers of Musical Instruments*.

(4) Hartman, G.; Danin, A. Isotopic Values of Plants in Relation to Water Availability in the Eastern Mediterranean Region. *Oecologia* **2010**, *162* (4), 837–852. https://doi.org/10.1007/s00442-009-1514-7.

(5) Tahmasebi, F.; Longstaffe, F. J.; Zazula, G.; Bennett, B. Nitrogen and Carbon Isotopic Dynamics of Subarctic Soils and Plants in Southern Yukon Territory and Its Implications for Paleoecological and Paleodietary Studies. *PLoS One* **2017**, *12* (8), e0183016.

(6) Milligan, H. E.; Pretzlaw, T. D.; Humphries, M. M. Stable Isotope Differentiation of Freshwater and Terrestrial Vascular Plants in Two Subarctic Regions. *Ecoscience* **2010**, *17* (3), 265–275.

(7) Winfield, K.; Scott, M.; Graysn, C. *Global Status of Dalbergia and Pterocarpus Rosewood Producing Species in Trade*; 2016.

(8) International Consortium o Combating Wildlife Crime. Best Practice Guide for Forensic Timber Identification INTERNATIONAL CONSORTIUM ON COMBATING WILDLIFE CRIME. *St/Nar/52* **2016**, 226.

(9) Duca, D.; Pizzi, A.; Rossini, G.; Mengarelli, C.; Foppa Pedretti, E.; Mancini, M. Prediction of Hardwood and Softwood Contents in Blends of Wood Powders Using Mid-Infrared Spectroscopy. *Energy and Fuels* **2016**, *30* (4), 3038–3044. https://doi.org/10.1021/acs.energyfuels.5b02994.

(10) Chen, H.; Ferrari, C.; Angiuli, M.; Yao, J.; Raspi, C.; Bramanti, E. Qualitative and Quantitative Analysis of Wood Samples by Fourier Transform Infrared Spectroscopy and Multivariate Analysis. *Carbohydr. Polym.* **2010**, *82* (3), 772–778. https://doi.org/10.1016/j.carbpol.2010.05.052.

(11) Zhou, C.; Jiang, W.; Cheng, Q.; Via, B. K. Multivariate Calibration and Model Integrity for Wood Chemistry Using Fourier Transform Infrared Spectroscopy. *J. Anal. Methods Chem.* **2015**, *2015*, 1–9. https://doi.org/10.1155/2015/429846.

(12) Via, B. K.; Zhou, C.; Acquah, G.; Jiang, W.; Eckhardt, L. Near Infrared Spectroscopy Calibration for Wood Chemistry: Which Chemometric Technique Is Best for Prediction and Interpretation? *Sensors (Switzerland)* **2014**, *14* (8), 13532–13547. https://doi.org/10.3390/s140813532.

(13) Almeida, M. R.; Fidelis, C. H. V.; Barata, L. E. S.; Poppi, R. J. Classification of Amazonian Rosewood Essential Oil by Raman Spectroscopy and PLS-DA with Reliability Estimation. *Talanta* **2013**, *117*, 305–311. https://doi.org/10.1016/j.talanta.2013.09.025.

(14) Amusant, N.; Beauchène, J.; Digeon, A.; Chaix, G. Essential Oil Yield in Rosewood (Aniba Rosaeodora Ducke): Initial Application of Rapid Prediction by near Infrared Spectroscopy Based on Wood Spectra. *J. Near Infrared Spectrosc.* **2016**, *24* (6), 507–515. https://doi.org/10.1255/jnirs.1241.

(15) Soares, L. F.; Da Silva, D. C.; Bergo, M. C. J.; Coradin, V. T. R.; Braga, J. W. B.; Pastore, T. C. M. AVALIAÇÃO DE ESPECTRÔMETRO NIR PORTÁTIL E PLS-DA PARA A DISCRIMINAÇÃO DE SEIS ESPÉCIES SIMILARES DE MADEIRAS AMAZÔNICAS EVALUATION OF A NIR HANDHELD DEVICE AND PLS-DA FOR DISCRIMINATION OF SIX SIMILAR. *Quim. Nov.* **2017**, *40* (4), 418–426. https://doi.org/10.21577/0100-4042.20170014.

(16) Silva, D. C.; M, P. T. C.; F, S. L.; S, de B. F. A.; J, B. M. C.; H, C. V. T.; Bahia, G. A.; Herrera, S. M.; Beltetón, C. C.; B, B. J. W. Determination of the Country of Origin of True Mahogany (Swietenia Macrophylla King) Wood in Five Latin American Countries Using Handheld NIR Devices and Multivariate Data Analysis. *Holzforschung*. 2018, p 521. https://doi.org/10.1515/hf-2017-0160.

(17) Fellin, M.; Negri, M.; Zanuttini, R. Multi-Elemental Analysis of Wood Waste Using Energy Dispersive X-Ray Fluorescence (ED-XRF) Analyzer. *Eur. J. Wood Wood Prod.* **2014**, *72* (2), 199–211. https://doi.org/10.1007/s00107-013-0766-4.

(18) Cremers, D.; Radziemski, L. *Handbook of Laser-Induced*; 2013. https://doi.org/10.1002/9781118567371.

(19) Arnaud, C. H. Chemical and Engineering News : "news Edition" of the American Chemical Society. The American Chemical Society April 2012, pp 36–37.

(20) Musazzi, S. *Springer Series in Optical Sciences 182 Laser-Induced Breakdown Spectroscopy*; 2014.

(21) Harmon, R. S.; Russo, R. E.; Hark, R. R. Applications of Laser-Induced Breakdown Spectroscopy for Geochemical and Environmental Analysis: A Comprehensive Review. *Spectrochim. Acta - Part B At. Spectrosc.* **2013**, *87* (September 2018), 11–26. https://doi.org/10.1016/j.sab.2013.05.017.

(22) Manard, B. T.; Schappert, M. F.; Wylie, E. M.; McMath, G. E. Investigation of Handheld Laser Induced Breakdown Spectroscopy (HH LIBS) for the Analysis of Beryllium on Swipe Surfaces. *Anal. Methods* **2019**, *11* (6), 752–759. https://doi.org/10.1039/c8ay02473a.

(23) Senesi, G. S.; Manzari, P.; Tempesta, G.; Agrosì, G.; Touchnt, A. A.; Ibhi, A.; De Pascale, O. Handheld Laser Induced Breakdown Spectroscopy Instrumentation Applied to the Rapid Discrimination between Iron Meteorites and Meteor-Wrongs. *Geostand. Geoanalytical Res.* **2018**, *42* (4), 607–614. https://doi.org/10.1111/ggr.12220.

(24) Manard, B. T.; Wylie, E. M.; Willson, S. P. Analysis of Rare Earth Elements in Uranium Using Handheld Laser-Induced Breakdown Spectroscopy (HH LIBS). *Appl. Spectrosc.* **2018**, *72* (11), 1653–1660. https://doi.org/10.1177/0003702818775431.

(25) Senesi, G. S.; Manzari, P.; Consiglio, A.; De Pascale, O. Identification and Classification of Meteorites Using a Handheld LIBS Instrument Coupled with a Fuzzy Logic-Based Method. *J. Anal. At. Spectrom.* **2018**, *33* (10), 1664–1675. https://doi.org/10.1039/c8ja00224j.

(26) Harmon, R. S.; Hark, R. R.; Throckmorton, C. S.; Rankey, E. C.; Wise, M. A.; Somers, A. M.; Collins, L. M. Geochemical Fingerprinting by Handheld Laser-Induced Breakdown Spectroscopy. *Geostand. Geoanalytical Res.* **2017**, *41* (4), 563–584. https://doi.org/10.1111/ggr.12175.

(27) Piorek, S. Rapid Sorting of Aluminum Alloys with Handheld LIBS Analyzer. **2019**, *10*, 348–354.

(28) Pospíšilová, E.; Novotný, K.; Pořízka, P.; Hradil, D.; Hradilová, J.; Kaiser, J.; Kanický, V. Depth-Resolved Analysis of Historical Painting Model Samples by Means of Laser-Induced Breakdown Spectroscopy and Handheld X-Ray Fluorescence. *Spectrochim. Acta - Part B At. Spectrosc.* **2018**, *147* (May), 100–108. https://doi.org/10.1016/j.sab.2018.05.018.

(29) Harmon, R. S.; Throckmorton, C. S.; Hark, R. R.; Gottfried, J. L.; Wörner, G.; Harpp, K.; Collins, L. Discriminating Volcanic Centers with Handheld Laser-Induced Breakdown Spectroscopy (LIBS). *J. Archaeol. Sci.* **2018**, *98* (July), 112–127. https://doi.org/10.1016/j.jas.2018.07.009.

(30) Scaffidi, J.; Angel, S. M.; Cremers, D. A. Emission Enhancement Mechanisms in Dual-Pulse LIBS. *Anal. Chem.* **2006**, *78* (1), 24–32. https://doi.org/10.1021/ac069342z.

(31) Marquardt, B. J.; Goode, S. R.; Michael Angel, S. In Situ Determination of Lead in Paint by Laser-Induced Breakdown Spectroscopy Using a Fiber-Optic Probe. *Anal. Chem.* **1996**, *68* (6), 977–981. https://doi.org/10.1021/ac950828h.

(32) Angel, S. M.; Stratis, D. N.; Eland, K. L.; Lai, T.; Berg, M. A.; Gold, D. M. LIBS Using Dual- and Ultra-Short Laser Pulses. *Fresenius. J. Anal. Chem.* **2002**, *369* (3–4), 320–327. https://doi.org/10.1007/s002160000656.

(33) Michael Angel, S.; Bonvallet, J.; Lawrence-Snyder, M.; Pearman, W. F.; Register, J. Underwater Measurements Using Laser Induced Breakdown Spectroscopy. *J. Anal. At. Spectrom.* **2016**, *31* (1), 328–336. https://doi.org/10.1039/c5ja00314h.

(34) SciApps. No TitleZ-300 LIBS Analyzer https://www.sciaps.com/libs-handheld-laser-analyzers/z-300/ (accessed Apr 21, 2019).

(35) Bruker. EOS 500 https://www.bruker.com/products/x-ray-diffraction-and-elemental-analysis/libs/eos-500/overview.html (accessed Apr 21, 2019).

(36) BWTEK. NanoLIBS http://bwtek.com/products/nanolibs/ (accessed Apr 21, 2019).

(37) Inc., S. PORTA-LIBS-2000 and Plasma Monitor Configurations.

(38) Yang, Z.; Liu, Y.; Pang, X.; Li, K. Preliminary Investigation into the Identification of Wood Species from Different Locations by Near Infrared Spectroscopy. *BioResources* **2015**, *10* (4), 8505–8517. https://doi.org/10.15376/biores.10.4.8505-8517.

(39) Lancaster, C.; Espinoza, E. 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.* **2012**, *26* (9), 1147–1156. https://doi.org/10.1002/rcm.6215.

(40) Whittaker, E. T. On a New Method of Graduation. *Proc. Edinburgh Math. Soc.* **1922**, *41*, 63–75. https://doi.org/10.1017/S001309150000359X.

(41) Savitzky, A.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36* (8), 1627–1639. https://doi.org/10.1021/ac60214a047.

(42) Roger, J. M.; Chauchard, F.; Maurel, V. B.; Roger, J. M.; Chauchard, F.; Maurel, V. B. EPO-PLS External Parameter Orthogonalisation of PLS Application to Temperature-Independent Measurement of Sugar Content of Intact Fruits To Cite This Version : HAL Id : Hal-00464022 Introduction. **2010**.

(43) Alhazen; Smith, A. M. *Alhacen's theory of visual perception : a critical edition, with English translation and commentary, of the first three books of Alhacen's De aspectibus, the medieval Latin version of Ibn al-Haytham's Kitab al-Manazir*; American Philosophical Society: Philadelphia, 2001.

(44) Stevens, K. N.; Cover, T. M.; Hart, P. E. Nearest Neighbor. **1967**, *I*.

(45) Wold, S.; Martens, H.; Wold, H. The Multivariate Calibration Problem in Chemistry Solved by the PLS Method. In *Matrix Pencils*; Kågström, B., Ruhe, A., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1983; pp 286–293.

(46) Barker, M.; Rayens, W. Partial Least Squares for Discrimination. *J. Chemom.* **2003**, *17* (3), 166–173. https://doi.org/10.1002/cem.785.

(47) Vandeginste, B. G. M. Handbook of Chemometrics and Qualimetrics. Part B Part B. Elsevier: Amsterdam; Boston 1998.

(48) Belson, W. Matching and Prediction on the Principle of Biological Classification. *J. R. Stat. Soc.* **1953**, *8* (2), 8679–8681.

(49) Liu, H.; Gegov, A.; Cocea, M. Representation of Classification Rules. In *Rule Based Systems for Big Data: A Machine Learning Approach*; Springer International Publishing: Cham, 2016; pp 51–62. https://doi.org/10.1007/978-3-319-23696-4_5.

(50) Breiman, L.; Friedman, J.; Stone, C. J.; Olshen, R. A. *Classification and Regression Trees*; The Wadsworth and Brooks-Cole statistics-probability series; Taylor & Francis, 1984.

(51) Dormontt, E. E.; Boner, M.; Braun, B.; Breulmann, G.; Degen, B.; Espinoza, E.; Gardner, S.; Guillery, P.; Hermanson, J. C.; Koch, G.; et al. Forensic Timber Identification: It's Time to Integrate Disciplines to Combat Illegal Logging. *Biol. Conserv.* **2015**, *191*, 790–

798. https://doi.org/10.1016/j.biocon.2015.06.038.

TABLE 1

Table 1. Origins, class labels, and commercial source of wood samples

Species	Origin	Class Number	Number of Samples
Dalbergia baronii[a]	Madagascar	1	10
Dalbergia latifolia[b]	India, Malaysia	2	10
Dalbergia melanoxylon[b]	Tanzania	3	10
Dalbergia nigra[b]	Brazil	4	10
Dalbergia retusa[b]	Mexico	5	11
Dalbergia spruceana[a]	Brazil	6	10
Dalbergia stevensonii[b]	Honduras	7	9
Phoebe porosa[b]	Brazil	8	10
Swietenia macrophylla[c]	Central & South America	9	10

[a] Gilmer Wood Co., Portland, OR, USA
[b] Eisenbrand Inc. Exotic Hardwoods, Torrance, CA, USA
[c] Cook Woods, Klamath Falls, OR, USA

TABLE 2

Table 2. Ensemble Sensitivity and Selectivity of all PLS-DA models for *Dalbergia* Classification.

	Calibration (0.5 Cut / 1.0 Cut)		Cross-Validation (0.5 Cut / 1.0 Cut)		Prediction (0.5 Cut / 1.0 Cut)	
	Sensitivity	Selectivity	Sensitivity	Selectivity	Sensitivity	Selectivity
Class 1	1.00 / 1.00	1.00 / 1.00	0.71 / 0.88	1.00 / 0.98	1.00 / 1.00	1.00 / 1.00
Class 2	1.00 / 1.00	1.00 / 1.00	0.75 / 0.88	0.99 / 0.99	0.85 / 0.85	1.00 / 0.99
Class 3	1.00 / 1.00	1.00 / 1.00	1.00 / 0.91	0.99 / 0.99	0.85 / 1.00	1.00 / 1.00
Class 4	1.00 / 1.00	1.00 / 1.00	1.00 / 0.79	0.99 / 0.99	1.00 / 1.00	1.00 / 1.00
Class 5	1.00 / 1.00	1.00 / 1.00	0.92 / 0.97	1.00 / 0.99	1.00 / 1.00	1.00 / 1.00
Class 6	1.00 / 1.00	1.00 / 1.00	0.96 / 0.96	0.97 / 0.98	1.00 / 1.00	1.00 / 1.00
Class 7	1.00 / 1.00	1.00 / 1.00	0.96 / 1.00	0.98 / 0.99	1.00 / 1.00	0.97 / 1.00
Class 8	1.00 / 1.00	1.00 / 1.00	1.00 / 0.96	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00
Class 9	1.00 / 1.00	1.00 / 1.00	0.96 / 0.96	0.98 / 0.98	1.00 / 1.00	0.99 / 1.00
Average	**1.00 / 1.00**	**1.00 / 1.00**	**0.92 / 0.92**	**0.99 / 0.99**	**0.97 / 0.98**	**0.99 / 0.99**

TABLE 3

Table 3. Confusion Matrix for PLS-DA applied to 0.5 Cut-off Filtered Average LIBS Data. Shown values are for the calibration / cross validation / and test sets. Each value represents the average number of assignments across three replicate analyses, each with a unique training set and different calibration set. Values of '0' are replaced in the table with '.' for clarity of reading.

	N =	Pred. 1	Pred. 2	Pred. 3	Pred. 4	Pred. 5	Pred. 6	Pred. 7	Pred. 8	Pred. 9
True 1	8/8/2	8/5.7/2	. /. /.	. /. /.	. /.3/.	. /. /.	. /2/.	. /. /.	. /. /.	. /. /.
True 2	8/8/2	. /. /.	8/6/1.7	. /. /.	. /. /.	. /. /.	. /. /.	. /1/.	. /. /.	. /1/.3
True 3	8/8/2	. /. /.	. /. /.	8/8/1.7	. /. /.	. /. /.	. /. /.	. /. /.3	. /. /.	. /. /.
True 4	8/8/2	. /. /.	. /. /.	. /. /.	8/8/2	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.
True 5	9/9/2	. /. /.	. /. /.	. /.7 /.	. /. /.	9/8.3/2	. /. /.	. /. /.	. /. /.	. /. /.
True 6	8/8/2	. /. /.	. /. /.	. /. /.	. /.3/.	. /. /.	8/7.7/2	. /. /.	. /. /.	. /. /.
True 7	7/7/2	. /. /.	. /.3/.	. /. /.	. /. /.	. /. /.	. /. /.	7/6.7/2	. /. /.	. /. /.
True 8	8/8/2	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.	8/8/2	. /. /.
True 9	8/8/2	. /. /.	. /.3/.	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.	8/7.7/2

TABLE 4

Table 4. Confusion Matrix for PLS-DA applied to 1.0 Cut-off Filtered Average LIBS Data. Shown values are for the calibration / cross validation / and test sets. Each value represents the average number of assignments across three replicate analyses, each with a unique training set and different calibration set. Values of '0' are replaced in the table with '.' for clarity of reading.

	N =	Pred. 1	Pred. 2	Pred. 3	Pred. 4	Pred. 5	Pred. 6	Pred. 7	Pred. 8	Pred. 9
True 1	8/8/2	8/7/2	. /. /.	. /. /.	. /.3/.	. /. /.	. /.7/.	. /. /.	. /. /.	. /. /.
True 2	8/8/2	. /. /.	8/7/1.7	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.	. /1 /.3
True 3	8/8/2	. /./.	. /. /.	8/7.3/2	. /./.	. /.7/.	. /./.	. /. /.	. /. /.	. /. /.
True 4	8/8/2	. /1/.	. /. /.	. /. /.	8/6.3/2	. /. /.	. /.7/.	. /. /.	. /. /.	. /. /.
True 5	9/9/2	. /. /.	. /. /.	. /.3 /.	. /. /.	9/8.7/2	. /. /.	. /. /.	. /. /.	. /. /.
True 6	8/8/2	. /. /.	. /. /.	. /. /.	. /.3/.	. /. /.	8/7.7/2	. /. /.	. /. /.	. /. /.
True 7	7/7/2	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.	7/7/2	. /. /.	. /. /.
True 8	8/8/2	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.	. /.3/.	8/7.7/2	. /. /.
True 9	8/8/2	. /. /.	. /.3/.	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.	. /. /.	8/7.7/2

FIGURE 1

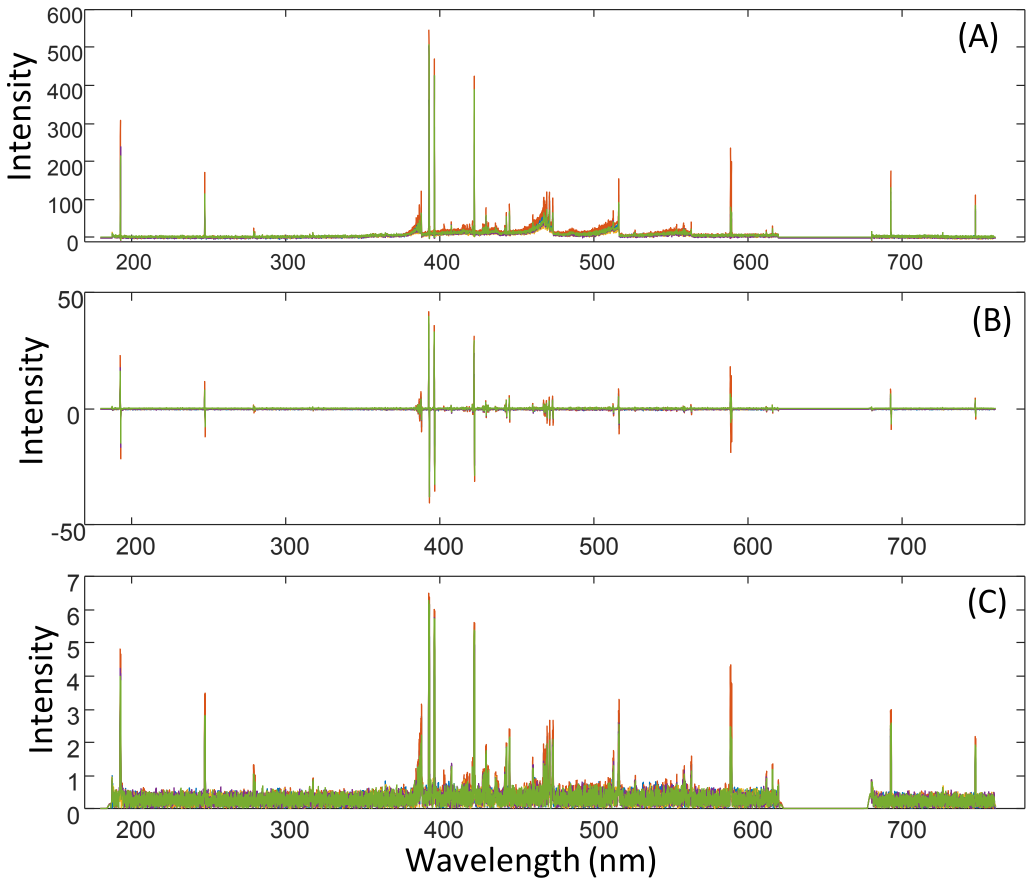


Figure 1. LIBS spectra of (A) Raw data, (B) the data after application of a Whittaker filter and the first spectral derivative, and (C) fully preprocessed and baseline corrected data. Spectra shown are 5 select spectra from randomly chosen timbers.

FIGURE 2

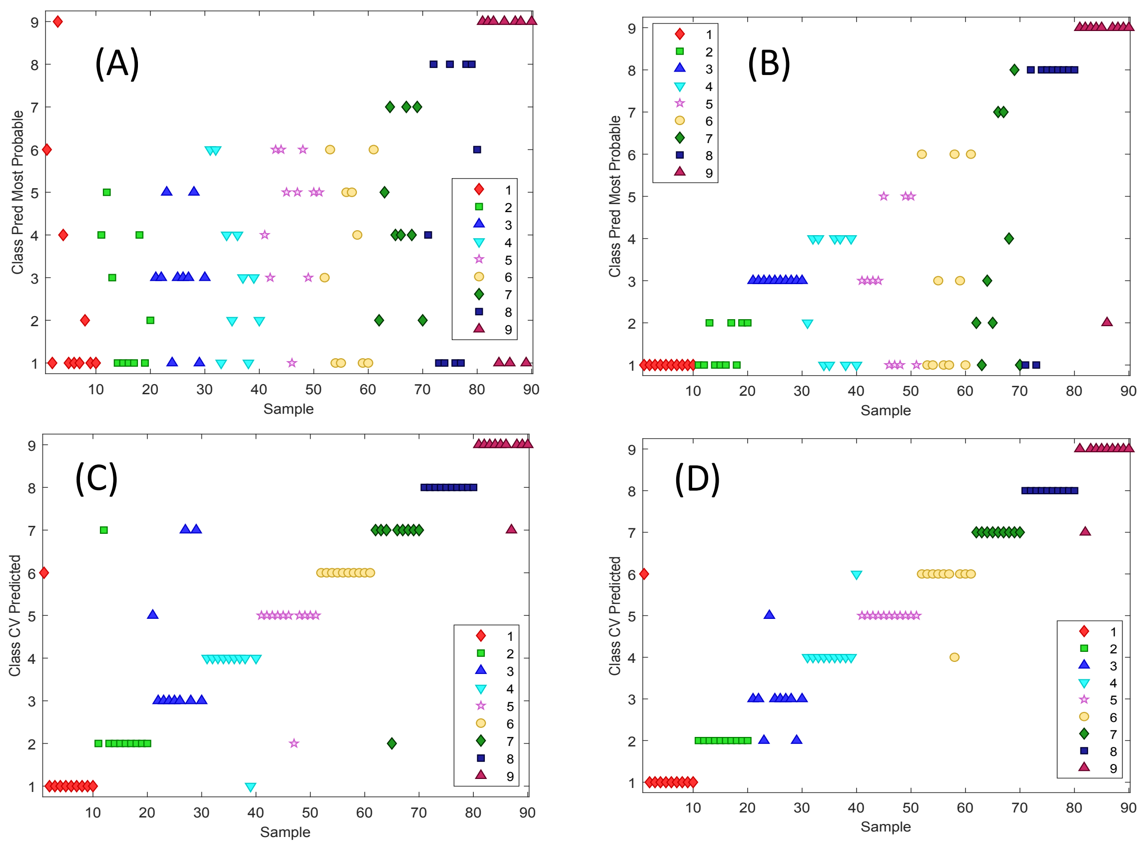


Figure 2. k-Nearest Neighbors plots showing (A) a KNN model of the raw data, (B) a KNN model of the preprocessed data with no variable selection, (C) a KNN model of the preprocessed data with a mean intensity cutoff greater than 0.5, and (D) a KNN model of the preprocessed data with a mean intensity cutoff greater than 1.0.

FIGURE 3

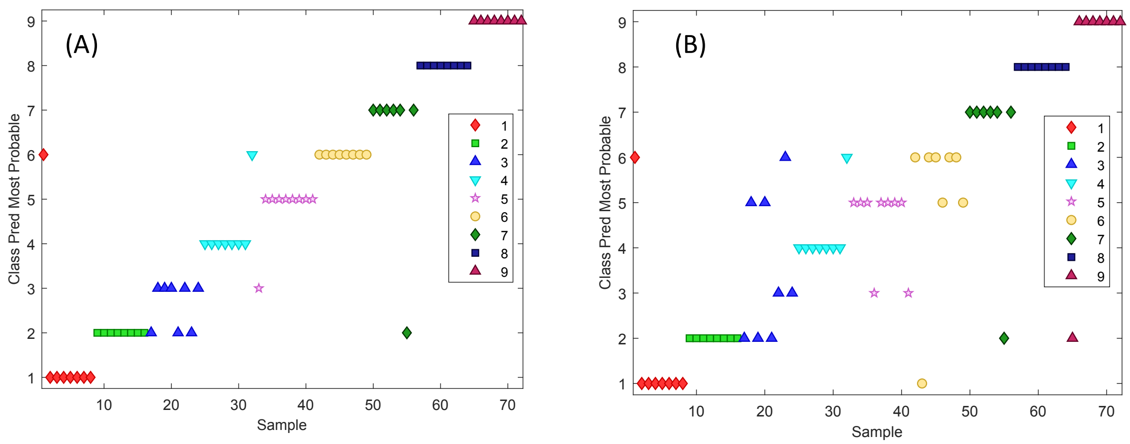
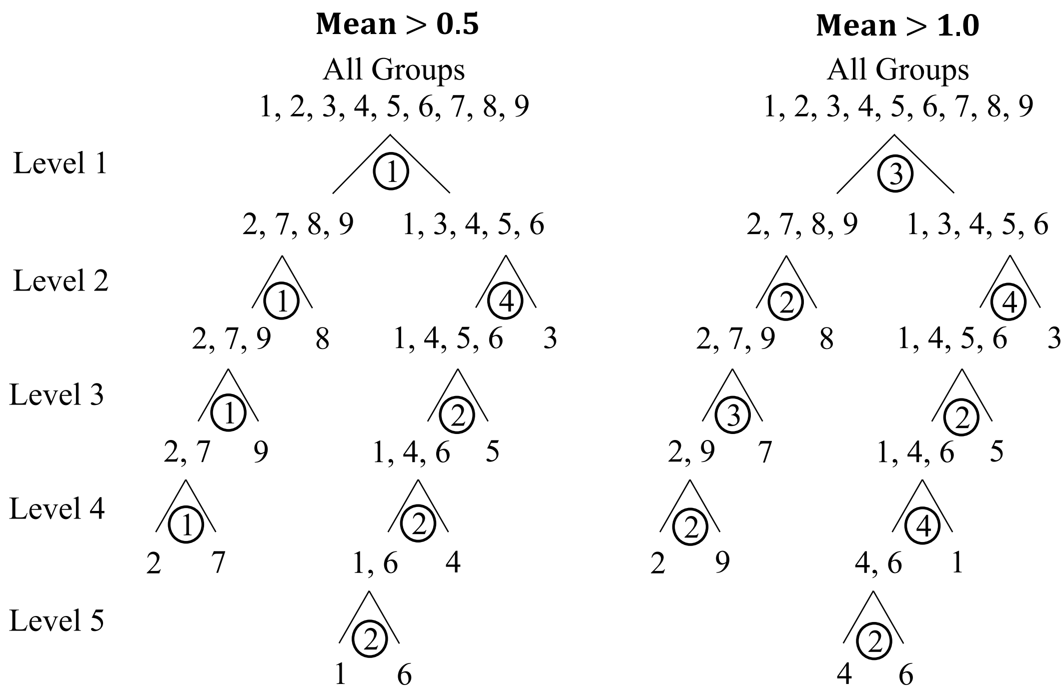
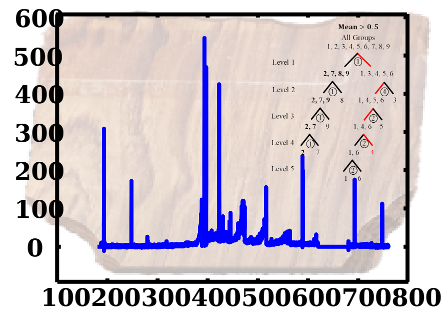


Figure 3. k-Nearest Neighbor models of the training sets (~8 samples/class) for (A) a mean intensity cutoff greater than 0.5 and (B) a mean intensity cutoff greater than 1.0.

SCHEME 1



Scheme 1. Decision tree model for calibration, training, and validation sets at both variable selection cutoffs. Circled number indicates number of latent variables used at each branch for the PLSDA model.



Seven *Dalbergia* and two *non-Dalbergia* hardwood species were successfully differentiated with PLS-DA and KNN chemometric models of LIBS spectra.