Ciência e Agrotecnologia
SCIENCE AND AGROTECHNOLOGY
Ciência e Agrotecnologia, 46:e001422, 2022
http://dx.doi.org/10.1590/1413-7054202246001422
Agricultural Sciences
eISSN 1981-1829

Near infrared spectroscopy as a tool to discriminate tannins from Amazonian species

Espectroscopia do infravermelho próximo como ferramenta para discriminar taninos de espécies amazônicas

Cristiano Souza do Nascimento[1*], Roberto Daniel de Araújo[2], Claudia Eugênio da Silva[3], Claudete Catanhede do Nascimento[2], Valdiek da Silva Menezes[1], Joaquim dos Santos[2]

[1]Instituto Nacional de Pesquisas da Amazônia/INPA, Programa de Pós-Graduação em Ciências de Florestas Tropicais/PPGCFT, Manaus, AM, Brasil
[2]Instituto Nacional de Pesquisas da Amazônia/INPA, Manaus, AM, Brasil
[3]Instituto Nacional de Pesquisas da Amazônia/INPA, Programa de Pós-Graduação em Botânica/PPGBOT, Manaus, AM, Brasil
*Corresponding author: s-nascimento@hotmail.com
Received in January 23, 2022 and approved in May 3, 2022

ABSTRACT

Near infrared spectroscopy (NIR) is a tool capable of providing efficient results for organic molecules of different materials. We developed a predictive model using Fourier Transform NIR Spectroscopy to distinguish the types of tannins in different forest species in the Amazon. Samples were obtained from different regions of the State of Amazonas/Brazil, and tests for tannins were performed, including obtaining NIRS spectra. The assembly of spectral data matrices versus analytes of interest was crossed with the results of traditional analyses. In addition, a calibration and validation set was constructed for condensed tannins, hydrolyzable tannins, and samples with no tannins. Finally, the performance of classification models was evaluated for sensitivity, identification index, and errors. The condensed tannin classes were detected in 63% of the species studied, followed by 34% of the species not containing tannin. The discriminant analysis produced groupings of classes, with a hit sensitivity index >90%. The developed model can be applied in studies of ecology, forestry and chemotaxonomy, with a focus on phenolic compounds such as tannins. The proposed methodology has advantages over the reference methods, reflected as a lower need for sample preparation, shorter analysis time, no use of reagents, and, consequently, no generation of waste.

Index terms: Condensed tannins; Amazon woods; NIRS; discriminant analysis; non-destructive methodology.

RESUMO

A espectroscopia no infravermelho próximo (NIR) é uma ferramenta capaz de fornecer resultados eficientes para moléculas orgânicas de diversos materiais. O estudo desenvolveu modelo preditivo com a espectroscopia NIR com Transformada de Fourier**,** para distinguir taninos em diferentes espécies florestais da Amazônia. Foram obtidas amostras de diferentes regiões do Estado do Amazonas/Brasil, testes para taninos e obtenção de espectros NIRS foram realizados. A montagem de matrizes de dados espectrais, versus analitos de interesse foram cruzados com os resultados das análises tradicionais. Um conjunto de calibração e validação foi construído para taninos condensados, taninos hidrolisáveis e amostras sem taninos. Ao final, avaliou-se o desempenho dos modelos de classificação quanto à sensibilidade, índice de identificação e erros. Em 63% das espécies estudadas, foi detectada a classe de tanino condensado, seguida de 34% para espécies que não continham tanino. A análise discriminante produziu agrupamentos com índice de sensibilidade de acerto >90%. O modelo desenvolvido pode ser aplicado em estudos de ecologia, silvicultura e quimiotaxonomia, com foco em compostos fenólicos como taninos. A metodologia proposta neste trabalho apresenta vantagens sobre os métodos de referência, destacando-se a menor necessidade de preparo das amostras, menor tempo de análise, não utiliza-se reagentes e consequentemente, não gera resíduos

Termos para indexação: Taninos condensados; madeiras da Amazônia; NIR; análise discriminante; metodologia não destrutiva.

INTRODUCTION

The Amazon is known for having the largest rainforest area. One of the last taxonomic surveys in the region recorded 14,003 species, 1,788 genera, 188 families of plant species (Cardoso et al., 2017). Studies of ecology and botany require know ledge of the chemical nature of the flora, in addition to industrial applications and biotechnological processes (Nascimento et al., 2021).

Tannins are polyphenols that are widely present in plants. They are present in abundance in tropical; certain botanical families such as Anacardiaceae, Leguminosae, Myrtaceae and Polinaceae are rich in this compound (Coq

et al., 2010; Simões et al., 2017). Hydrolyzable tannins (gallic and ellagic acid) provide resistance to plants against herbivores, whereas condensed tannins (catechin and proanthocyanidin) guarantee protection against pathogenic microorganisms; thus, strengthening the natural durability of the species. Industrially, tannins are used in the manufacture of leather, beverages, and wood adhesives (Monteiro et al., 2005; Grasel; Ferrão, 2016). Tannins are generally characterized by analytical methods such as high-performance liquid chromatography (HPLC), nuclear magnetic resonance (NMR) (13C), matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF), among others. These procedures, although efficient, are laborious and costly and use harmful solvents and colorimetry equipment. In addition, these, techniques are performed by the wet method and require extensive sample preparation, several reagents, glassware, and equipment. When the end activity targets large numbers of samples, studies are often inaccessible, necessitating the need for novel tools (Barbosa et al., 2006; Ricci et al., 2015).

Non-destructive methodologies allow estimating several of the plants material, without changing their structure, preserving the sample, and thus not compromising their use. The majority of these methods are indirect methods that are based on the investigation of correlations and adjustment of calibration models between the properties of interest and others that are easier to measure. The development of sophisticated statistical analyses and reliable software has led to the use of spectroscopy in the near-infrared region (NIRS) in several areas such as agriculture, forensic sciences and industry (Menezes et al., 2014; Souza et al., 2017).

NIR spectroscopy is an analytical tool that provides efficient results to analyze organic molecules quantitatively. It has been used in several industrial applications and scientific research, generating robust results in ecology, botany, and forestry (Ono; Hiraide; Amari, 2003; Tsuchikawa; Schwanninger, 2013). In the near-infrared, photon energy vibration is used, 13,000 to 4,000 cm^{-1}; more efficient spectroscopes are coupled to the Fourier transform (FT) (Pasquini, 2018). Because NIRS modelling allows spectral correlation with different wood properties, it is possible to estimate its characteristics rapidly, without destroying the samples and without using reagents (Fernandes et al., 2017; Nascimento et al., 2021).

Braga et al. (2011) used NIRS to characterize tropical woods, and their models presented low errors of discrimination, confirming the efficiency of this tool. Similarly, Tigabu et al. (2018) used near-infrared for studying the chemical composition of *Betula* sp. (polysaccharides, proteins, and fatty acids) and obtained 99% accuracy.

Falcão and Araujo (2011) confirmed the robustness of infrared spectroscopy (attenuated total reflectance) to distinguish different types of tannins in leather tanning. This uniqueness was possible from a complete comparison of the spectral profile of the samples with the reference tannin. Results from studies with NIRS state that their spectra can be considered as a "spectral signature" that can accurately the identify the investigated raw material (Durgante et al., 2013; Souza et al., 2017). In this context, we developed a predictive model of FT-NIR to distinguish tannins that occur in the xylem tissue (wood) of different forest species in the Amazon region.

MATERIAL AND METHODS

Forest species were obtained from different regions of the state of Amazonas, Brazil. For each species, wedges (heartwood-bark sense) and replicates were obtained from different individuals, totaling 100 species (Figure 1 and Table 1). Specimens in dimensions 20 × 20 × 30 mm were analyzed macroscopically and compared with material obtained from the Botanical Collection, namely, xylotheque/PCAC/INPA (Freitas; Vasconcellos, 2019), by the specialist J. A. Freitas to confirm the species, which were later registered and deposited in the collection.

The samples were broken into smaller pieces (Band saw-Videira), perforated (Chipper Pallmann-PZ8), and crushed (Willey mill) to obtain sawdust. Subsequently, the sawdust was sieved with a set of 0.84, 0.42, 0.25, and 0.18 mm screens (Ro-Tap/Testing Sieve Shaker, Model B). The standard granulometry for wet and spectroscopic analysis range from 0.41 to 0.25 mm, as recommended by the American Society for Testing and Materials (ASTM) (2021) for the chemical analysis of wood. The fraction standard was divided into two portions, namely portion 1 for obtaining hydroalcoholic extracts and detecting types of tannins and portion 2 for obtaining NIR spectra.

NIR spectra

Spectra were obtained from 10.00 g of wood sawdust (0.41–0.25 mm) in 'sample cup spinner' the FT-NIR Antaris II Thermo Scientific system. The data were collected in the RESULT software in the region between 10,000 and 4,000 cm^{-1} (resolution of 8 cm^{-1}, 96 scans, 16 scan spectrum/sample, automatic background internal), where each reading/sample was performed

in triplicate, totaling 996 spectra that were used for multivariate analysis. This step highlights the importance of standardizing the granulometry of samples, as well as the humidity control (0.41–0.25 mm; relative humidity < 60% at 20 °C), which reduces the possible effects of instrumental deviations during digitization (Nascimento; Varejão; Vianez, 2012; Silva et al., 2013).

Reference method for detecting of tannin types

Different types of tannins in forest species were detected by wet methodology (Matos, 2009; Varejão et al., 2012) using the catechin P.A., Merck (condensed tannins), and gallic acid P.A., Merck (hydrolyzable tannins) as the standards. Hydroalcoholic extracts were obtained from 10.00 g of each sample (100 mL of the solvent) in an ultrasonic bath (Unique) at 65 °C for 60 min. Subsequently, the extracts were analyzed for the presence/absence and type of tannins (condensed or hydrolyzable).

I: In a test tube, 2 mL of the extract and three drops of the alcoholic solution of $FeCl_3$ (10%) were added, shaken, and subsequently, a possible color variation or abundant precipitate was observed; II: In a round bottom flask, 25 mL of the extract was added, plus 10 mL of CH_2O + HCl solution (2:1). The solution was heated (90 °C) under reflux for 30 min. For both tests, a brown precipitate and a greenish solution were formed, confirming the presence of condensed tannins (pyrocatechol tannin). In addition, a solution with bluish tones was formed, confirming hydrolyzable tannins (pyrogallic tannin);

III: In a test tube, 2 mL of the extract and 2 mL of $Pb(C_2H_3O_2)_2$ solution (10%, neutral) were added, and stirred. The presence of hydrolyzable tannins was confirmed by the formation of flocculant precipitates (white salt). These tests were compared with a blank test performed with water and standard catechin (condensed tannin) and gallic acid (hydrolyzable tannin).

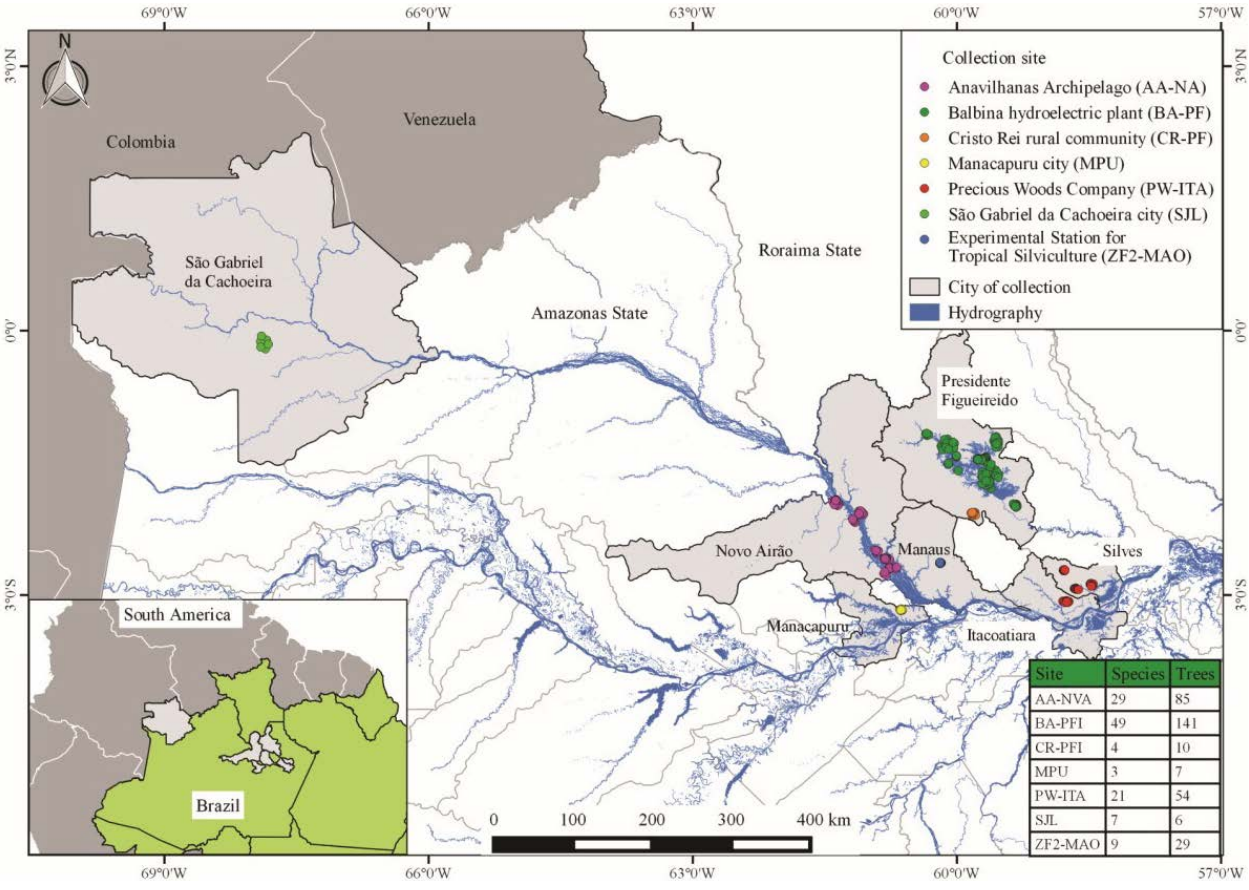

Figure 1: Geographic area of collections of forest species in the Amazon.

Table 1: Amazonian Forest species used in the study.

No.	Scientific name (popular name) - Family	Number of trees
1.	*Acosmium nitens* (Vogel) Yakovlev (Taboarana) - Fabaceae	03
2.	*Albizia inundata* (Mart.) Barneby & J.W. Grimes (Ingarana-da-várzea) - Fabaceae	03
3.	*Albizia subdimidiata* (Splitg.) Barneby & J.W. Grimes (Faveira-do-igapó) - Fabaceae	03
4.	*Aldina heterophylla* Benth. (Macucu-de-paca) - Fabaceae	05
5.	*Andira parviflora* Ducke (Sucupira vermelha) - Fabaceae	03
6.	*Aniba canelilla* (Kunth) Mez (Preciosa) - Lauraceae	03
7.	*Aniba duckei* Kostermans (Pau rosa) - Lauraceae	03
8.	*Aniba* sp. (Louro rosa) – Lauraceae	03
9.	*Aspidosperma album* (Vahl) Benoiste x Pichon (Piquiá marfim) - Apocinaceae	06
10.	*Aspidosperma obscurinervium* Azambuja (Carapanaúba) - Apocinaceae	03
11.	*Brosimum parinarioides* subsp. *amplicoma* (Ducke) C.C. Berg (Amapá doce) - Moraceae	03
12.	*Brosimum rubescens* Taub. *(*Pau rainha) - Moraceae	03
13.	*Buchenavia oxycarpa* (Mart.) Eichler. (Tanimbuca) - Combretaceae	03
14.	*Byrsonima crispa* A. Juss. (Murici) - Malphigiaceae	04
15.	*Calophyllum brasiliense* Cambess. (Jacareúba) - Clusiaceae	03
16.	*Campsiandra chigo-montero* Stergios (Acapurana) - Fabaceae	03
17.	*Cariniana integrifolia* Ducke (Tauari da Amazônia) Lecythidaceae	03
18.	*Caryocar glabrum* subsp. *parviflorum* Prance & M.F. Silva (Pequiarana) - Caryocaraceae	05
19.	*Caryocar* sp. (Pequi da Amazônia) - Caryocaraceae	03
20.	*Caryocar villlosun* (Aubl.) Pers. (Pequiá) - Caryocaraceae	03
21.	*Catostemma sclerophyllum* Ducke (Castanha de paca) - Malvaceae	03
22.	*Cedrela odorata* L. (Cedro) – Meliaceae	03
23.	*Ceiba pentandra* (L.) Gaertn. (Sumaúma) - Malvaceae	03
24.	*Clarisia racemosa* Ruiz & Pav. (Guariuba) - Moraceae	07
25.	*Clathrotropis nitida* (Benth.) Harms. *(*Acapu-do-igapó) - Fabaceae	03
26.	*Crudia amazonica* Benth. (Lombrigueiro) - Fabaceae	03
27.	*Cynometra spruceana* var. *phaselocarpa (*Hayne) Dwyer (Castanha-de-burro) - Fabaceae	03
28.	*Didymopanax morototoni* (Aubl.) Decne. & Planch. (Morototó) - Araliaceae	03
29.	*Dinizia excelsa* Ducke (Angelim pedra) - Fabaceae	06
30.	*Diplotropis martiusii* Benth. (Sucupira preta) - Fabaceae	03
31.	*Diplotropis* sp. (Sucupira) – Fabaceae	03
32.	*Dipteryx odorata* (Aubl.) Willd. (Cumaru) - Fabaceae	08
33.	*Dipteryx polyphylla* Huber (Cumarurana) - Fabaceae	03
34.	*Duckeodendron cestroides* Kuhlm. (Pupunharana) - Solanaceae	03
35.	*Emmotum fagifolium* Desv. ex Ham. (Pau de remo) - Icacinaceae	02
36.	*Enterolobium schomburgkii* (Benth.) Benth. (Sucupira amarela) - Fabaceae	03
37.	*Erisma uncinatum* Warm. (Quarubarana) - Vochysiaceae	03
38.	*Eschweilera coriacea* (DC.) S.A. Mori (Mata-matá) - Lecythidaceae	03
39.	*Eschweilera odora* (Poepp. ex O. Berg) Miers (Mata-matá preto) - Lecythidaceae	04

Continue...

Table 1: Continuation.

No.	Scientific name (popular name) - Family	Number of trees
40.	*Eschweilera truncata* A.C. Sm. (Mata-matá) - Lecythidaceae	02
41.	*Euxylophora paraenses* Huber (Pau amarelo) - Rutaceae	02
42.	*Goupia glabra* Aubl. (Cupiúba) – Goupiacea	03
43.	*Guarea trichilioides* L. (Gitó) – Meliaceae	03
44.	*Handroanthus serratifolius* (Vahl) S.O. Grose (Pau - D'Arco) - Bignoniaceae	03
45.	*Heterostemon mimosoides* Desf. (Haiari) - Fabaceae	03
46.	*Hura creptans* L. (Assacu) – Euphobiaceae	03
47.	*Hymenaea courbaril* L. (Jatobá) – Fabaceae	05
48.	*Hymenaea intermedia* Ducke (Jatobá) - Fabaceae	03
49.	*Hymenaea parvifolia* Huber (Jutaí) - Fabaceae	03
50.	*Hymenolobium excelsum* Ducke (Angelim da mata) - Fabaceae	04
51.	*Inga disticha* Benth. (Ingá-chinelo) - Fabaceae	03
52.	*Iryanthera* sp. (Ucuuba vermelha) - Myristicaceae	03
53.	*Iryanthera tricornis* Ducke (Ucuuba puna) - Myristicaceae	03
54.	*Jacaranda copaia* (Aubl.) D. Don. (Caroba) - Bignoniaceae	03
55.	*Licaria aritu* Ducke *(*Louro aritu) - Lauraceae	03
56.	*Licaria canella* (Meisn.) Kosterm. (Louro chumbo) - Lauraceae	03
57.	*Macrolobium acaciifolium* (Benth.) Benth. (Arapari) - Fabaceae	03
58.	*Macrolobium angustifolium* (Benth.) Cowan (Arapari) - Fabaceae	03
59.	*Manilkara amazanica* (Huber) A. Chev. (Maçaranduba) -Sapotaceae	03
60.	*Manilkara bidentata (*A.DC.) A. Chev. (Maçaranduba) -Sapotaceae	03
61.	*Manilkara huberi* (Ducke) Standl. (Maçaranduba) -Sapotaceae	03
62.	*Maquira guianensis* Aubl. (Muiratinga) - Moraceae	03
63.	*Marmaroxylon racemosum* (Ducke) Record (Angelim rajado) - Fabaceae	03
64.	*Martiodendron elatum* (Ducke) Gleason (Muirapixuna) - Fabaceae	03
65.	*Mezilaurus ita-uba* (Meisn.) Taub. exMez (Itaúba) - Lauraceae	03
66.	*Micrandropsis scleroxylon* (W.A. Rodrigues) (Piãozinho) - Euphorbiaceae	05
67.	*Minquartia guianensis* Aubl. (Aquariquara) Olacaceae	03
68.	*Mora paraensis* (Ducke) Ducke (Pracu*úba*) - Fabaceae	03
69.	*Nectandra cuspidata* Nees & Mart. (Louro preto) Lauraceae	03
70.	*Ochroma pyramidale* (Cav. ex Lam.) Urb. (Pau de balsa) - Malvaceae	03
71.	*Ocotea cymbarum* Kunth (Louro chumbo) - Lauraceae	03
72.	*Osteophloeum platyspermum* (Spruce ex A.DC.) Warb. (Ucubarana) - Miristicaceae	03
73.	*Panopsis rubescens* (Pohl) Pittier (Pau de rato) - Proteaceae	03
74.	*Parkia discolor* Benth. (Bico-de-arara) - Fabaceae	03
75.	*Peltogyne venosa* subsp. *densiflora* (Benth.) M.F. Silva (Pau roxo) - Fabaceae	03
76.	*Pentaclethra macroloba* (Willd.) Kuntze (Paracaxi) - Fabaceae	03
77.	*Piptadenia suaveolens* Griseb. *(*Faveira) - Fabaceae	03
78.	*Platonia insignis* Mart. (Bacuri Açu) - Clusiaceae	03

Continue...

No.	Scientific name (popular name) - Family	Number of trees
79.	*Platymiscium ulei* Harms (Macacaúba) - Fabaceae	03
80.	*Pouteria guianensis* Aubl. (Abiurana) Sapotaceae	03
81.	*Protium paraense* Cuatrec. (Breu) - Burseraceae	03
82.	*Protium puncticulatum* J.F. Macbr. (Breu vermelho) - Burseraceae	04
83.	*Qualea paraenses* Ducke (Mandioqueira) - Vochysiaceae	04
84.	*Roupala montana* Aubl. (Louro faia) - Proteaceae	03
85.	*Sacoglottis guianensis* Benth. (Uxi-verdadeiro) - Humiriaceae	03
86.	*Scleronema micranthum* (Ducke) Ducke (Cardeiro) - Malvaceae	04
87.	*Sextonia rubra* (Mez) van der Werff (Louro gamela) - Lauraceae	04
88.	*Simarouba amara* Aubl. (Marupá) - Simaroubaceae	08
89.	*Stryphnodendron guianense* (Aubl.) Benth. (Barbatimão) - Fabaceae	03
90.	*Swartzia argentea* Benth. (Acapu do igapó) - Fabaceae	03
91.	*Swartzia laevicarpa* Amshoff *(*Saboarana) - Fabaceae	03
92.	*Swartzia macrocarpa* Benth. (Muirapixuna) - Fabaceae	03
93.	*Swartzia panacoco* (Aubl.) Cowan (Coração de negro) - Fabaceae	03
94.	*Swartzia polyphylla* DC. (Arabá) - Fabaceae	03
95.	*Tachigali paniculata* Aubl. (Tachi preto) - Fabaceae	03
96.	*Vatairea guianensis* Aubl. (Fava-bolacha) - Fabaceae	03
97.	*Virola* sp. (Ucuúba vermelha) - Miristicaceae	03
98.	*Virola surinamensis* (Rol. ex Rottb.) Warb. (Ucuúba) - Miristicaceae	03
99.	*Vismia guianensis* (Aubl.) Pers. (Lacre) - Hypericaceae	03
100.	*Vochysia maxima* Ducke (Quaruba) -Vochysiaceae	03
	Total samples	332

Data processing and multivariate analysis

The TQ Analyst software was used for multivariate analysis. Before this examination, the usual pre-processing technique, multiplicative scatter correction (MSC), was performed to reduce the influence of several sources not related to the physical or chemical information carried by the raw spectra. Spectra were studied to evaluate the similarity and clustering tendency. Subsequently, the data were subjected to principal component analysis (PCA).

The assembly of the spectral data matrices versus the analyte of interest (classes of tannins) was crossed with the results of the reference method, and calibration/validation sets were constructed for condensed tannins (CT), hydrolyzable tannins (HT), and samples without the presence of tannins (WT).

The NIR spectral range was selected by the software, which configured a region capable of covering the entire spectral zone of its standards. The classification model developed consisted of a discriminant analysis that used the Mahalanobis algorithm (distance) to indicate the grouping between samples. This algorithm works as a metric that determines the distance between a vector and a distribution. The principle is that an observation is assigned to the class that is closest to the base on the Mahalanobis distance.

All calibration and validation were developed in sets of spectral data centered on the average of each species, where the sample universe was composed of 75% (238 samples) of the species for calibration and 25% (94 samples) of species for validation (Table 2). Random selection was performed to define the species selected for each model.

The performance of the classification models was assessed using the following classification parameters:

- Sensitivity $Sn = \frac{TP}{TP + FN} \times 100$, where TP (true positive) = number of samples determined correctly in the class; FN (false negative) = It is the number of samples determined incorrectly in the class or does not belong to the class.

- Calibration error $CE = \frac{Ecal}{Tcal} \times 100$, where ECal = number of samples classified wrong in the calibration set; TCal = Total of calibrated samples.

- Validation error $VE = \frac{EVal}{TVal} \times 100$, where EVal = number of samples classified wrong in the validation set; TVal = Total sample validated.

- Maximum classification error $MCE = \frac{\sum E}{CA} \times 100$, where $\sum E$ = sum of errors (ECal+EVal); CA = total sample set.

- Identification Index $II = \frac{N1}{N2} \times 100$, where N1 = number of samples correctly identified in the validation set; N2 = total number of samples used in the validation set.

Using the PAST software version 4.08 (Hammer; Harper; Ryan, 2001), cluster analysis (CA) was developed, pairing the data using the Cosine algorithm, which evaluated the similarity between the botanical families (universes of the calibrated and validated samples) and the tannin classes.

RESULTS AND DISCUSSION

This study presents a model for the classification of tannins present in the wood. The spectral universe analyzed is shown in Figure 2, where it is possible to observe the spectral similarity. The overlap between the spectra indicates the chemical composition of the species (characteristic CH and OH bonds), with the need to pre-process the spectra. The multivariate calibration detected possible differences, resulting in the extraction of useful parameters from each sample. It is possible to detect a small difference in the absorption in the bands of 4,500–4,000, 6,100–5,500, and 7,500–7,000 cm^{-1} even without spectral treatment. The spectral region 9,882–4,292 cm^{-1} was selected because this band represents chemical properties that can be used to discriminate classes of tannins in the Amazonian woods.

Grasel and Ferrão (2016) used the NIRS tool to discriminate types of tannins in commercial tannic extracts. They observed that regions 6,250–5,555 and 4,545–4,116 cm^{-1} were related to absorption in the first overtone of the bands CH, CH_2, and CH_3. In contrast, Teye et al. (2013) indicated the region of 9,000–5,000 cm^{-1}, the carbonyl groups, CH elongation, and the deformations CH, SH, NH, CH_2, and CH_3, corresponding to the bonds of structures polyphenolics, alkaloids, and terpenes. The highest absorption peaks × wavelength (cm^{-1}) characteristics can provide information for the classification of tannins. This explanation may open possibilities for groupings for condensed tannins, hydrolyzable tannins and without tannins; however, it may also be related to the species, the so-called "spectral signature" or "fingerprint" of the material (Ricci et al., 2015; Souza et al., 2017).

Principal Component Analysis (PCA)

Using raw spectra (without treatments), the principal component analysis (PCA) was used to distinguish the types of tannins in the Amazonian species. For that, linear combinations of the spectral variables, absorbance × wavelength, were used. The computational analysis generated 10 PCs, with PC1 being responsible for the greatest variation in the dataset. An initial observation of variables indicated a possible separation of the species in terms of types of tannins.

The sum of PC 1 × PC 2 × PC 3 describes 98.84% of the 100 species used in the study (Figure 3). Grouping of the spectral species showed PCA graphs with a tendency for grouping of tannin classes. In general, species "without tannins" showed a behavior of agglomerating in the central axis of the score, and the other groups (condensed and hydrolyzable tannins) were initially dispersed.

The PC1 algorithm describes the maximum of the spectral information that is used in predictions/modeling. According to Luna et al. (2013), discrimination above 80% carries most of the chemical composition information in the NIR region. The xylem tissue studied here presented considerable differences in chemical properties from the number of primary metabolites such as cellulose, hemicellulose, and lignin, as well as the quantity/quality of the extractives (secondary metabolic). Discriminant analysis (DA) indicated the differentiation of variables (Figure 3).

Table 2: Summary of the chemometric parameters used in the FT-NIR modeling.

Chemometric method	Spectral regions of analysis	Spectral treatment	Smoothing	Data set
Discriminant analysis	9,882–4,292 cm^{-1}	None 1st derivative 2nd derivative	MSC; Savitzky-Golay (Data point = 7; Polyn. order = 5); Norris (Segm. length = 3; gap = 2)	Calibration = 238 samples (714 spectra) Validation = 94 samples (282 spectra)

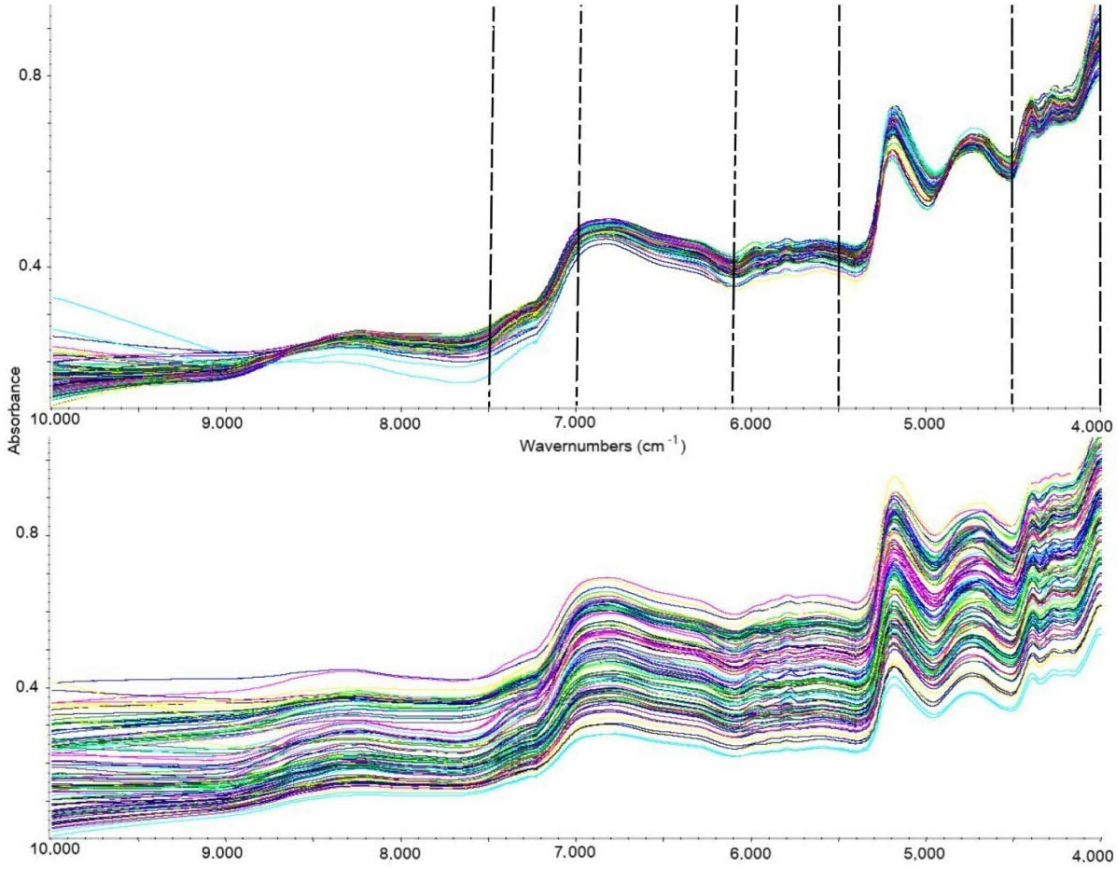

Figure 2: Sample universe with an averages spectra of 100 species (332 samples) of wood from the Amazonian region in the range of 10,000-4,000 cm^{-1}. At the bottom are raw spectra, and at the top are pre-processed spectra with multiplicative scatter correction (MSC) with Savitzky-Golay filters.

Types of tannins in the Amazonian woods

The reference method (wet method) was used to detect the types of tannins present in the Amazonian woods. Out of 100 species studied, 63 had condensed tannins, and three contained hydrolyzable tannins (Table 3).

Using the same reference as the study (wet method), Barbosa et al. (2006) detected condensed tannins in 100% of Fabaceae tree species. Another study by Varejão et al. (2009) reached the qualification of 70% of the species with condensed tannins; however, the researchers studied the precision of the technique because concentrations < 10 ppm can make it difficult to detect tannins in reactive tests (colorimetric). In addition, Coq et al. (2010) and Oliveira et al. (2010) performed research on tannins from tropical species included in the present study, such as *Caryocar glabrum, Goupia glabra, Hymenaea courbaril, Peltogyne venous, Platonia insignis, Simarouba amara*, and *Virola surinamensis*.

Tannins produced by plants provide biological protection, defense against solar radiation, and regulate the water flow (Dominy; Lucas; Wright, 2003; Taiz et al., 2017). Species in which tannins were not detected (Table 3) were *Ceiba pentadra*, *Didymopax morototoni*, *Ematum fagifolium*, *Hura creptans*, and *Simarouba amara*, among others. These are woods of low density, white in color, and with low natural resistance to xylophagous organisms.

Discriminant Analysis (DA)

The central question of this study was: Is it possible to use FT-NIR spectroscopy to discriminate the type of tannins present in the xylem tissue? After the PCA results were obtained and tannins were detected, the NIR spectra were used as the data input for multivariate classification. The result of DA (25/75) generated a predictive model with 99.9% of described variability and a 98% accuracy rate when spectra without derivation were used.

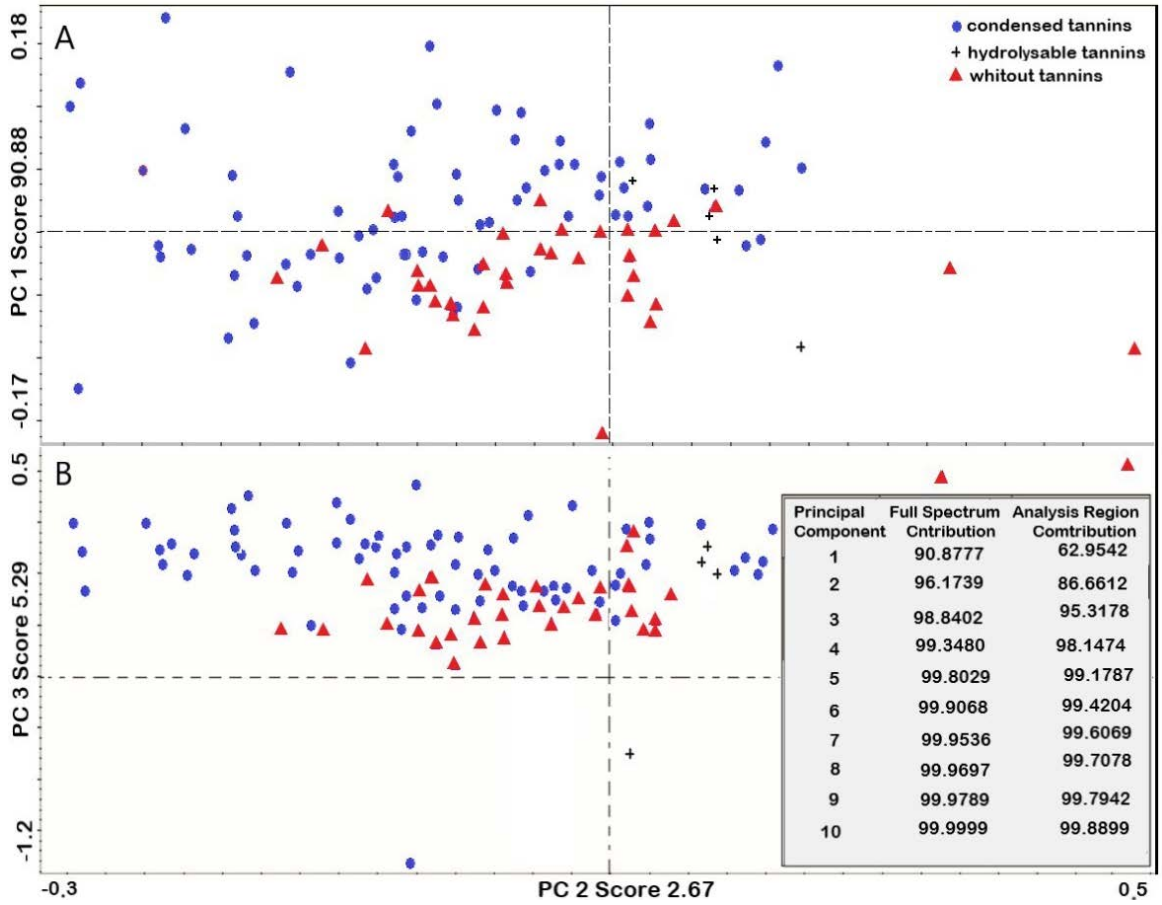

Figure 3: Results of the principal component analysis: A: PC 1 × PC 2; B: PC 3 × PC 2.

The resulting DA results separated and classified the condensed tannins (CT), hydrolyzable tannins (HT), and without tannins (WT) adequately for most species. Figure 4 shows the grouping calculated using the Mahalanobis distance (DM) with the calibration dataset. DM is an algorithm that can explain the grouping or removal of individuals. The species *Enterolobium schomburgkii* (Fabaceae) and *Maquira guianensis* (Moraceae) were outside the CT group, although close to other species of this class.

Table 4 presents the results of sensitivity, errors, and II when applied mathematically. The best model is indicated when there are low errors and high indices of identification and sensitivity. In a general analysis, the best classification model is the one that used only the Savitzky–Golay filter in the set of spectra, followed by the model with second derivative spectral treatment, and Norris filter, because these are within the acceptable limit for non-destructive prediction, which would be sensitivity and II,> 90% and errors (ECal, Eval, and MCE) < 10%.

The efficiency of FT-NIRS in discriminating the wood with and without tannins is associated with the chemical and physical properties of samples, in which DA functions as a pattern recognition tool (Teye et al., 2013). According to Ricci et al. (2015), the spectral range 6,250–4,500 cm^{-1} is characterized by the band associated with the -OH bond (aromatic structure) attributed to condensed tannins. Allied to this parameter, the exploratory analysis of the PC describes the variations in the spectra with possible chemical concentrations, thereby potentializing the separation of the groups.

Morozova, Elizarova, and Pleteneva (2013) performed studies with NIR spectroscopy in conjunction with DA and DM to explain the correct grouping of their samples. They performed a survey of scientific papers (1987–2010) that demonstrated the efficiency of the Mahalanobis distance in the distance-based pattern recognition method and concluded that the DA algorithm presents reliable relationships when considering the qualitative characteristics of samples.

Table 3: Results of tests to detect types of tannins in Amazonian woods.

Class	Species	Number of sample
Condensed tannins	*Acosmium nitens; Albizia inundata; A. subdimidiata; Aldina heterophylla; Andira parviflora; Aspidosperma album; A. obscurinervium; Brosimum rubescens; Buchenavia oxycarpa; Calophyllum brasiliense; Campsiandra chigo-montero; Caryocar* sp.; *C. villlosun; Catostemma sclerophyllum; Cedrela odorata; Clathrotropis nitida; Crudia amazanica; Cynometra spruceana; Dinizia excelsa; Diplotropis* sp.*; D. martiusii; Dipteryx odorata; D. polyphylla; Enterolobium schomburgkii; Eschweilera coriacea; E. truncata; Goupia glabra; Guarea trichilioides; Handroanthus serratifolia; Heterostemon mimosoides; Hymenaea courbaril; H. intermedia; H. parvifolia; Hymenolobium excelsum; Inga disticha; Macrolobium acaciifolium; M. angustifolium; Manilkara huberi; Maquira guianensis; Marmaroxylon racemosum; Mezilaurus ita-uba; Minquartia guianensis; Mora paraensis; Ocotea cymbarum; Parkia discolor; Peltogyne venosa; Pentaclethra macroloba; Piptadenia suaveolens; Platonia insignis; Platymiscium ulei; Pouteria guianensis; Protium paraense; P. puncticulatum; Scleronema micranthum; Sextonia rubra; Stryphnodendron guianense; Swartzia argentea; S. laevicarpa; S. macrocarpa; S. panacoco; S. polyphylla; Tachigali paniculata; Vatairea guianensis.*	Species 63 Botanical genus 49 Family 13
Hydrolyzable tannins	*Cariniana integrifolia*; *Eschweilera odora*; *Manilkara bidentata*.	Species 3 Botanical genus 3 Family 2
Without tannins	*Aniba* sp.*; A. canelilla; A. duckei; Brosimum parinarioides; Byrsonima crispa; Caryocar glabrum; Ceiba pentandra; Clarisia racemosa; Didymopanax morototoni; Duckeodendron cestroides; Emmotum fagifolium; Erisma uncinatum; Euxylophora paraensis; Hura creptans; Iryanthera* sp.*; I. tricornis; Jacaranda copaia; Licaria aritu; L. canella; Manilkara amazonica; Martiodendron elatum; Micrandropsis scleroxylon; Nectandra cuspidata; Ochroma pyramidale; Osteophloeum platyspermum; Panopsis rubescens; Qualea paraensis; Roupala montana; Sacoglottis guianensis; Simarouba amara; Virola* sp.*; V. surinamensis; Vismia guianensis; Vochysia maxima*.	Species 34 Botanical genus 29 Family 17*

* Five families with condensed tannins and without tannins.

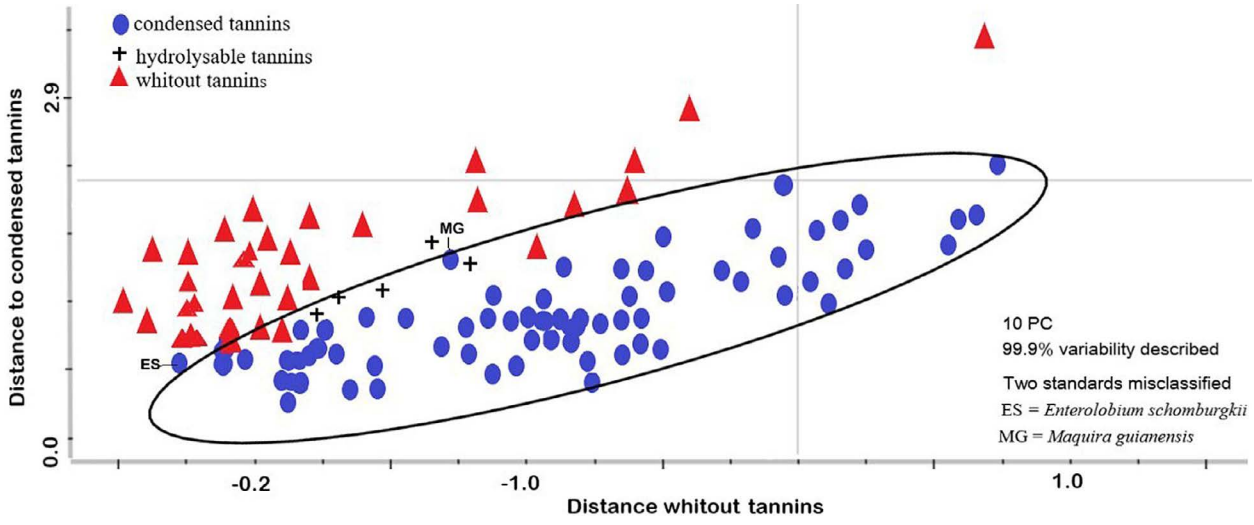

Figure 4: Results of the discriminant analysis of 100 Amazonian woods regarding the tannin class.

Table 4: Evaluation parameters of chemometric models.

Treatment spectral	Sensitivity (%)	Error of calibration (%)	Error of validation (%)	Error total (%)	Identification index (%)
Smoothing Savitzky-Golay	98.25	2.30	0.00	1.75	100.00
1st derivative Smoothing Savitzky-Golay	88.60	11.49	11.11	11.40	88.89
1st derivative Smoothing Norris	89.47	9.19	14.81	12.28	85.19
2nd derivative Smoothing Savitzky-Golay	84.21	16.09	14.81	15.79	85.19
2nd derivative Smoothing Norris	90.35	10.34	7.41	9.65	92.59

The NIR spectrum is represented by the number of waves × absorbance, which is the result of the amount of energy absorbed by characteristic bonds OH, CH, CN, and CO. Spectral studies are facilitated in the presence of a reference library (peaks, a digital signature of compounds). Otherwise, bands are interpreted by comparing between standard spectra (Schwanninger; Rodrigues; Fackler, 2011; Ricci et al., 2015).

A group of 27 species was randomly selected (TQ analyst) to compose the validation set. All samples used in this set were classified correctly, confirming a 100% II (Table 5).

The system that used DM provided two results; the first (predictive Class 1) indicated the distance of the sample to the center of the nearest group, and predictive Class 2 indicated the closeness of the sample to the center of the next group. The CT group had a central distance in the range of 0.522–0.791, WT of 1.135–1.720, and HT 2.070–2.310. The species *E. schomburgkii* and *M. guianensis* that constituted the calibration set were referred to in the laboratory as CT; however, in predictive Class 1, they were classified as WT. However, these samples were classified correctly when comparing the distances of the next group.

An analysis of the results presented in Table 5 and Figure 4 shows a strong tendency of *E. Schomburgkii* and *M. guianensis* belonging to the CT group. According to Morozova, Elizarova, and Pleteneva (2013), shorter distances indicate spectral similarity with a class, whereas higher numbers indicate spectral dissimilarity. NIR spectroscopy is based on the measurement of wavelength × absorbance, which is a variable closely associated with chemical concentration (Budinova; Dominak; Strother, 2008; Prades et al., 2012). DA results together with DM presented in this study demonstrated that they are satisfactory resources to justify the classifications of types of tannins in the Amazonian woods.

In Figure 5, the spectra of standard wood-condensed tannins (CT), standard wood-without tannins (WT), and *E. Schomburgkii* and *M. guianensis* woods that were erroneously classified in the DA (non-derived spectrum) are shown. This figure provides a better distinction of the patterns and can explain the tendency to group the samples in a different class from the reference. Has the analysis carried out in the laboratory (wet) been altered and/or registered incorrectly? Could the samples be considered an outlier, and should they be excluded and/or continued in the modeling? Did the spectra show noises that could compromise the quality of discrimination?

The most plausible justification we found is that the possible grouping of species in the class WT is related to the physical, chemical, and anatomical properties of these species. For example, low concentration of extractives, basic density, very porous wood, and being white woods, i.e., having low concentration and/or absence of tannins in the woody fabric (Santiago et al., 2018). Such justification, at first, partially explains the ability of the multivariate model to discriminate wood with and without tannins.

Bands with greater heights are influenced by the concentration of chemical compounds of the investigated matrix. In Figure 5, the species *E. Schomburgkii* (green line) is closer to the standard spectrum CT (red line). This pattern is the spectral means of samples rich in polyphenols. Similarly, the spectrum of *M. guianensis* (purple lines) overlaps the WT pattern (blue line) in almost all NIR extensions. In general, the wood classified without tannins is whitish, indicating the absence of tannins or a very low concentration of tannins. Therefore, the bands used in the NIR distinction are influenced by sample concentrations.

Table 5: Results of the classification of wood species according to the tannin class (validation set).

Species	Reference class	Predictive class 1	Predictive class 2
Albizia inundata; A. subdimidiata; Andira parviflora; Campsiandra chigo-montero; Cedrela odorata; Clathrotropis nitida; Crudia amazonica; Cynometra spruceana; Dipteryx odorata; Heterostemon mimosoides; Macrolobium angustifolium; Manilkara amazonica; Mora paraensis; Ocotea cymbarum; Platonia insignis; Protium puncticulatum; Sacoglottis guianensis; Stryphnodendron guianense; Swartzia laevicarpa; Tachigali paniculata; Vatairea guianensis	Condensed tannins	Condensed tannins	Without tannins
Manilkara bidentata	Condensed tannins	Condensed tannins	Hydrolyzable tannins
Eschweilera odora	Hydrolyzable tannins	Hydrolyzable tannins	Condensed tannins
Aniba duckei; Martiodendron elatum; Qualea paraensis; Simarouba amara	Without tannins	Without tannins	Condensed tannins
Enterolobium schomburgkii; Maquira guianensis**	Condensed tannins	Without tannins	Condensed tannins

*Species used in calibration.

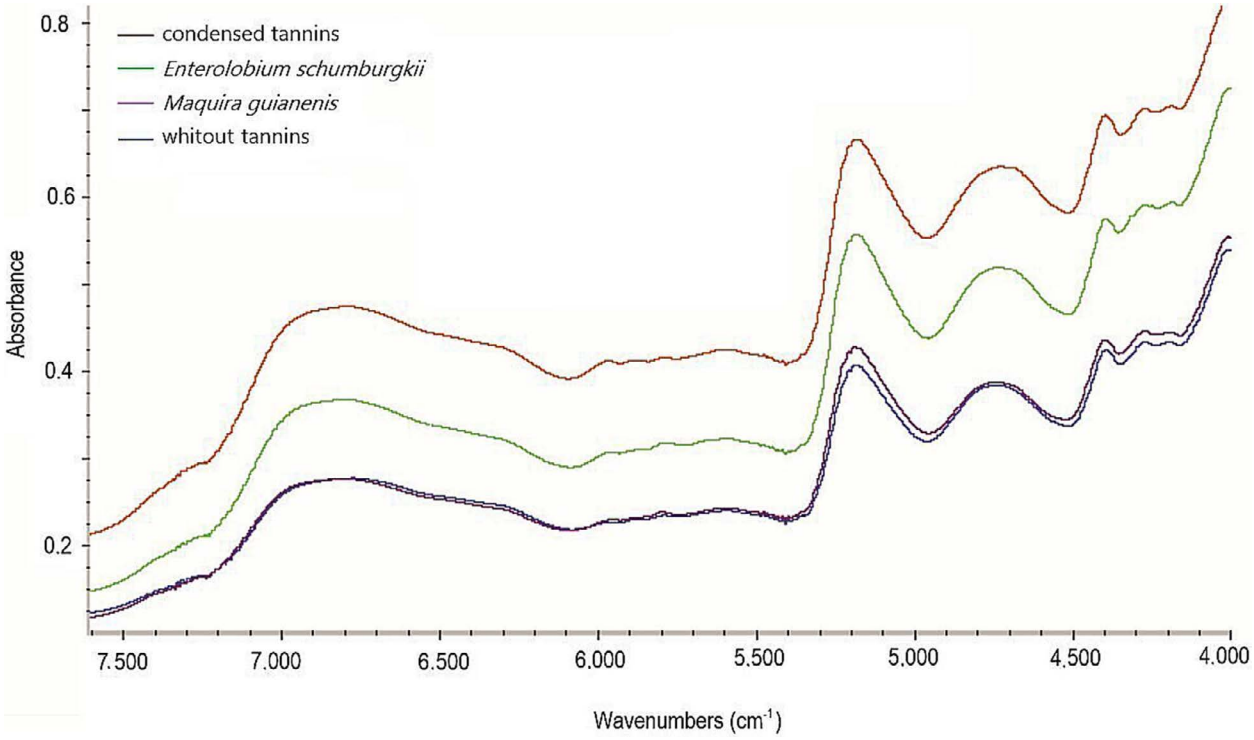

Figure 5: Profile of the NIR spectra obtained from standard wood-condensed tannin (red line), and standard wood-without tannin (blue line), and the *Enterolobium schomburgkii* species (green line) and *Maquira guianensis* (purple line), in the range of 7,500-4,000 cm^{-1}.

Cluster Analysis of tannin classes

The purpose of this analysis was to group spectral patterns from similar populations. The elements of the same group were homogeneous with each other, considering the variables (characteristics) measured in them. The inter-point distances between all samples contained in the dataset and their characteristics are represented by a two-dimensional graph called a dendrogram. Dendrogram helps to visualize clusters and similarities between samples and/or variables (Härdle; Simar, 2007; Sádecká; Tóthová; Májek, 2009).

The studied woods were grouped into 27 families, out of which 15 had individuals with tannin (CT and HT), 12 families were without tannin (WT), and 5 families had individuals with CT, and WT. In addition, DA, as a pattern recognition tool, can provide visible cluster trends, as shown in Figure 6.

The similarity between samples was the clustering parameter. Initially, the following four groups were formed with very similar characteristics (similarity ~1): WT (Araliaceae, Euphobiaceae, Hyperiaceae, Humiriaceae, Icacinaceae, Malphigiaceae, Miristicaceae, Proteaceae, Rutaceae, Simaroubaceae, Solanaceae, and Vochysiaseae), CT (Apocinaceae, Burseraceae, Clusiaceae, Combretaceae, Fabaceae, Goupiaceae, Meliaceae and Olacaceae), WT/CT (Bignoniaceae, Caryocaceae, Lauraceae, Malvaceae, and Moraceae) the fourth group was HT (Lecythidaceae and Sapotaceae) with a similarity of 0.88.

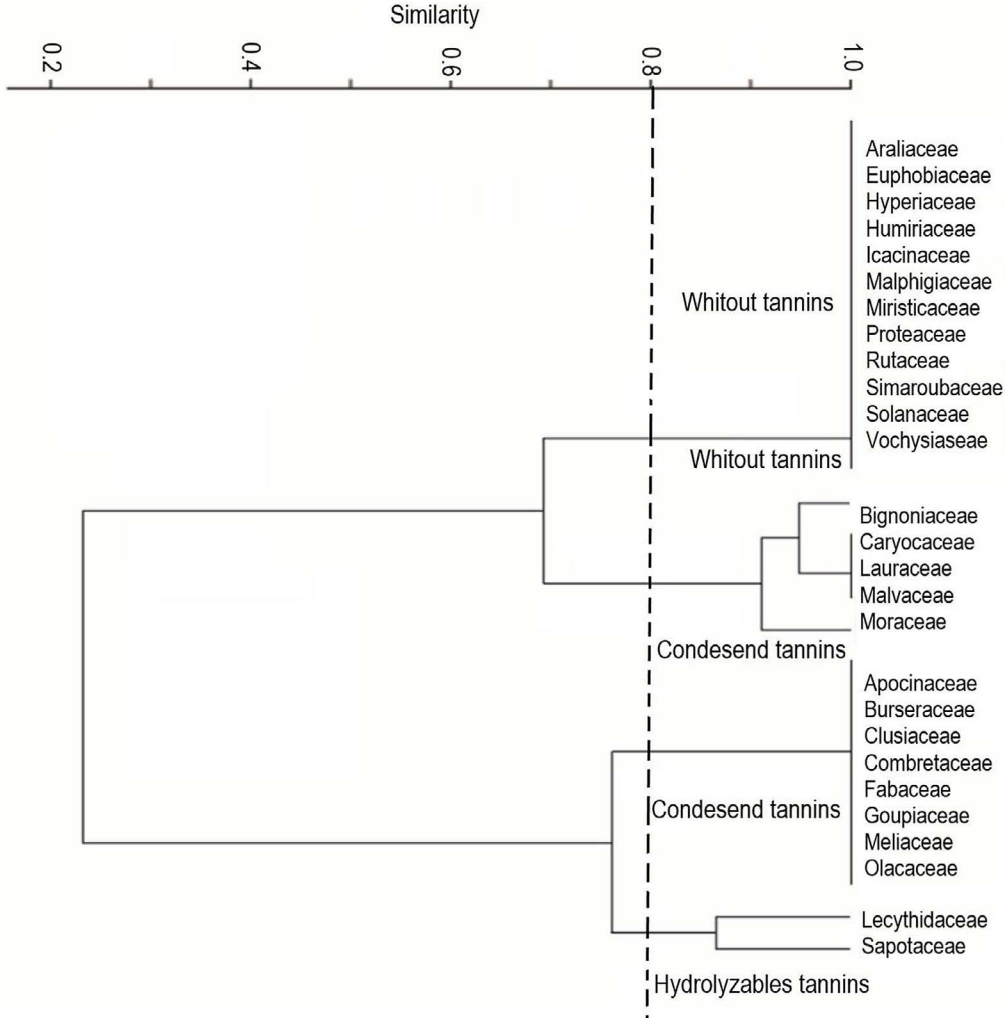

Figure 6: Dendrogram of the reflectance spectra obtained by the cluster analysis using Euclidean distance, in relation to families.

The groupings can be explained by the chemical and physical properties of each sample (Sandak; Sandak; Negri, 2011). In the case under study, different types of tannins can explain the approximation, whereas their absence indicates another group, according to the contributions of the first three PCA (98.84%) to the total variations in the data.

Souza et al. (2017) confirmed the efficiency of PCA and cluster analysis in explaining the separation of three plant species related to the number of phenolic compounds and flavonoids in the region 8,663–3,757 cm^{-1}. Grasel and Ferrão (2016) reported the formation of six groups associated with condensed/hydrolyzable tannins × plant origin. In studies of chemosystems, the presence of tannins can explain botanical groupings associated with environmental and evolutionary factors.

CONCLUSIONS

A combination of NIR spectroscopy with multivariate analysis revealed satisfactory results with a probability of 90% in distinguishing various types of tannins. The developed model can be used as a tool in ecology, forestry, and chemotaxonomy, with a focus on phenolic compounds such as tannins. The proposed methodology has advantages over the reference methods, such as the lower need for sample preparation, shorter analysis time, no use of reagents, and, consequently, no generation of waste**.**

AUTHOR CONTRIBUTION

Conceptual Idea: Nascimento, C.S.; Araújo, R.D.; Methodology design: Silva, C.E.; Data collection: Nascimento, C.S.; Araújo, R.D.; Data analysis and interpretation: Menezes, V.S.; Nascimento, C.C. and Writing and editing: Nascimento, C.S.; Nascimento, C.C.; Santos, J.

ACKNOWLEDGMENTS

This study was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Finance Code 001/Ph.D. Candidate CFT/INPA to 88882.4444 19/2019-01, Fundação de Amparo a Pesquisa do Estado do Amazonas – FAPEAM/Ph.D Candidate CFT/INPA 000284/2019. Thanks to MCTI/CNPq/FAPEAM Project "INCT Madeiras da Amazônia" for financial support, Geislayne Mendonça Silva for edit pictures and Ana Julia O. Godoy and Mateus N. da Silva reviewing the English text.

REFERENCES

AMERICAN SOCIETY FOR TESTING AND MATERIALS - ASTM. Annual book of ASTM standards. West Coshohocken, Pennsylvania, United States: ASTM, 2021. 850p.

BARBOSA, A. P. et al. Leguminosas florestais da Amazônia Central. Fitos, 1:47-57, 2006.

BRAGA, J. W. B. et al. The use of near infrared spectroscopy to identify solid wood specimens of *Swietenia macrophylla*. IAWA Journal, 32:285-296, 2011.

BUDINOVA, G.; DOMINAK, I.; STROTHER, T. FT-NIR Analysis of Czech Republic beer: A qualitative and quantitative approach. Application Thermo-Scientific, 5172:1-4. 2008.

CARDOSO, D. et al. Amazon plant diversity revealed by a taxonomically verified species list. Proceedings do the National Academy of Science, 114(40):10695-10700, 2017.

COQ, S. et al. Interspecific variation in leaf litter tannins drives decomposition in a tropical rainforest of French Guiana. Ecology, 91:2080-2091, 2010.

DOMINY, N. J.; LUCAS, P. W.; WRIGHT, S. J. Mechanics and chemistry of rain forest leaves. Journal of Experimental Botany, 54:2007-2014, 2003.

DURGANTE, F. M. et al. Species spectral signature: Discriminating closely related plant species in the Amazon with NIRS Leaf. Forest Ecology and Management, 291:240-248, 2013.

FALCÃO, L.; ARAUJO, M. E. M. Tannins characterization in new and historic vegetable tanned leather fibers by spot tests. Journal of Culture Heritage, 12:149-156, 2011.

FERNANDES, C. et al. Physical, chemical, and mechanical properties of *Pinus sylvestris* wood at five sites in Portugal. IForest, 10:669-679, 2017.

FREITAS, J. A.; VASCONCELLOS, F. J. Identificação de madeiras comerciais da Amazônia. Manaus, Amazonas, Brazil: INPA, 2019. 85p.

GRASEL, F. S.; FERRÃO, M. F. A rapid and non-invasive method for the classification of natural tannin extracts by NIRS and PLS-DA. Analytical Methods, 8:644-649, 2016.

HAMMER, O.; HARPER, D. A. T.; RYAN, P. D. A. PAST: Paleontological Static's software package for education and data analysis. Palaeontologia Electronica, 4:1-9, 2001.

HÄRDLE, W.; SIMAR, L. Applied multivariate statistical analysis. Berlin, Germany, Springer, 2007. 570p.

LUNA, A. S. et al. Rapid characterization of transgenic and non-transgenic soybean oils by chemometric. Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy, 100:115-119, 2013.

MATOS, F. J. A. Introdução a fitoquímica. Fortaleza, Ceará, Brazil: UFC, 2009. 125p.

MENEZES, C. M. et al. Direct determination of tannins in *Acacia mearnsii* bark using near-infrared spectroscopy. Analytical Methods, 6:8299-8305, 2014.

MONTEIRO, J. M. et al. Taninos: uma abordagem da química à ecologia. Química Nova, 28:892-896, 2005.

MOROZOVA, M.; ELIZAROVA, T.; PLETENEVA, T. Discriminant analysis and Mahalanobis distance in the assessment of drug's batch-to-batch dispersion and quality threshold establishment. European Scientific Journal, 9:8-25, 2013.

NASCIMENTO, C. S. et al. Characterization of technological properties of matá-matá wood (*Eschweilera coriacea* [DC.] S.A. Mori, *E. odora* Poepp. [Miers] and *E. truncata* A.C. Sm.) by Near Infrared Spectroscopy. iForest, 14:400-407, 2021.

NASCIMENTO, C. S.; VAREJÃO, M. J. C.; VIANEZ, B. F. Espectroscopia FT-NIR na predição de extrativos e polifenóis totais em cascas de espécies florestais da Amazônia. In: VIANEZ B. F. et al. (eds). Potencial tecnológico de madeiras florestais da Amazônia Central. Manaus, AM, Brazil: INPA, p.213-224, 2012.

OLIVEIRA, L. S. et al. Natural resistance of woods to *Phanerochaete chrysosporium* degradation. International Biodeterioration & Biodegradation, 64(8):711-715, 2010.

ONO, K.; HIRAIDE, M.; AMARI, M. Determination of lignin, holocellulose, and organic solvent extractives in fresh leaf, litter fall, and organic material on forest floor using NIR spectroscopy. Journal of Forestry Research, 8:191-198, 2003.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives: A review. Analytica Chimica Acta, 1026:8-36, 2018.

PRADES, C. et al. Discriminant analysis of geographical origin of cork planks and stoppers by NIR spectroscopy. Journal of Wood Chemistry and Technology, 32:66-85, 2012.

RICCI, A. et al. ATR-MID spectroscopy and chemometrics for the identification and classification of tannins. Applied Spectroscopy, 69:1243-1250, 2015.

SÁDECKÁ, J.; TÓTHOVÁ, J.; MÁJEK, P. Classification of brandies and wine distillates using front face fluorescence spectroscopy. Food Chemistry, 117:491-498, 2009.

SANDAK, A.; SANDAK, J.; NEGRI, M. Relationship between near-infrared spectra and the geographical provenance of timber. Wood Science and Technology, 5:35-48, 2011.

SANTIAGO, S. B. et al. Colagem de madeira de eucalipto com adesivos naturais. Revista Matéria, 23:1-12, 2018.

SCHWANNINGER, M.; RODRIGUES, J. C.; FACKLER, K. A review of band assignments in near infrared spectra of wood and wood components. Journal of Near Infrared Spectroscopy, 19:287-308, 2011.

SILVA, A. R. et al. Assessment of total phenols and extractives of mahogany wood by near infrared spectroscopy (NIRS). Holzforschung, 67:1-8, 2013.

SIMÕES, C. M. O. et al. Farmacognosia: Do produtos natural ao medicamento. Porto Alegre, RS, Brazil: Artmed, 2017. 580p.

SOUZA, M. et al. Predição dos teores de compostos fenólicos e flavonóides na parte aérea das espécies *Secalecereale* L., *Avena strigosa* L. e *Raphanus sativus* L. por meio de espectroscopia NIR. Química Nova, 40:1074-1081, 2017.

TAIZ, L. et al. Pant physiology and development. 6th edition. Sunderland, Massachusetts, U.S.A: Sinauer Associates, 2017. 888p.

TEYE, E. et al. Rapid differentiation of *Ghana cocoa* beans by FT-NIR spectroscopy coupled with multivariate. Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy, 114:183-189, 2013.

TIGABU, M. et al. Visible+Near Infrared Spectroscopy as taxonomic tool for identifying birch species. Silva Fennica, 52:1-13, 2018.

TSUCHIKAWA, S.; SCHWANNINGER, M. A review of recent near-infrared research for wood and paper (Part 2). Journal of Applied Spectroscopy Reviews, 48:560-587, 2013.

VAREJÃO, M. J. C. et al. Leguminosas florestais da Amazônia central. II. Prospecção das classes químicas nas cascas de espécies arbóreas. In: VIANEZ, B. F. et al. (eds.). Potencial tecnológico de madeiras florestais da Amazônia Central. Manaus, AM, Brasil: INPA, p.57-65, 2012.

VAREJÃO, M. J. C. et al. Madeiras amazônicas e os efeitos nocivos ao homem. Amazônia: Ciência & Desenvolvimento, 5(9):173-186, 2009.