Microchemical Journal 225 (2026) 117971

Contents lists available at ScienceDirect

Microchemical Journal

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

A non-destructive approach using spectral fingerprinting and chemometrics in the authentication and quality prediction of fallen tree wood

Cristiano S. do Nascimento [a,*], Jelmir C. de Andrade [b], Claudia Eugenio da Silva [a], Vinicius Kartnaller [b], Moacir Alberto A. Campos [a], Roberto D. de Araújo [a], Adrya da S. Figueiredo [a], José Carlos R. Soares [c], Marcia C. Ramos de Oliveira [d], Niro Higuchi [a]

[a] National Institute for Research in the Amazonia, Manaus, Amazonas 69067-375, Brazil
[b] Federal University of Rio de Janeiro (UFRJ), Institute of Chemistry (IQ), Cidade Universitária, Rio de Janeiro, RJ 21941-909, Brazil
[c] Embrapa Western Amazon, Manaus, Amazonas 69010-970, Brazil
[d] Program in Technological Education – PPGET, Federal Institute of Education, Science and Technology of Amazonas, 69020-120, Manaus, Amazonas, Brazil

ARTICLE INFO

Keywords:
Nondestructive tool
Partial least squares regression, Amazonian woods
Technological properties
Forest management, wood authentication

ABSTRACT

The Amazon rainforest harbors a stock of naturally fallen trees whose rational use is limited mainly by the difficulty of accurately identifying their taxa and, consequently, assigning technological value to this resource. This study proposes a rapid, non-destructive strategy that authenticates the origin of these woods and predicts their quality by integrating Fourier Transform Near-Infrared spectroscopy with chemometrics. Exploratory analysis applied to spectra from ten Central Amazon taxa revealed clear spectral patterns for some species, especially in bands at 5284–5426 and 7123–7359 cm^{-1}, associated with C—O and C—H vibrations of components such as lignin and hemicellulose. Using these fingerprints, a one-class classification approach implemented in MATLAB authenticated the woods with precision of 99.1–100%. Once identity was secured, the NIR spectra were used in PLS regression models developed in TQ Analyst software to predict properties chemical, physical and mechanical. Although the classification and regression steps were performed in different software environments due to the specific requirements of each algorithm, they were applied sequentially to the same spectral dataset, forming an integrated workflow. The models showed strong predictive ability, with coefficients of determination for prediction generally above 0.88, low prediction errors, and ratio of performance to deviation values mostly above 3, indicating suitability for screening and quantitative estimation. However, due to severe logistical constraints of sampling naturally fallen trees in the Amazon, these models were built with a limited test set and rely solely on internal validation; thus, these high-performance metrics should be interpreted cautiously as an exploratory proof-of-concept. While this study successfully demonstrates the feasibility of the method, rigorous external validation with larger, independent datasets is required before full implementation.

1. Introduction

The Amazon rainforest, a global epicenter of biodiversity, hosts a vast array of tree species of significant economic and ecological value. Natural disturbances, such as storms and windthrows, result in substantial volumes of fallen trees, presenting a largely untapped resource for sustainable value chains [1,2]. However, leveraging this resource depends on overcoming two primary challenges: the accurate authentication of the wood and the efficient assessment of its inherent quality for industrial applications [3,4]. Wood identification, based on anatomical features, is relatively efficient for well-preserved samples. However, it loses accuracy when the surface is degraded, charred, or colonized by fungi, and is often unfeasible for logs that have been fallen for extended periods [5,6].

Chemical methods (such as chromatography or marker extraction) and genetic approaches (DNA barcoding) can also be used, but they are more expensive, require sample preparation, and, in the case of DNA, are affected by material degradation over time. This set of constraints makes classical identification slow, destructive, and poorly suited for large-scale screening of fallen timber in the forest, creating an opportunity for non-destructive spectroscopic techniques coupled with chemometrics [7–9].

* Corresponding author.
E-mail addresses: s-nascimento@hotmail.com, cristiano.nascimento@inpa.gov.br (C.S. Nascimento).

https://doi.org/10.1016/j.microc.2026.117971
Received 13 February 2026; Received in revised form 26 March 2026; Accepted 5 April 2026
Available online 8 April 2026

In addition to classical approaches, several advanced non-destructive technologies have been explored for wood identification and quality assessment. Imaging-based systems such as X-ray computed tomography enable visualization of internal structure and defects without sectioning [10,11], with high-resolution micro-CT supporting taxonomic identification through diagnostic features like vessels and rays [12]. In industrial settings, CT integrated with artificial intelligence is increasingly used in automated log-scanning for defect detection and quality classification. Computer vision approaches based on deep learning have also been developed to classify wood species using macroscopic anatomical patterns from digital images [13]. Additionally, spectroscopic techniques such as ATR-FTIR and Raman spectroscopy provide detailed molecular fingerprints of lignocellulosic materials, showing promise for species discrimination [14,15]. However, many of these technologies require expensive instrumentation, controlled conditions, or complex data acquisition, limiting their applicability in remote regions like the Amazon. In contrast, near-infrared (NIR) spectroscopy offers rapid acquisition, minimal sample preparation, and portable instrumentation, enabling in situ screening of wood materials.

Wood quality in fallen trees exhibits marked variability due to prolonged exposure to humidity and microbial degradation, complicating both species authentication and quality prediction [16,17]. In this context, NIR spectroscopy stands out as a fast, non-destructive tool capable of capturing spectral fingerprints rich in information on O—H, C—H, and C—O functional groups from lignocellulosic matrices [18–20]. While most studies have focused on predicting density, moisture, and chemical properties from commercial plantations or processed samples [21–23], the present study advances the field by applying this technology to a neglected and highly complex resource: naturally fallen trees in the Amazon rainforest. Unlike previous approaches, our workflow integrates botanical authentication with simultaneous prediction of physical, mechanical, and energy-related properties directly on solid wood and under suboptimal conditions [24,25].

Therefore, the objective of this study was to investigate the feasibility of a non-destructive methodology based on FT-NIR spectroscopy and chemometrics to (i) authenticate fallen wood from the Central Amazon at the species/genus level using Data-Driven Soft Independent Modeling of Class Analogy (DD-SIMCA), and (ii) predict the physical and chemical parameters of the wood using PLSR. This study represents a significant expansion and advancement over our previous works: here, for the first time, we integrate authentication and quality prediction into a unified workflow, substantially broaden the taxonomic scope to five families and ten species, and introduce DD-SIMCA as a statistically rigorous one-class classification approach for timber authentication. This integrated strategy provides decision support for the allocation, reuse, and valuation of this abundant, yet often neglected, forest resource within sustainable management programs.

2. Materials and methods

2.1. Study area and sampling of the fallen tree woods

The research was conducted with samples collected in mature terra-firme forests located in the upper Solimões River region, in Amazonas, Brazil (3° 11′ 8.9” S, 67° 8′59.4” W; 3° 27′ 0.9” S, 67° 33′15” W) (Fig. 1). This area is characterized by a humid tropical climate, with an average annual rainfall of approximately 2600 mm, an average annual temperature of 29 ± 2 °C, and an average altitude of 65 m above sea level. The topography is predominantly semi-flat, while the soils are mostly yellow latosols, recognized for their clayey texture, low permeability, and low fertility. The predominant vegetation is classified as dense alluvial rainforest, typical of the varied ecosystems of the Amazon basin [26].

In the field, each fallen tree was treated as an individual sample. Detailed measurements were recorded, including total height and

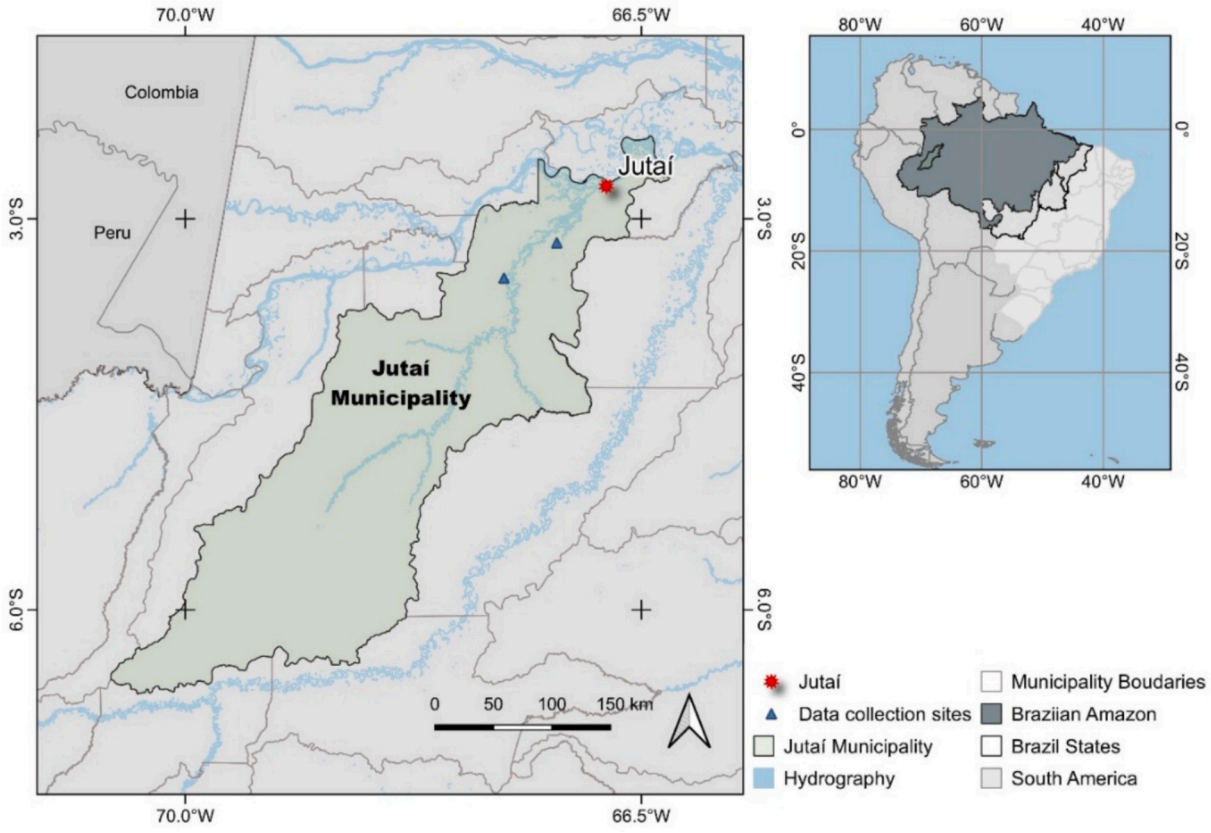

Fig. 1. Geographic location of the collection area for naturally fallen trees.

diameter at breast height (DBH). The phytosanitary condition of the trunks was assessed by visual inspection, which revealed no signs of degradation, such as hollows, rot, or infestations by xylophagous organisms. Wood samples were transported to the Wood Laboratory of the National Institute for Amazonian Research (INPA) in Manaus, Brazil, for storage and subsequent analysis. Wedges extracted from disks collected at DBH height were sent to the Wood Anatomy and Identification Laboratory at INPA, where taxonomic confirmation was performed by specialist J.A. Freitas using the comparison technique, and anatomical features were described.

For this study, the ten most frequently occurring species in the collection area were selected, with three individuals per species (Table 1). The selection of three individuals per species follows established recommendations in wood anatomy and technology, which consider this number sufficient to capture the essential variability of qualitative and quantitative features for taxonomic and technological characterization [5,27,28]. This approach is also consistent with previous chemometric studies employing NIR spectroscopy for wood differentiation, in which three to five individuals per taxon have been successfully used to build robust classification models [6,29]. Sampling was constrained by the natural occurrence and abundance of fallen trees in the terra-firme forest; therefore, only species with at least three fallen individuals were included to ensure a balanced experimental design.

From each individual, four samples measuring 20 × 20 × 30 mm were prepared, resulting in a total of 120 samples for analysis.

2.2. NIR instrumentation and spectral acquisition

The woods were acclimatized (relative humidity < 60% at 20 °C), and spectra were obtained from the surface of the radial plane of the samples in an FT-NIR spectrometer (near-infrared with Fourier transform), Antaris II Thermo Scientific equipped with a halogen source, diffuse reflectance accessory, paired indium-gallium-arsenic high sensitivity, stability detector, and Michelson interferometer. The system uses the RESULT™ Integration software that operates in the range of 10,000–4000 cm^{-1} with a spectral resolution of 8 cm^{-1} and 16 scans, where up to 750 data points were obtained in this spectral range. Three readings were taken on the selected face of each sample, resulting in a final $n = 360$.

2.3. Chemometric framework

2.3.1. Exploratory and classification Chemometric modeling

The chemometric analysis was conducted in two sequential stages, each implemented in different software environments optimized for the respective algorithms. First, exploratory analysis and one-class classification (DD-SIMCA) were performed using MATLAB 2023b® (The MathWorks Inc., Natick, USA) with the GAMMA-GUI toolbox [30], which provides a user-friendly interface for multivariate analysis. Second, quantitative prediction models were developed using TQ Analyst™ software, which offers specialized tools for PLS regression. Although the preprocessing strategies were optimized independently for each software environment, as required by the specific algorithms and their respective capabilities, the same spectral dataset was used throughout. This sequential workflow ensures that the classification and regression steps are complementary, even though implemented in different platforms.

2.3.2. Spectral preprocessing

To avoid model overfitting and artificially inflated performance due to shared spectral structure, the three spectra acquired per tree were averaged to obtain a single mean spectrum per individual prior to chemometric analysis. Therefore, the allocation of samples into calibration and validation sets was performed strictly at the individual tree level, ensuring complete data independence during the validation step. To minimize interference and improve the performance of the classification

Table 1
Species selected for the study.

Botanical family	Natural fallen trees	Typical characteristics of wood	Tangential cut view of the wood
Fabaceae	*Hymenolobium excelsum* Ducke	High-density wood, with reddish-brown heartwood on a yellow-brown background, with light streaks. The sapwood has a yellow-gray hue. The grain is interlocked; the texture is coarse and has no distinctive smell or flavor.	
	Inga alba (Sw.) Willd.	Medium-density wood, with color ranging from light brown to reddish, with straight or irregular (interlocking) grain and generally coarse texture. It is odorless and tasteless.	
	Tachigali paniculata Aubl.	Moderately heavy wood, soft to the cut, light brown to brown heartwood; fine texture; wavy grain; accentuated shine; smooth to the touch; imperceptible odor and taste	
Lauraceae	*Sextonia rubra* (Mez) van der Werff	Moderately heavy wood, with heartwood presenting tones ranging from pinkish-brown to reddish-brown. The wood has straight grain, texture ranging from medium to coarse, with no distinctive figure. When green, it emits a slight unpleasant odor, which disappears after drying.	
	Mezilaurus Itauba (Meisn.) Taub. ex Mez	Heavy wood, with heartwood ranging from olive yellow to brownish yellow, which acquires a light beige tone after drying, with an average thickness of 3.5 cm. It has a straight grain, medium texture, and a discreetly marked figure.	
Lecythidaceae	*Eschweilera odora* (Poepp. ex O. Berg) Miers	High-density wood, with heartwood ranging from light-brown to dark brown, often with stripes, while the sapwood has a yellowish tone. It has a straight grain, medium texture, and an unpleasant odor when fresh, which disappears after the drying process. The flavor is neutral and indistinct.	
	Couratari stellata A.C.Sm.	Wood of moderate weight, with heartwood of dark brown coloration and yellowish sapwood. It has a straight grain,	

(continued on next page)

Table 1 (*continued*)

Botanical family	Natural fallen trees	Typical characteristics of wood	Tangential cut view of the wood
Moraceae	*Maquira coriacea* (H.Karst.) C.C. Berg	medium texture, subtle aroma, and a slight bitter taste. Wood of medium density, with indistinct heartwood and sapwood. The color varies from yellowish white to light yellowish brown. The growth rings are not very distinct. The grain is straight to cross-grained, reversed. It has a medium texture, no visible pattern, and a weak shine.	
	Brosimum rubescens Taub.	Wood classified as heavy, with heartwood of a reddish hue with light yellow nuances and cream sapwood, which can reach between 15 and 20 cm. It has an interlocking grain, medium texture, an attractive decorative pattern, and is odorless and tasteless.	
Myristicaceae	*Virola surinamensis* (Rol. exRottb.) Warb.	Low-density wood, heartwood is distinguished from sapwood by a cream color, lighter sapwood, fine to medium texture and regular grain.	

models and correct any interference that could not be experimentally controlled, various transformations were applied to the spectral data, including the first derivative with Savitzky-Golay (S-G) smoothing, the second derivative with S-G smoothing, multiplicative scatter correction (MSC), standard normal variate (SNV), and the combination of first and second derivatives using S-G + MSC and SNV. Three pre-processing methods were used: mean centering, autoscaling, and Pareto scaling, resulting in 93 possible combinations (31 transformations × 3 pre-processings) [20].

2.3.3. Principal component analysis (PCA)

PCA was used to investigate the natural behavior of the wood sample data, with a particular focus on evaluating the grouping tendencies of the samples according to species and botanical families. This multivariate technique reduces the dimensionality of the spectral dataset by generating linear combinations of the original variables (wavenumbers) that retain the maximum variance in the multivariate space. The resulting variables, known as principal components (PCs), are orthogonal to one another and represent successive directions of greatest variability in the data structure. In the context of wood analysis, the components are obtained through Singular Value Decomposition (SVD) of the spectral data matrix (X), producing the score matrix (T), loading matrix (L), and residual matrix (E) for a defined number of A components [31]. This approach enables the identification of natural clustering patterns among the wood samples and allows for a detailed interpretation of how specific spectral variables contribute to the separation between species and families.

2.3.4. Data-driven–soft independent modeling of class analogy (DD-SIMCA)

DD-SIMCA is an adaptation of SIMCA [32], proposed by Zontov et al. [33], which uses PCA decomposition to process a training matrix consisting of pre-processed average spectra of wood samples. After obtaining the score matrix (T) and loading matrix (L) from the PCA decomposition, two distances are calculated for each training sample: the score distance (SD) and the orthogonal distance (OD). These distances represent population statistics, with critical values defined for each. When a new sample is considered, it is projected onto the model, and its SD and OD values are compared with the known critical levels to determine its membership to the target class [34,35]. The SD, also known as leverage, is the squared Mahalanobis distance from the model center to the sample, calculated in the score subspace. The OD, in turn, measures the residual variance by summing the squared residuals for the PCs used in the PCA model. The total distance for a sample is then calculated as a weighted sum of SD and OD.

The acceptance threshold for classifying a sample into the target class is based on type I error (α). The acceptance region of the model is determined by calculating the critical distance using a chi-squared distribution with the appropriate degrees of freedom (DoF), which are estimated based on the training dataset [35,36]. The accuracy of the model is quantified through the DoF and its relationship to the chi-squared distribution. The acceptance area is visualized in an acceptance plot, where the distance of each sample is compared with the critical threshold. Samples with distances below this threshold are considered part of the target class, while those above are classified as outliers. For a more detailed mathematical derivation, it is recommended to refer to the articles by Pomerantsev and Rodionova [36] and Pomerantsev [37], which cover the complete mathematical formulations and methods for calculating the necessary parameters.

To ensure the robustness of the results and the reliability of the DD-SIMCA models, leave-one-out cross-validation (LOOCV) was employed. This method allows the performance of the model to be evaluated without relying on any specific sample by using each sample individually for validation while the remaining samples are used for training. During this process, several performance parameters were assessed, including cross-validation sensitivity and the Figures of Merit (FoMs) obtained during the prediction stage. These indicators were monitored throughout all LOOCV steps, providing a comprehensive and detailed analysis of the models' discriminant ability. To ensure that the model performance was not inflated by overfitting resulting from preprocessing optimization, the DD-SIMCA validation followed rigorous control criteria. The number of Principal Components (PCs) was optimized by evaluating the behavior of the error metrics as a function of the PCs, rather than using arbitrary limits of explained variance. In addition, diagnostic tools such as the extreme plot were employed to confirm that the model respected the expected statistical uncertainty envelope, ensuring the exclusive modeling of the intrinsic chemical variance of the evaluated classes.

2.3.5. Quantitative prediction models for wood properties

A separate PLS model was constructed for each property using the TQ Analyst™ software. The number of latent variables (LVs) for each PLS model was automatically selected by the TQ Analyst™ software using its internal cross-validation algorithm. The dataset was partitioned into calibration (2/3 of the samples) and external validation (1/3 of the samples) subsets using the Kennard-Stone algorithm; however, to ensure independence between the sets and avoid any bias arising from potential similarity among samples from the same tree, sample selection was performed at the tree level, meaning all four specimens originating from the same tree were allocated entirely to the same subset. This strategy ensures that there is no overlap of genetic or micro-environmental factors between the calibration and validation groups, thereby conferring greater robustness and realism to the predictive performance estimates. Prior to modeling, all spectral data were mean-centered. The models

were evaluated using raw data as well as data pre-processed with first- and second-order derivatives, applied both with and without Norris or Savitzky-Golay smoothing filters (see Table 4). During calibration, outlier detection was performed by analyzing residuals and leverage; samples with a standardized residual exceeding ±2.5 were removed, which occurred only for the ash model.

To obtain the reference values (Y-variables) for the calibration of the PLS model, the following standard procedures were employed, divided into two parts:

I. Chemical properties

Determination of total extractives content - ASTM D1105, D1107 [38]: Two successive extractions were performed with solvents of different polarities. In the first extraction (FE), 5.00 g of sawdust (60 mesh) was transferred to a cellulose cartridge in a Soxhlet extractor, using approximately 180 mL of toluene-ethanol solution (1:2) under reflux for 8 h until complete extraction. The cartridges were dried in an oven (100 °C) and stored in a desiccator. The flask containing the extract was concentrated, dried, and weighed to constant weight. The FE content was calculated as $FE\% = (W_f - W_i)/W_d \times 100$, where W_i = initial weight of the flask, W_f = final weight of the flask with extract, and W_d = weight of the dry sample. The second extraction (SE) repeated the procedure using 95% ethanol. The total extractives content (TEx) corresponds to the sum of FE and SE.

Determination of solubility in hot water - ASTM D1110 [38]: The extracted sawdust (FE and SE) was dried and extracted with boiling distilled water in a water bath for 4 h, with the water being changed every hour. The material was filtered in a Gooch crucible, dried in an oven (100 °C/24 h) and weighed until constant weight. The hot water solubility (HWS) was calculated as $HWS\% = (W_1 - W_2)/W1 \times 100$, where W_1 = initial weight and W_2 = weight after extraction.

Determination of lignin content - ASTM D1106 [38]: A 1.00 g sample of extractive-free wood (FE, SE, and HWS) was treated with 72% H_2SO_4 for 2 h in a cold bath. Then, 560 mL of water were added, and the mixture was kept in a water bath (100 °C) for 4 h. After filtration in a Gooch crucible, the material was washed with heated distilled water, dried (100 °C/12 h), and weighed to constant weight. The Klason lignin content was determined as $Lignin\% = W_2/W_1 \times 100$, where W_1 = initial dry weight and W_2 = weight of the obtained lignin.

Determination of holocellulose content [39]: A 1.00 g sample of extractive-free sawdust was refluxed with 20 mL of 3% HNO_3 (80 °C/30 min), filtered, and treated with 25 mL of 3% NaOH for further digestion. The residue was washed with ethanol, ethanol-water solution (1:1), and water until the filtrate reached neutral pH. The resulting whitish material was dried (100 °C) and weighed to constant weight. The holocellulose content was calculated as $Holocellulose\% = W_2/W_1 \times 100$, where W_1 = initial dry mass and W_2 = final dry mass.

Ash content determination - ASTM D1102 [38]: Exactly 1.00 g of sawdust (60 mesh) was placed in a porcelain crucible, dried in an oven (100 °C/1 h) and incinerated in a muffle furnace with gradual heating to 580–600 °C. The crucible was weighed until constant weight. The ash content was calculated as Ash% = Weight of ash/Weight of dry sawdust × 100.

Determination of tannins and other polyphenols [40]: A 2.00 g sample of sawdust (40 mesh) was refluxed with water (90 °C/60 min) in two successive extractions. The combined extract was diluted to 500 mL. A 100 mL aliquot was treated with 10 mL of 40% formaldehyde and 5 mL of concentrated HCl under reflux (30 min) using the Stiasny method. The precipitate was filtered, dried, and weighed. The tannin and polyphenol content (TPf) was calculated as $TPf\% = TMT/W_d \times 100$, where TMT = total mass of tannins and W_d = dry weight of the wood.

All determinations were performed in triplicate, and the mean values were used as reference data for PLS modeling.

II. Physical-mechanical properties.

Determination of moisture content (MC) - ASTM D2016 [38]: A 1.00 g sample of sawdust (60 mesh) was oven-dried (100 ± 2 °C for 4 h) and weighed until constant weight. The moisture content was calculated as $MC\% = (W_w - W_d)/Wd \times 100$, where W_w = wet mass and W_d = dry mass. Three replicates were performed, as recommended by the standard.

Determination of wood density (WD) - NBR 7190 [41]: Specimens (20 × 20 × 30 mm) were water-saturated to obtain the green volume, weighed, and oven-dried (100 ± 3 °C) until constant weight. The wood density was calculated as $WD = W_d/V_s$, where W_d = dry mass (g) and V_s = saturated volume (cm^3). Ten replicates were used for this determination, as recommended by the standard.

Determination of static bending (MOR and MOE) - BS 373 [42]: Specimens (20 × 20 × 300 mm) were conditioned (20 ± 2 °C, 65 ± 5% RH for 7 days) until the moisture content reached approximately 12%. The tests were performed on an INSTRON 1123 universal testing machine (20 t capacity, crosshead speed of 1.0 mm/min) with central loading and a distance between supports of 280 mm. The modulus of rupture (MOR) was calculated as $MOR = 3 \cdot P \cdot L/2 \cdot b \cdot h^2$, where P = maximum load (N), L = distance between supports (280 mm), b = width (mm) and h = thickness (mm). The modulus of elasticity (MOE) was calculated as $MOE = \Delta P \cdot L^3/4 \cdot \Delta y \cdot b \cdot h^3$, where ΔP = change in load within the proportional limit and Δy = corresponding change in deflection (mm). Ten replicates were performed for this determination, following the standard recommendation.

Determination of the higher heating value (HHV) - ASTM D2015 [38]: Approximately 0.80 g of sawdust (60 mesh) was placed in the combustion capsule of the bomb calorimeter. After oxygen injection, the bomb was immersed in the calorimeter vessel containing water and the ignition was initiated. The result HHV was obtained in Cal/g after approximately 10 min. Three replicates were performed, as recommended by the standard.

All determinations were made on replicates defined by each standard, and the average values were used as reference data for PLS modeling.

Additionally, porosity (Φ) [43] and Janka hardness (f_{H90}), transverse to the fibers [44] were estimated using regression equations, with wood density as the dependent variable, allowing for a broader assessment of wood properties.

2.4. Model performance

To ensure the robustness of the results and the reliability of the DD-SIMCA models, leave-one-out cross-validation (LOOCV) was employed. This method allows the performance of the model to be evaluated without relying on any specific sample by using each sample individually for validation while the remaining samples are used for training. During this process, several performance parameters were assessed, including cross-validation sensitivity and the Figures of Merit (FoMs) obtained during the prediction stage. These indicators were monitored throughout all LOOCV steps, providing a comprehensive and detailed analysis of the models' discriminant ability.

In addition, metrics such as sensitivity (SEN, specificity (SPE), accuracy (ACC), and the Matthews correlation coefficient (MCC) were calculated to strengthen the reliability of the evaluation. The inclusion of the MCC as a complementary metric is particularly relevant in this context, as the number of samples in the calibration class (modeled species) is lower than the number of samples in the validation class (other species), characterizing an imbalanced scenario. The MCC takes into account all elements of the confusion matrix (true positives, false positives, true negatives, and false negatives), providing a more balanced and reliable assessment of model performance, and avoiding the biased interpretations that can arise from traditional metrics [45]. These metrics were calculated based on the values of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), providing a comprehensive evaluation of the reliability and accuracy of the model classifications, considering different significance level values α (0.01 and 0.05) (Eqs. 1 to 4) [20,34,46].

$$SEN = \frac{TP}{TP + FN} \quad (1)$$

$$SPE = \frac{TN}{TN + FP} \quad (2)$$

$$ACC = \frac{TN + TP}{TN + TP + FN + FP} \quad (3)$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (4)$$

The DD-SIMCA models were developed to assess whether it was possible to capture the unique spectral signature of each wood species. To this end, a specific model was built for each species, considering species X as the target class, while all other species were treated as the external class. This approach simulates a real-world inspection scenario, in which the goal is to authenticate the identity of a sample from a fallen tree, confirming whether or not it belongs to the modeled species.

For the PLSR models, statistical quality was assessed based on the number of latent variables (LVs) used, the coefficient of determination (R^2), the root mean squared error (RMSE) for calibration (RMSEC) and prediction (RMSEP), the prediction bias, and the ratio between performance and deviation (RPD).

3. Results and discussion

3.1. Visual inspection of spectral information

Fig. 2A shows the average NIR absorbance spectra for the ten wood species evaluated, with $n = 12$ samples per species. The spectra reveal significant variations in both the shape and intensity of the bands, reflecting differences in the chemical composition of the woods, particularly in the content and structure of lignin and hemicellulose, which are major components of the plant cell wall [10,15]. The species *H. excelsum* stands out with higher absorbance values in key spectral regions, such as 9200–9000 cm^{-1}, 8800–8200 cm^{-1}, and 7350–7200 cm^{-1}, which are attributed to C—H and C—O deformations and stretching vibrations associated with lignin and phenolic compounds (Table 2). This elevated spectral response suggests a higher density of aromatic functional groups, such as guaiacyl and syringyl units.

H. excelsum and *E. odora* also presented pronounced peaks between 6900 and 6700 cm^{-1} and 5322–5014 cm^{-1}, regions attributed to aliphatic C—H bonds and C—O stretching of guaiacyl/syringyl rings, respectively, indicating complex lignocellulosic profiles rich in aromatic structures. These characteristics are also observed, to a lesser extent, in *S. rubra*, whose spectral intensity in similar regions suggests a comparable chemical composition.

In contrast, species such as *I. alba*, *V. surinamensis*, and *T. paniculata* presented lower spectral intensities across the entire analyzed range, which may indicate lower lignin and hemicellulose content, and potentially lower density and natural durability. On the other hand, *B. rubescens* and *T. paniculata* presented more balanced spectra, with well-defined bands in the ranges of 4944–4539 cm^{-1} and 4447–4442 cm^{-1}, associated with C—O and C—H vibrations of aromatic rings. This balanced profile may reflect a more homogeneous distribution of functional groups, which could be associated with favorable physical-mechanical properties.

Fig. 2B presents the average NIR absorbance spectra organized by botanical family, grouping the species. Similar trends to those observed in the species-level analysis can be seen, with notable variations in band intensities among the groups. The Lecythidaceae family shows the highest absorbance intensities in regions such as 7350–7200 cm^{-1} and 5322–5014 cm^{-1}, suggesting a higher concentration of aromatic lignin and phenolic compounds, a pattern consistent with the performance of *E. odora*, a representative species of this family. The Lauraceae and Moraceae families presented intermediate spectral intensities, particularly in the 7900–7700 cm^{-1} and 6900–6700 cm^{-1} ranges, associated with methylene and aliphatic groups from lignin and hemicellulose [6,47].

In contrast, the Myristicaceae family displayed the lowest absorbance values across nearly all spectral regions, especially within 9200–9000 cm^{-1}, 8800–8200 cm^{-1}, and 7350–7200 cm^{-1}, as indicated in Table 2. This suggests a lower density of structural functional groups, such as C—H and C—O bonds, consistent with the low physical-mechanical values observed for *V. surinamensis*, the representative species of the family. Table 2 provides the essential vibrational assignments for interpreting these spectral regions, linking specific bands to potential chemical compounds, and thereby enabling robust chemical inferences from the raw spectral data [18,48].

Overall, the average spectra by species and by family reveal that, despite the spectral overlap commonly observed in the NIR region, these similarities reflect chemical composition affinities among taxonomic

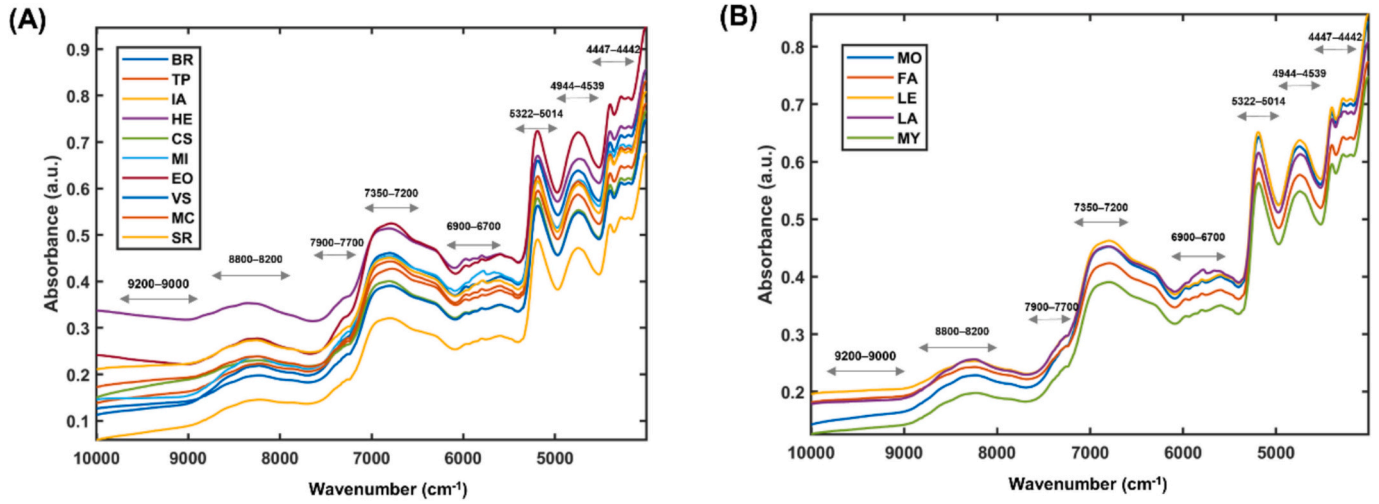

Fig. 2. Average near-infrared absorbance spectra obtained for wood samples: (A) Average spectra by species, highlighting variations in band intensities associated with differences in wood chemical composition among the species; (B) Average spectra grouped by botanical family, showing characteristic spectral patterns. *Brosimum rubescens* (BR), *Tachigali paniculata* (TP), *Inga alba* (IA), *Hymenolobium excelsum* (HE), *Couratari stellata* (CS), *Mezilaurus itauba* (MI), *Eschweilera odora* (EO), *Virola surinamensis* (VS), *Maquira coriacea* (MC), and *Sextonia rubra* (SR); Moraceae (MO), Fabaceae (FA), Lecythidaceae (LE), Lauraceae (LA), and Myristicaceae (MY).

Table 2
Wavenumber ranges (cm^{-1}) observed in the NIR spectra and their respective vibrational assignments, with possibly associated compounds.

Wavenumber Range (cm^{-1})	Associated Vibration	Possible Associated Compound	Reference
9200–9000	C–O–C + C–H deformation	Lignin, Hemicellulose	Nascimento et al. [15]
8800–8200	C–H deformation	Lignin, Hemicellulose	Nascimento et al. [15]
7900–7700	C–H stretch (methylene)	Lignin, Hemicellulose	Nascimento et al. [15]
7350–7200	C–O stretch, C—H stretch	Lignin, Phenolic Compounds	Eugenio da Silva et al. [6]
6900–6700	C–H stretch (aliphatic)	Lignin, Hemicellulose	Eugenio da Silva et al. [6]
5322–5014	C–O stretch (guaiacyl/syringyl)	Lignin (guaiacyl/syringyl)	Eugenio da Silva et al. [6]
4944–4539	C–O stretch (syringyl/guaiacyl)	Lignin (guaiacyl/syringyl)	Eugenio da Silva et al. [6]
4447–4442	C–H stretch (aromatic ring)	Lignin (aromatic)	Eugenio da Silva et al. [6]
7092–7057	O–H stretch (phenolic hydroxyl)	Lignin (phenolic)	Sandak et al. [36]
6944–6838	C–H stretch & C—H def (methyl/methylene)	Methyl/methylene groups	Sandak et al. [33]
6874–6844	O–H stretch (hydrogen-bonded)	Lignin (hydrogen bonded)	Sandak et al. [36]
6440–6008	C–H stretch (aromatic ring) & C═O stretch	Lignin (aromatic), hemicellulose	Eugenio da Silva et al. [6]
5980–5950	C–H stretch (aromatic ring)	Lignin (aromatic)	Sandak et al. [36]
5935–5848	C–H stretch (methylene group)	Lignin, Hemicellulose	Eugenio da Silva et al. [6]
5700–5583	C–H stretch	Lignin, Hemicellulose	Eugenio da Silva et al. [6]
5265–5068	C–O stretch (guaiacyl ring)	Lignin (guaiacyl)	Nascimento et al. [15]
4910–4872	C–H stretch	Lignin (guaiacyl)	Sandak et al. [33]
4686–4560	C–H stretch & C—C stretch	Lignin, Hemicellulose	Sandak et al. [33]
4452–4411	C–O stretch (guaiacyl/syringyl)	Lignin (guaiacyl/syringyl)	Kelley et al. [37]
4280–4195	C–H stretch	Lignin, Hemicellulose	Sandak et al. [33]
4105	C–O stretch (primary alcohol, guaiacyl C—H)	Lignin (guaiacyl)	Schwanninger et al. [35]
4014	C–H stretch & C—C stretch	Lignin, Hemicellulose	Sandak et al. [36]

groups. At the same time, the subtle yet systematic differences in the intensity and shape of spectral bands indicate the presence of unique spectral signatures for each botanical group. These underlying patterns, while not easily discernible by visual inspection alone, can be effectively elucidated through chemometric techniques such as Principal Component Analysis, which is discussed in the following section as a tool for revealing natural groupings and identifying discriminant variables among species and families.

3.2. Exploratory analysis using PCA

Prior to the PCA, the raw NIR spectra were preprocessed using the first derivative (Savitzky-Golay algorithm, 2nd order polynomial, 9-point window) followed by MSC. This specific preprocessing sequence was selected because it successfully minimized baseline drifts and resolved overlapping bands, providing the most effective exploratory separation behavior among the evaluated taxa.

The PCA was applied to explore natural grouping patterns among the wood samples based on the NIR spectra, assessing whether the chemical differences represented in the vibrational absorption bands are sufficiently pronounced to separate the groups in the multivariate space. The PCA was conducted at two levels: by species and by botanical family, with the results shown in the score plots of Fig. 3 **A** and **B,** respectively. The first three PCs captured 84.52% of the total variance of the spectral data, indicating that most of the chemical information was preserved in this three-dimensional projection.

The analysis of the PC1 scores for both cases reveals a clear separation between the samples grouped by species (Fig. 3A) and by botanical family (Fig. 3B). In the first case, *M. itauba* and some samples of *S. rubra* project into the negative quadrant of PC1, while the other species remain in the positive quadrant, indicating a well-defined separation. This behavior is consistent with the pattern observed in Case 2, where the samples from the Lauraceae family, to which *S. rubra* and *S. rubra* belong, also predominantly concentrate in the negative region of PC1. However, in Case 2, the separation is even more pronounced, reflecting the unique chemical composition of the Lauraceae family. The chemical explanation for this separation can be found in the PC1 loadings (Fig. 3C), which identify the spectral regions responsible for the separation at 4250–4412 cm^{-1}, 5284–5364 cm^{-1}, 7123–7397 cm^{-1}, and 5754–5847 cm^{-1}, associated with C—O and C—H vibrations of lignin (guaiacyl/syringyl units), hemicellulose, and aromatic phenols [18,47]. This indicates that the Lauraceae species investigated in this study differ from the other species by presenting lower intensity in the bands with positive loadings, which corresponds to a lower relative contribution of lignin and phenolic compounds to their overall chemical profile.

The score plots for PC2 show a well-defined separation between species and botanical families, explaining 26.01% of the total variance. In Case 1, the species *H. excelsum*, *B. rubescens*, and more strongly *E. odora* projected into the negative region of PC2, while the remaining species were predominantly grouped in the positive quadrant. In Case 2, the separation was even more refined, with the Lecythidaceae family being projected mainly in the negative region of PC2 but forming two distinct subgroups with samples of *C. stellata* and *E. odora*, whose spectral characteristics differ despite belonging to the same family. Additionally, some samples from the Fabaceae and Moraceae families were also distributed in the negative quadrant, reflecting behavior similar to what was observed in Case 1, where species such as *H. excelsum* and *B. rubescens* presented spectral characteristics shared with Lecythidaceae.

The observed separation was explained by the PC2 loadings (Fig. 3D), which highlight spectral bands such as 5284–5426 cm^{-1} and 7123–7359 cm^{-1}, associated with C—O and C—H vibrations of lignocellulosic compounds like lignin and hemicellulose [46,49]. The samples of *H. excelsum*, *B. rubescens*, and especially *E. odora* showed higher intensity in these bands, indicating a chemical composition rich in aromatic lignin and condensed phenolic compounds, which justifies their negative projection in PC2. The Lecythidaceae, Fabaceae, and Moraceae families, also projected in the negative region of PC2, show spectral similarities in these regions, associated with the presence of substituted aromatic structures and aliphatic lignin chains. On the other hand, the species grouped in the positive quadrant of PC2 present a spectral profile with lower contributions of phenolic compounds, reflecting a simpler or less diversified lignocellulosic composition.

PC3, which explains 14.25% of the total variance, played an important role in differentiating species and botanical families. In Case 1, although PC3 did not show broad and well-defined groupings, it was able to more strongly group *M. coriacea* and some samples of *T. paniculata* in the positive quadrant, while the other species were grouped in the negative quadrant. In Case 2, the Myristicaceae family, represented by *V. surinamensis*, was projected in the positive quadrant, standing out from the other families. This indicates that *V. surinamensis* shares spectral characteristics with *M. coriacea* and *T. paniculata*, but with unique behavior.

The PC3 loadings (Fig. 3E) indicate that the most influential spectral bands for the formation of this component are the 5257–5438 cm^{-1} and

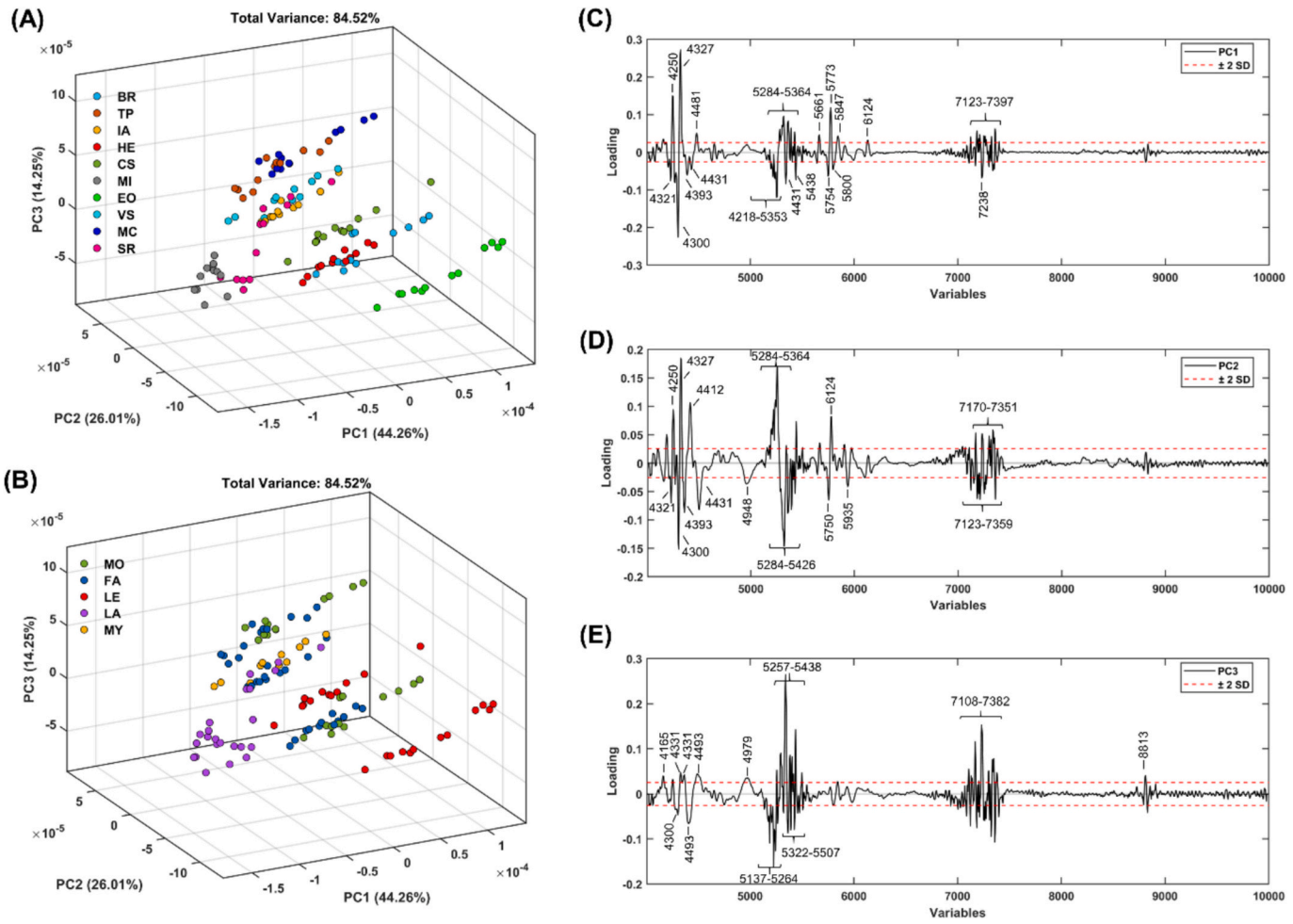

Fig. 3. PCA analysis of the wood samples grouped by species and botanical family. (A) Score plot for the samples grouped by species, for the first three principal components. (B) Score plot for the samples grouped by botanical family, for the same three principal components. (C), (D), and (E) Loadings for components PC1, PC2, and PC3, respectively, showing the most influential spectral bands for the formation of each component. The discriminant spectral regions are highlighted with the corresponding wavenumber values (cm^{-1}). The dashed line represents the ±2 standard deviation (±2 SD) limits for each component. *Brosimum rubescens* (BR), *Tachigali paniculata* (TP), *Inga alba* (IA), *Hymenolobium excelsum* (HE), *Couratari stellata* (CS), *Mezilaurus itauba* (MI), *Eschweilera odora* (EO), *Virola surinamensis* (VS), *Maquira coriacea* (MC), and *Sextonia rubra* (SR).

7108–7382 cm^{-1} regions, associated with C—O and C—H vibrations in phenolic compounds and lignin, which were the most prominent in the positive loadings [6,48,49].

These bands suggest that MC, *T. paniculata*, and *V. surinamensis* present a chemical profile enriched in substituted ligno-aromatic structures, with higher intensity in these regions. In contrast, the most significant negative loadings in the 5137–5264 cm^{-1} and 5322–5507 cm^{-1} regions, associated with overtones of methyl (–CH_3) and methylene (–CH_2–) groups, as well as vibrations of guaiacyl/syringyl aromatic rings, were observed in other families such as Lauraceae and Lecythidaceae, suggesting a more homogeneous or less enriched chemical composition in aromatic compounds [18,49,50]. The projection of *V. surinamensis* in the positive quadrant of PC3, while less pronounced, suggests its spectral profile is more closely aligned with *M. coriacea* and *T. paniculata*, likely due to specific variations in its lignocellulosic and phenolic constituents.

3.3. Woods authentication using DD-SIMCA

To achieve a more accurate and robust classification of the samples belonging to the target species, the supervised DD-SIMCA approach was applied. This technique allows modeling only the class of interest (target class) and is particularly suitable for scenarios in which the main goal is to determine whether a sample belongs to a predefined class, even in the absence of information about other classes [34]. Accordingly, ten independent models were built, each corresponding to one of the analyzed species, considered individually as the target class (Fig. 4). This strategy enabled the evaluation of the algorithm's ability to capture the characteristic spectral patterns of each species while simultaneously discriminating the remaining samples, treated as the external class in each model.

The adopted approach followed what is referred to as the “rigorous” model, meaning that the models were built exclusively based on samples from the target class, without including information from external classes during the calibration step (Table 3). The optimization of each model was focused on minimizing the Type I error (α), which corresponds to the incorrect rejection of true target class samples (false negatives). In this context, a fixed α value of 0.01 was used, meaning that at most one false rejection is tolerated for every 100 genuine samples, thereby ensuring high sensitivity (1 – α) [36,37]. This choice is particularly relevant in real-world applications, where the primary objective is to confirm with high confidence whether a sample belongs to the modeled species. Furthermore, the optimal number of principal components (PCs) in each DD-SIMCA model was selected based on the best sensitivity value observed during cross-validation and calibration, aiming to reduce the false negative rate. This criterion is essential in

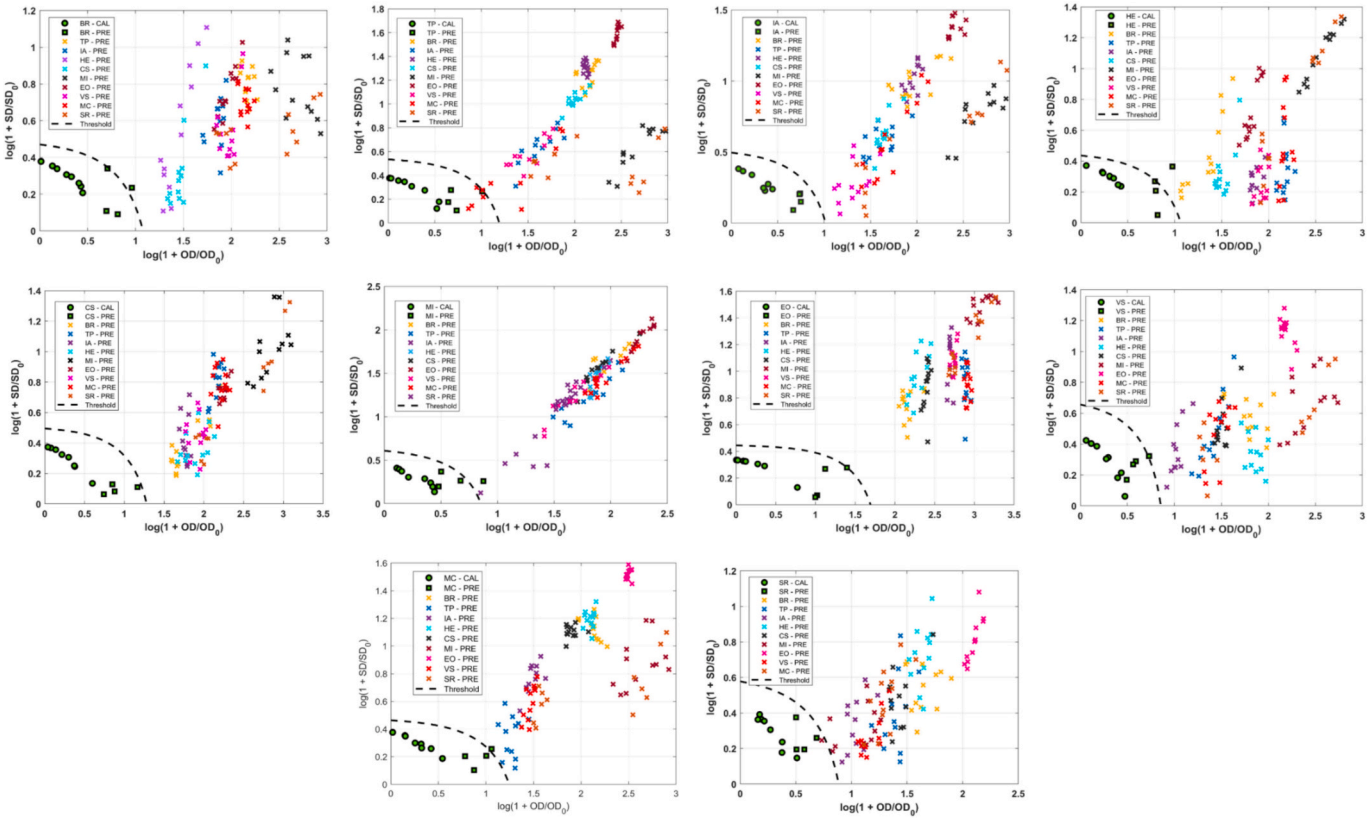

Fig. 4. Acceptance/rejection plots of the DD-SIMCA models generated for the ten wood species using the "rigorous" approach, in which the model is calibrated exclusively with samples from the target class. Each plot represents the space defined by the orthogonal distances (OD) and scores (SD), with dashed lines indicating the critical limits (α = 0.01). Target class samples are shown as green circles (calibration) and green squares (validation). Samples from external classes (other species) are displayed as colored "X"s, indicating their species of origin. Calibration samples (CAL), Prediction samples (PRE), *Brosimum rubescens* (BR), *Tachigali paniculata* (TP), *Inga alba* (IA), *Hymenolobium excelsum* (HE), *Couratari stellata* (CS), *Mezilaurus itauba* (MI), *Eschweilera odora* (EO), *Virola surinamensis* (VS), *Maquira coriacea* (MC), *Sextonia rubra* (SR). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Performance metrics of the DD-SIMCA models developed for each of the ten wood species, using the rigorous approach (α = 0.01).

Target Class	PCS n°	CV	TRAIN			TEST								
		SEN (%)	TP	FN	SEN (%)	FP	TN	TP	FN	SPE (%)	SEN (%)	EFF (%)	ACC (%)	MCC (%)
Hymenolobium excelsum	6	60.1	8	0	100	0	107	3	1	100	75.0	86.6	99.1	86.2
Inga alba	4	71.4	7	0	100	0	108	4	0	100	100	100	100	100
Tachigali paniculata	5	62.5	8	0	100	4	103	4	0	96.2	100	98.1	96.4	69.3
Sextonia rubra	4	73.2	8	0	100	1	106	4	0	99.5	100	99.1	99.1	89.0
Mezilaurus itauba	4	70.5	8	0	100	0	107	3	1	100	75.0	86.6	99.1	86.2
Eschweilera odora	5	62.0	8	0	100	0	107	4	0	100	100	100	100	100
Couratari stellata	5	75.0	8	0	100	0	107	4	0	100	100	100	100	100
Maquira coriacea	5	68.5	8	0	100	0	107	3	1	100	75.0	86.6	99.1	86.2
Brosimum rubescens	4	75.0	8	0	100	0	107	3	1	100	75.0	86.6	99.1	86.0
Virola surinamensis	5	62.5	8	0	100	0	107	4	0	100	100	100	100	100

Principal Components (PCs), false negative (FN), true positive (TP), false positive (FP), true negative (TN), specificity (SPE), sensitivity (SEN), efficiency (EFF), accuracy (ACC), and Matthews correlation coefficient (MCC).

practical applications such as botanical authentication, origin certification, or quality control, where confidence in the correct identification of the modeled class (species) is crucial.

The methodological framework presented here represents a significant advancement beyond our group's previous applications of FT-NIR to Amazonian woods. While earlier studies focused either on wood property prediction for the genus *Eschweilera* [51] or on species discrimination within individual families using LDA [6], the present work introduces three fundamental innovations: (i) expanded taxonomic scope, encompassing ten species from five families, ensuring PLSR models are calibrated across a broader range of chemical and physical variability; (ii) methodological evolution, the introduction of DD-SIMCA as a statistically rigorous one-class classifier, which offers objective acceptance thresholds based on Type I error control (α = 0.01) and is better suited for forensic and traceability applications where the universe of potential species is vast and incompletely known [33]; and (iii) workflow integration, for the first time, uniting species authentication and quality prediction into a single, sequential process that mirrors real-world decision-making: taxonomic identity must be established before appropriate end-use can be determined. By transforming fragmented analytical capabilities into a unified decision-support framework, this study moves beyond proof-of-concept toward operational applicability for sustainable management of fallen Amazonian timber.

Fig. 4 presents the acceptance/rejection plots obtained for the ten

models generated using DD-SIMCA. In each plot, the samples are projected in the space defined by orthogonal distances (OD) and score distances (SD), with dashed lines delimiting the model's critical acceptance limits ($\alpha = 0.01$). Target class samples are represented by green circles (calibration) and green squares (prediction), while samples from external classes appear as "X" symbols, color-coded according to their actual species. Overall, all models showed excellent performance during the calibration phase, with sensitivity (SEN) and accuracy (ACC) of 100%, indicating that the spectral profiles were effective in capturing intraclass variability and distinguishing the target class. However, acknowledging the limitation imposed by the restricted number of available samples, the models were tested under a more challenging condition, simulating a realistic scenario, by being developed with only 12 target class samples, distributed between calibration and prediction. This approach aimed to evaluate the algorithm's ability to identify the unique spectral patterns of each species even with limited data, and to assess its classification performance in practical authentication situations.

During the prediction phase, most models maintained high performance metrics. However, four species, *Brosimum rubescens*, *Hymenolobium excelsum*, *Mezilaurus itauba*, and *Maquira coriacea*, each had one target class sample incorrectly rejected, resulting in reduced sensitivity (75.0%) for these classes and contributing to a decrease in overall accuracy to 99.1%. Two primary factors may explain this occurrence. First, traditional taxonomic identification, based on subtle visual and anatomical characteristics, is vulnerable to human error, particularly with morphologically similar species. Second, the observed result may be associated with biodegradation, given that many of the fallen trees presented some degree of degradation.

These results highlight the potential of DD-SIMCA models as an objective and reliable tool to support or validate visual taxonomic classifications, particularly in quality control or traceability processes involving tropical woods. To strengthen the evaluation, the Matthews Correlation Coefficient (MCC) was included as a robust performance metric, particularly relevant in scenarios involving imbalanced classes. The MCC values obtained ranged from 69.3% to 100%, demonstrating that, despite sample size limitations, most models exhibited satisfactory classification performance. However, the results also highlight the need to expand the dataset in future studies in order to further enhance the robustness of the models in the face of natural intraspecies variability.

Among the evaluated models, the performance of *Tachigali paniculata* and *Sextonia rubra* stands out due to the occurrence of false positives (FP), which negatively impacted the specificity of the models. The model for TP incorrectly accepted four external class samples, while SR had one FP, resulting in specificities of 96.2% and 95.5%, respectively, and an overall accuracy still above 96.4%. These results suggest that samples accepted by the model despite being misclassified deserve scrutiny: their high chemical similarity to the target class may indicate an initial taxonomic error, a issue also documented by Eugenio da Silva et al. [6] in their work on Myristicaceae woods.

In some DD-SIMCA models, a significant proximity was observed between the OD and SD distances of certain external class samples and the acceptance region of the target class, highlighting the algorithm's ability to capture real spectral similarities associated with the lignocellulosic composition of the woods. For example, the model for *T. paniculata* presented a high rate of false positives due to its strong spectral similarity with *M. coriacea*, as previously evidenced by PCA, reflected in intense bands at 5257–5438 cm^{-1} and 7108–7382 cm^{-1} (C—O and C—H of substituted lignins). Similarly, *H. excelsum* presented spectra similar to those of *B. rubescens*, both presenting a strong contribution from aromatic structures (5284–5426 cm^{-1} and 7123–7359 cm^{-1}), possibly due to shared bands associated with hemicellulose and aliphatic chains (5137–5264 cm^{-1}). On the other hand, the models for *Couratari stellata* and *Eschweilera odora* revealed greater spectral distance from the other species, both in OD and SD, indicating a distinctive chemical pattern for these woods and reinforcing the robustness of these models for separation and authentication.

Nevertheless, it is important to emphasize that, due to the limited sample size available, this approach was primarily conducted as an exploratory investigation to determine whether the algorithm exhibits a reliable tendency to capture the necessary spectral information for both classification and regression tasks. Therefore, while these preliminary findings demonstrate the feasibility of the method, this spectral tendency must be investigated more precisely and robustly using a significantly larger sample set in future studies.

3.4. Prediction of physicochemical properties using PLS models

The development of a rapid analytical methodology capable of predicting the physicochemical parameters of fallen wood represents an efficient alternative to traditional methods, which are generally more time-consuming and require extensive laboratory preparation. Such an approach enables continuous and real-time monitoring of the quality and reuse potential of these materials, thereby facilitating their screening, classification, and appropriate destination. In this context, the NIR spectra obtained for the wood samples were calibrated using PLS regression based on previously optimized models. Table 4 presents the main statistical parameters of the PLSR models developed, summarizing their predictive performance and the robustness of the calibrations employed.

The interpretation of PLSR model performance requires careful consideration of the property ranges represented in the calibration set. While the R^2 values reported in Table 4 are high (generally >0.88), it should be noted that several properties present relatively narrow natural ranges within our sample set, for example, lignin content varies from approximately 27% to 32%, and basic density from 0.43 to 0.74 g/cm^3. This raises an important question: do high R^2 values in narrow-range calibrations indicate genuine predictive power or statistical artifacts?

The chemometrics literature provides guidance on this point. Santos et al. [52] developed NIR-PLSR models for Kappa number prediction using a range of only 10–17 (a similarly narrow interval) and obtained R^2 values of 92.1% with RPD of 3.6. They concluded that such models are sufficiently accurate to be used in screening programs and in quality control. This is consistent with the widely accepted interpretation that RPD values >3 indicate models suitable for screening applications [6]. All our models meet this criterion, with RPD values ranging from 3.05 (ash) to 5.82 (modulus of rupture).

Furthermore, the RMSEP values we report provide a more operationally meaningful metric than R^2 alone. For lignin content, an RMSEP of 0.99% represents a relative prediction error of approximately 3–4% relative to the mean value. For basic density, an RMSEP of 0.103 g/cm^3 corresponds to relative errors of 14–24% depending on the species, which is adequate for preliminary material sorting but would need improvement for precise engineering applications.

It is also important to recognize that the reference methods used to generate the calibration data carry their own measurement uncertainty. As discussed by Faber et al. [53], the 'apparent' prediction error (RMSEP) includes both the true prediction uncertainty and the reference method error. This means that the actual predictive ability of the NIR models may be somewhat better than the raw statistics suggest, although we have chosen to maintain a conservative reporting approach throughout this study.

Nevertheless, we acknowledge that the relatively narrow property ranges limit the extrapolation capacity of our models beyond the calibrated values. The models should be viewed as robust screening tools suitable for preliminary classification of fallen timber within the range of properties encountered in our sample set. Future work should aim to expand the calibration range by including samples from a broader geographical distribution and from trees presenting greater phenotypic variation.

The PLSR models built from NIR spectra presented strong predictive performance for both chemical and physical–mechanical properties of

Table 4
Statistical parameters for the Partial Least Squares (PLS) regression models for the prediction of wood properties.

Properties	Analyte	LVs	R^2c	R^2p	RMSEC	RMSEP	RPD
Chemical	Ethanol-toluene extractives	10	0.99	0.93	0.097	0.415	4.09
	Total polyphenols	7	0.95	0.90	2.480	4.680	4.66
	Hot water extractives	7	0.98	0.93	0.676	1.280	3.22
	Lignin	4	0.98	0.97	0.457	0.990	5.43
	Holocellulose	7	0.99	0.89	0.122	1.385	3.85
	Ash	8	0.97	0.88	0.110	0.470	3.05
	Moisture	3	0.99	0.93	0.150	0.918	3.53
Physical-Mechanical	Higher heating value	6	0.95	0.93	49,000	62,000	3.99
	Wood density	3	0.91	0.90	0.071	0.103	3.69
	Modulus of elasticity	10	0.90	0.88	1250	1910	4.09
	Modulus of rupture	4	0.95	0.93	16,100	28,000	5.82

LVs: Number of Latent Variables; R^2c: Determination coefficient of calibration; R^2p: Determination coefficient of prediction; RMSEC: Root Mean Square Error of Calibration; RMSEP: Root Mean Square Error of Prediction; RPD: Ratio of Performance to Deviation.

fallen wood. Chemical attributes such as lignin, total polyphenols, and ethanol–toluene extractives showed particularly robust models, with high R^2p (0.90–0.97) and RPD values exceeding 4, reflecting strong calibration stability and low prediction errors. The physical–mechanical properties also performed well, especially modulus of rupture (RPD = 5.82) and modulus of elasticity (RPD = 4.09), demonstrating the capacity of NIR–PLSR to capture key mechanical behavior relevant to material classification and reuse.

Additionally, porosity [43] and Janka hardness, transverse to the fibers [44], were estimated using regression equations, with wood density as the dependent variable, allowing for a broader assessment of wood properties.

The analyzed fallen-tree samples covered a broad range of physical-mechanical behavior, which was sufficient to support classification and calibration modeling. Moisture content was narrowly distributed (11.56–13.08%, with no significant differences), while porosity (Φ) varied widely (50.67–71.67%). Less-dense woods such as *Tachigali paniculata* and *Virola surinamensis* presented higher porosity and correspondingly lower stiffness and strength. Basic density (ρ_{bas}) reached its maximum in *Eschweilera odora* (0.74 g/cm^3) and remained high in *Brosimum rubescens*, *Hymenolobium excelsum*, and *Mezilaurus itauba* (~0.60 g/cm^3). Mechanical properties spanned E_{M0} of 8600–13,675 MPa and f_M of 66.06–136.85 MPa, with *E. odora* and *H. excelsum* at the upper end; Janka hardness peaked in *E. odora* (10,110N), supporting demanding uses such as flooring. The higher-density, high-strength woods are suited for structural or load-bearing applications [54], whereas lighter species are more appropriate for interior or secondary uses [55,56]. The higher heating value (HHV) ranged from 4248 to 4959 kcal/kg, indicating good energy potential for several species (Table 5).

This physical diversity is mirrored by chemical variability, which underpins natural durability and processing behavior [57,58]. *M. itauba* presented the highest cellulose content (65.23%) and elevated extractives (5.79%), factors that favor dimensional stability and biological resistance. Lignin contents ≥30.20% were observed in *E. odora*, *H. excelsum*, *Sextonia rubra*, and *T. paniculata*, indicating enhanced natural durability despite differences in density (Table 5). Total polyphenols and extractives were high in *H. excelsum*, *Maquira coriacea*, and *M. itauba*, further contributing to resistance against fungi and insects. Conversely, ash content, which reflects mineral load, was higher in *Couratari stellata*, *B. rubescens*, and *M. coriacea*, with potential implications for tool wear and reduced energy efficiency [59,60]. Together, these chemical signatures, ranging from lignin-rich, durable matrices to mineral-dense woods, directly influence their technological characteristics. The accurate prediction of these multivariable characteristics represents the key challenge addressed by the chemometric models explored in this work.

A systematic optimization and construction of the PLS models [61,62] allowed for accurate prediction of the chemical and physical properties of the fallen tree samples based on their FT-NIR spectra. The grid search approach was successful in identifying the optimal combination of spectral preprocessing and data scaling for each specific property, highlighting the need for an individualized modeling strategy. This property-specific calibration was necessary to maximize the predictive ability of the models.

Taken together, the predictive models provide a robust, quantitative framework for rapid wood characterization. The estimated chemical profiles, from lignin-rich, durable matrices to mineral-dense woods, offer qualitative insights into potential technological performance and natural durability. For instance, woods like *M. itauba*, predicted to have high cellulose and extractives, or *E. odora* and *H. excelsum*, estimated to have high lignin, can be flagged for applications requiring dimensional stability or biological resistance. Conversely, the model estimates for higher ash content in species like *C. stellata* provide a preliminary warning for potential issues like tool wear. These results confirm that FT-NIR spectroscopy, supported by a robust chemometric modeling approach, offers a rapid and non-destructive tool for the estimation and comprehensive screening of wood from fallen trees. The high level of accuracy achieved for key mechanical and chemical properties offers a valuable alternative to traditional destructive techniques. This approach not only optimizes the efficiency of the material evaluation process but also promotes a more efficient and sustainable utilization of this valuable but frequently overlooked forest resource by enabling a data-driven, preliminary sorting based on estimated property profiles.

The proposed workflow integrates two complementary chemometric approaches implemented in different software environments. First, the DD-SIMCA models (MATLAB) authenticate wood samples by verifying their taxonomic identity based on spectral fingerprints. Second, the PLS models (TQ Analyst) predict key technological properties from the same spectra. Although the preprocessing strategies and modeling algorithms differ between the two software platforms—as each was optimized for its specific analytical goal, the sequential application of these tools to the same dataset provides a coherent framework for decision-making. This integration allows for the rapid screening of fallen timber, combining taxonomic authentication with quality assessment in a non-destructive manner.

4. Conclusions

This study demonstrates the feasibility of a non-destructive methodology based on FT-NIR spectroscopy combined with chemometric approaches for the authentication and quality prediction of wood from naturally fallen trees in the Central Amazon. The classification models (DD-SIMCA) presented satisfactorily performance, achieving high accuracy in species authentication even under conditions of chemical similarity, confirming the robustness of spectral fingerprinting for taxonomic discrimination. In parallel, the PLSR models provided reliable predictions for important physical, mechanical, and chemical properties, supporting the use of this approach as a rapid screening tool for evaluating wood quality and potential end uses.

Table 5
Physical-mechanical and chemical properties of wood from natural fallen trees in Central Amazonia, estimated by near-infrared (NIR) spectroscopy.

Properties		Physical -	Mechanical				Chemicals				
Natural fallen trees	Moisture (%)	ρ_{bas} (g/cm^3)	Φ (%)	HHV (Kcal/g)	E_{M0} (MPa)	f_M (MPa)	f_{H90} (N)	Extractives (%)	Hot water (%)r	Total poly phenols (%)	Lignin (%)	Holo cellulose (%)	Ash (%)
	12.48^a	0.60^{ab}	60.33^b	4775^b	12810^{ab}	136.85^a	6168^b	4.36^b	7.25^b	2.63^a	31.44^a	46.92^e	1.15^b
Hymenolobium excelsum	(±0.56)	(±0.08)	(±0.56)	(±43)	(±269)	(±9.76)	(±189)	(±0.56)	(±0.60)	(±0.75)	(±1.58)	(±0.83)	(±0.52)
	12.61^a	0.57^{ab}	62.33^b	4794^b	12530^{ab}	87.96^{def}	5378^b	4.86^{ab}	5.82^c	0.94^b	29.90^{ab}	51.56^{de}	0.58^c
Inga alba	(±0.31)	(±0.02)	(±0.31)	(±53)	(±170)	(±6.21)	(±160)	(±0.76)	(±0.77)	(±0.46)	(±0.25)	(±1.12)	(±0.16)
	11.56^a	0.43^b	71.67^a	4846^a	10445^{bcd}	77.18^{ef}	2846^c	5.80^a	8.90^b	0.41^c	31.74^a	59.03^b	1.51^a
Tachigali paniculata	(±0.22)	(±0.01)	(±0.47)	(±51)	(±629)	(±1.63)	(±95)	(±0.68)	(±0.90)	(±0.15)	(±1.27)	(±2.68)	(±0.45)
	12.37^a	0.60^{ab}	60.00^b	4959^a	12850^{ab}	86.06^{ef}	6250^b	3.91^b	11.93^a	1.12^b	30.79^a	60.03^b	0.54^c
Mezilaurus itauba	(±0.65)	(±0.06)	(±3.77)	(±67)	(±919)	(±1.43)	(±139)	(±1.07)	(±0.78)	(±0.24)	(±0.11)	(±1.07)	(±0.21)
	11.74^a	0.55^{ab}	63.67^{ab}	4868^a	11275^{abc}	95.83^{cde}	5164^b	5.79^a	12.38^a	2.22^a	28.70^{ab}	65.23^a	0.51^c
Sextonia rubra	(±0.05)	(±0.12)	(±8.01)	(±70)	(±1.025)	(±1.39)	(±250)	(±0.75)	(±0.54)	(±0.12)	(±0.59)	(±0.23)	(±0.10)
	13.08^a	0.53^b	50.67^c	4758^b	11250^{abc}	111.52^{bc}	4530^{bc}	1.76^c	2.37^d	0.18^d	30.20^a	48.83^{de}	1.31^b
Couratari stellata	(±0.25)	(±0.02)	(±3.77)	(±49)	(±353)	(±2.79)	(±426)	(±0.59)	(±0.13)	(±0.07)	(±0.55)	(±0.78)	(±0.37)
	12.53^a	0.74^a	65.00^{ab}	4615^b	13675^a	121.44^{ab}	10110^a	1.05^c	1.22^d	0.75c	29.19^{ab}	51.16^{de}	1.76^a
Eschweilera odora	(±0.57)	(±0.06)	(±1.41)	(±79)	(±728)	(±5.81)	(±168)	(±0.45)	(±0.88)	(±0.21)	(±0.43)	(±1.10)	(±0.48)
	12.51^a	0.60^{ab}	60.33^b	4674^b	12270^{abc}	117.83^{abc}	6083^b	5.87^a	8.70^b	2.22^a	29.60^{ab}	57.21^{bc}	1.64^a
Brosimum rubescens	(±0.59)	(±0.01)	(±0.47)	(±79)	(±381)	(±0.06)	(±172)	(±0.63)	(±0.77)	(±0.12)	(±0.76)	(±1.30)	(±0.28)
	12.27^a	0.57^{ab}		4930^a	8600^d	96.90^{bcde}	5503^b	2.29^c	2.07^d	0.17^d	29.20^{ab}	51.98^{cd}	1.85^a
Maquira coriacea	(±1.12)	(±0.03)	62.00^b (±1.88)	(±33)	(±678)	(±5.37)	(±648)	(±0.61)	(±1.57)	(±0.06)	(±0.50)	(±0.66)	(±0.34)
	12.72^a	0.45^b	70.33^a	4248^c	9630^{cd}	66.06^f	3129^c	3.02^b	8.00^b	0.58^c	27.04^b	53.63^{de}	0.66^c
Virola surinamensis	(±0.93)	(±0.02)	(±0.47)	(±50)	(±890)	(±8.58)	(±105)	(±0.93)	(±0.75)	(±0.22)	(±0.77)	(±1.58)	(±0.20)

ρ_{bas} = wood density; Φ = porosity; HHV = higher heating value; E_{M0} = modulus of elasticity; f_M = modulus of rupture; f_{H90} = Janka hardness. Value in parentheses = standard deviation; means followed by the same letter in the same column do not differ from each other by Tukey's test at 95% probability ($p < 0.05$).

These results highlight the potential of FT-NIR spectroscopy as a practical and efficient tool for the valorization of fallen timber, enabling data-driven allocation of this underutilized resource and contributing to more sustainable forest management strategies. However, it is important to emphasize that, due to the severe logistical constraints of sampling in the Amazon rainforest, the current models were developed using a limited sample size and rely solely on internal validation. Consequently, this study was primarily conducted as an exploratory proof-of-concept to determine the algorithm's tendency to capture the necessary spectral information. While the models demonstrated strong predictive performance and high classification accuracy within the scope of this study, some aspects warrant further investigation to expand their applicability. The sample size per species, although consistent with standard practices in wood anatomy and chemometric studies, may not capture the full extent of intraspecific variability across the geographical distribution of these species. Additionally, the narrow range observed for certain properties suggests that future efforts should focus on incorporating broader natural variability through expanded sampling across diverse forest sites and environmental conditions. Such endeavors will enable model refinement and independent validation, ultimately enhancing robustness and generalization capacity for large-scale operational applications, supporting the transition of this methodology from proof-of-concept to a fully operational tool for sustainable forest management and circular bioeconomy initiatives.

CRediT authorship contribution statement

Cristiano S. do Nascimento: Writing – review & editing, Writing – original draft, Validation, Investigation, Data curation, Conceptualization. **Jelmir C. de Andrade:** Writing – review & editing, Validation, Methodology, Formal analysis. **Claudia Eugenio da Silva:** Writing – original draft, Visualization, Investigation, Data curation. **Vinicius Kartnaller:** Validation, Software, Formal analysis, Conceptualization. **Moacir Alberto A. Campos:** Visualization, Software, Methodology, Conceptualization. **Roberto D. de Araújo:** Visualization, Methodology, Conceptualization. **Adrya da S. Figueiredo:** Visualization, Methodology, Conceptualization. **José Carlos R. Soares:** Writing – original draft, Methodology, Formal analysis, Conceptualization. **Marcia C. Ramos de Oliveira:** Methodology, Formal analysis, Conceptualization. **Niro Higuchi:** Supervision, Resources, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank the MCTI/CNPq/FAPEAM/Brazil – "INCT-Madeiras da Amazônia;" CNPq/MCTI/FNDCT/Brazil - No.19/2024, PROA-O papel da Floresta amazônica no processone adaptação à mudança do climática – ciclagens da água e carbono" for financial support, . A.J.O. Godoy for data analysis and English language review. The authors also thank CAPES/Brazil for their support, linked to PPGCFT, PPGBOT and PPGECO/INPA/MCTI.

Data availability

Data will be made available on request.

References

[1] R.I. Negrón-Juárez, J.A. Holm, D.M. Marra, S.W. Rifai, W.J. Riley, J.Q. Chambers, C.D. Koven, R.G. Knox, M.E. McGroddy, A.V. Di Vittorio, J. Urquiza-Muñoz,

R. Tello-Espinoza, W.A. Muñoz, G.H.P.M. Ribeiro, N. Higuchi, Vulnerability of amazon forests to storm-driven tree mortality, Environ. Res. Lett. 13 (2018) 054021, https://doi.org/10.1088/1748-9326/aabe9f.
[2] H. Angelo, A.N. Almeida, E.A.T. Matriacardi, C.F. Roseti, R.O. Gaspar, E.P. Miguel, P.G.A. Vasconcelos, Determinants of profit in sustainable forest management in the brazilian amazon, Afr. J. Agric. Res. 11 (2016) 4498–4503, https://doi.org/10.5897/AJAR2016.11764.
[3] F.G. Higuchi, N.T. Nakajima, S.A. Machado, J. Santos, A.J.N. Lima, N. Higuchi, The use of natural fallen trees in mature tropical amazonian forest for volumetric regression analysis and volume equations adjustment, Aust. J. Basic Appl. Sci. 9 (2015) 342–348.
[4] K. Lima, A. Castro, J. Baptista, U. Silva, Wood-logging process management in eastern Amazonia (Brazil), Sustainability 12 (2020) 7571, https://doi.org/10.3390/su12187571.
[5] J.A. Freitas, F.J. Vasconcellos, Identificação de Madeiras Comerciais da Amazônia, INPA, 2019.
[6] C. Eugenio da Silva, C.S. Nascimento, J.A. Freitas, R.D. Araújo, C.C. Nascimento, N. Higuchi, Application of infrared spectroscopy in the identification of fallen trees from the amazon rainforest (myristicaceae), For. Sci. 70 (2024) 102–112, https://doi.org/10.1093/forsci/fxae001.
[7] S. Tsuchikawa, H. Kobori, A review of recent application of near infrared spectroscopy to wood science and technology, J. Wood Sci. 61 (2015) 213–220, https://doi.org/10.1007/s10086-015-1467-x.
[8] L. Jiao, Y. Lu, T. He, J. Guo, Y. Yin, DNA barcoding for wood identification: global review of the last decade and future perspective, IAWA J. 41 (2020) 620–643, https://doi.org/10.1163/22941932-bja10041.
[9] V.R. Castro, P.G. Surdi, S.A. Fernandes, M.S. Berger, A.J.V. Zanucio, J.C. Zanucio, S.O. Araujo, Chemical composition of heartwood and sapwood of tectona grandis characterized by CG/MS-PY, Sci. Rep. 12 (2022) 18441, https://doi.org/10.1038/s41598-022-22800-1.
[10] S. Tsuchikawa, T. Inagaki, T. Ma, Application of near-infrared spectroscopy to forest and wood products, Curr. For. Rep. 9 (2023) 401–412, https://doi.org/10.1007/s40725-023-00203-3.
[11] F.B.B.D. Santos, C.C.D. Nascimento, D.R. Galbraith, G.L.R.C. Silva, R.D.D. Araujo, C.S. Nascimento, M.D.P. Lima, Use of impulse tomography in the evaluation of manilkara huberi (maçaranduba) managed of the amazon rainforest, wood, Mater. Sci. Eng. 18 (2023) 975–985, https://doi.org/10.1080/17480272.2022.2098054.
[12] S. Dierickx, S. Genbrugge, H. Beeckman, W. Hubau, P. Kibleur, J.V. den Bulcke, Non-destructive wood identification using x-ray μCT scanning: which resolution do we need? Plant Methods 20 (2024) 98, https://doi.org/10.1186/s13007-024-01216-0.
[13] A.R. Geus, A.R. Backes, A.B. Gontijo, G.H.Q. Albuquerque, J.R. Souza, Amazon wood species classification: a comparison between deep learning and pre-designed features, Wood Sci. Technol. 55 (2021) 857–872, https://doi.org/10.1007/s00226-021-01282-w.
[14] L. Schimleck, J. Dahlen, L.A. Apiolaza, G. Downes, G. Emms, R. Evans, J. Moore, L. Pâques, J. Van den Bulcke, X. Wang, Non-destructive evaluation techniques and what they tell us about wood property variation, Forests 10 (2019) 728, https://doi.org/10.3390/f10090728.
[15] H. Pereira, A.J.A. Santos, O. Anjos, Fibre morphological characteristics of Kraft pulps of acacia melanoxylon estimated by NIR-PLS-R models, Materials 9 (1) (2016) 8, https://doi.org/10.3390/ma9010008.
[16] D. Fengel, G. Wegener, Wood: Chemistry, Ultrastructure, Reactions, Walter de Gruyter, Berlin, 2003.
[17] M.S. Vieira, C.C. Nascimento, C. Higuchi, K.M.M. Pacheco, Evaluation of the quality of wood from naturally fallen trees in the central Amazon, Int. J. Innov. Educ. Res. 8 (2020) 564–579, https://doi.org/10.31686/ijier.vol8.iss8.2564.
[18] C.S. Nascimento, I.A. Cruz, R.D. Araújo, J.C.R. Soares, C. Eugenio da Silva, C. C. Nascimento, J. Santos, N. Higuchi, A rapid and nondestructive method for the prediction of lignin content in tropical amazon woods using FT-NIR spectroscopy, J. Indian Acad. Wood Sci. 21 (2024) 123–134, https://doi.org/10.1007/s13196-024-00331-8.
[19] C. Pasquini, Near infrared spectroscopy: a mature analytical technique with new perspectives - a review, Anal. Chim. Acta 1026 (2018) 8–36, https://doi.org/10.1016/j.aca.2018.04.004.
[20] J.C. Andrade, D. Galvan, L. Effting, C. Lelis, F.L. Melquiades, E. Bona, C.A. Conte-Junior, An easy-to-use and cheap analytical approach based on NIR and chemometrics for tomato and sweet pepper authentication by non-volatile profile, Food Anal. Methods 16 (2023) 567–580, https://doi.org/10.1007/s12161-022-02439-4.
[21] L.M. Santos, E.A. Amaral, E.M. Nieri, E.V.S. Costa, P.F. Trugilho, N. Calegário, P.R. G. Hein, Estimating wood moisture by near infrared spectroscopy: testing acquisition methods and wood surfaces qualities, wood, Mater. Sci. Eng. 16 (2021) 336–343, https://doi.org/10.1080/17480272.2020.1768143.
[22] B.A. Loureiro, T.G. Arriel, F.M. Guedes Ramalho, P.R.G. Hein, P.F. Trugilho, Nir-based models for estimating selected physical and chemical wood properties from fast-growing plantations, iForest 15 (2022) 372–380, https://doi.org/10.3832/ifor4030-015.
[23] Z. Yuan, H. Wang, L. Cai, T. Sun, Rapid and non-destructive detection of wood density based on NIR hyperspectral imaging technology and moisture correction methods, Spectrochim. Acta A Mol. Biomol. Spectrosc. 341 (2025) 126410, https://doi.org/10.1016/j.saa.2025.126410.
[24] Y. Li, M. Li, X. Li, Z. Liu, A. Ming, H. Lan, S. Ye, The abundance and structure of deadwood: a comparison of mixed and thinned chinese fir plantations, Front. Plant Sci. 12 (2021) 614695, https://doi.org/10.3389/fpls.2021.614695.
[25] R. Javier-Astete, J. Melo, J. Jimenez-Davalos, G. Zolla, Classification of amazonian fast growing tree species and wood chemical determination by FTIR and multivariate analysis, Sci. Rep. 13 (2023) 7827, https://doi.org/10.1038/s41598-023-35107-6.
[26] R.W. Myster, A comparison of the forest soils in the peruvian amazon: Terra firme, palm, white sand and Igarapé, J. Soil Sci. Environ. Manage. 8 (2017) 130–134, https://doi.org/10.5897/JSSEM2017.0644.
[27] I.A.W.A. Committee, IAWA list of microscopic features for hardwood identification, IAWA Bull. 10 (1989) 219–332.
[28] I.A.W.A. Committee, IAWA list of microscopic features for softwood identification, IAWA J. 25 (1) (2004) 1–70.
[29] R. Dias, S.L. Nam, A.P. de la Mata, M. Williams, I. Duchesne, M. Lamothe, N. Isabel, J.J. Harynuk, Classification of wood species in trade using metabolomic profiling by GC×GC-TOFMS, Wood Sci. Technol. 59 (2025), https://doi.org/10.1007/s00226-025-01657-3.
[30] D. Galvan, T.L. Conte, J.M.M. Liduário, J.C. Andrade, P. Casarin, E. Bona, Aplicativo Gamma-Gui: Uma interface gráfica amigável para análise exploratória de dados no MATLAB, Quím. Nova 48 (2025), https://doi.org/10.21577/0100-4042.20250206 e–20250206.
[31] P. Geladi, P. Linderholm, Principal component analysis, in: S. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, 2nd ed., Elsevier, 2020 https://doi.org/10.1016/B978-0-12-409547-2.14892-9.
[32] S. Wold, Pattern recognition by means of disjoint principal components models, Pattern Recogn. 8 (1976) 127–139.
[33] Y.V. Zontov, O.Ye. Rodionova, S.V. Kucheryavskiy, A.L. Pomerantsev, DD-SIMCA – A MATLAB GUI tool for data driven SIMCA approach, Chemom. Intell. Lab. Syst. 167 (2017) 23–28, https://doi.org/10.1016/j.chemolab.2017.05.010.
[34] O.Ye. Rodionova, A.L. Pomerantsev, Chemometric tools for food fraud detection: the role of target class in non-targeted analysis, Food Chem. 317 (2020) 126448, https://doi.org/10.1016/j.foodchem.2020.126448.
[35] A.L. Pomerantsev, O.Y. Rodionova, On the type II error in SIMCA method, J. Chemom. 28 (2014) 518–522, https://doi.org/10.1002/cem.2610.
[36] A.L. Pomerantsev, O.Y. Rodionova, Concept and role of extreme objects in PCA/SIMCA, J. Chemom. 28 (2014) 429–438, https://doi.org/10.1002/cem.2506.
[37] A.L. Pomerantsev, Acceptance areas for multivariate classification derived by projection methods, J. Chemom. 22 (2008) 601–609, https://doi.org/10.1002/cem.1147.
[38] American Society for Testing and Materials - ASTM, Annual Book of ASTM Standards, Section 4: Construction, Volume 04.10: Wood, ASTM, West Conshohocken, 2021.
[39] A. Ramadan, A. Nasser, Specific gravity, fiber length and chemical components of conocarpus erectus as affected by tree spacing, J. Agric. Environ. 7 (2008) 52–59.
[40] C.S. do Nascimento, R.D. de Araújo, C.E. da Silva, C.C. do Nascimento, V. da S. Menezes, J. dos Santos, Near infrared spectroscopy as a tool to discriminate tannins from amazonian species, Ciênc. Agrotec 46 (2022) e001422, https://doi.org/10.1590/1413-7054202246001422.
[41] Associação Brasileira de Normas Técnicas—ABNT, Design of wooden structures — ABNT NBR 7190, ABNT, Rio de Janeiro, 2022.
[42] British Standard – BS, Methods of Testing Small Clear Specimens of Timber, BS 373, British Standards Institution, London, 1999.
[43] J.G. Haygreen, J.L. Bowyer, Forest Product and Wood Science, 3rd ed, Iowa State University Press, 1996.
[44] H.J.B. de Araújo, Relações funcionais entre propriedades físicas e mecânicas de madeiras tropicais brasileiras, Floresta 37 (2007), https://doi.org/10.5380/rf.v37i3.9937.
[45] C. Halimu, A. Kasem, S.H. Newaz, Empirical comparison of area under ROC curve (AUC) and Mathew correlation coefficient (MCC) for evaluating machine learning algorithms on imbalanced datasets for binary classification, in: ACM International Conference Proceeding Series, 2019, pp. 1–6.
[46] L.Q. de Souza Leão, J.C. de Andrade, G.M. Marques, C.C. Guimarães, R.F. Vieira de Albuquerque, A.S.E. Silva, K.P. de Araujo, M.P. de Oliveira, A.F. Gonçalves, H. F. Figueiredo, D.L. Lira, M.A. Alves, C.A. Conte-Junior, P.F. de Aquino, Rapid prediction of cervical cancer and high-grade precursor lesions: an integrated approach using low-field ^{1}H NMR and chemometric analysis, Clin. Chim. Acta 574 (2025) 120346, https://doi.org/10.1016/j.cca.2025.120346.
[47] A. Sandak, J. Sandak, M. Negri, Relationship between near-infrared (NIR) spectra and the physico-mechanical properties of wood, Wood Sci. Technol. 45 (2011) 35–48, https://doi.org/10.1007/s00226-010-0313-y.
[48] M. Schwanninger, J.C. Rodrigues, K. Fackler, A review of band assignments in near infrared spectra of wood and wood components, J. Near Infrared Spectrosc. 19 (2011) 287–308, https://doi.org/10.1255/jnirs.955.
[49] J. Sandak, A. Sandak, R. Meder, Assessing trees, wood and derived products with near infrared spectroscopy: hints and tips, J. Near Infrared Spectrosc. 24 (2016) 485–505, https://doi.org/10.1255/jnirs.12.
[50] S.S. Kelley, T.G. Rials, R. Snell, L.H. Groom, A. Sluiter, Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood, Wood Sci. Technol. 38 (2004) 257–276, https://doi.org/10.1007/s00226-003-0213-5.
[51] C.S. Nascimento, C.C. Nascimento, R.D. Araújo, J.C.R. Soares, N. Higuchi, 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 (2021), https://doi.org/10.3832/ifor3748-014.
[52] A.J.A. Santos, O. Anjos, R. Simões, J. Rodrigues, H. Pereira, Kappa number prediction of acacia melanoxylon unbleached Kraft pulps using NIR-PLSR models with a narrow interval of variation, BioResources 9 (2014) 6735–6744.

[53] N.M. Faber, F.H. Schreutelkamp, H.W. Vedder, Estimation of prediction uncertainty for a multivariate calibration model, Spectrosc. Eur. 16 (2004) 17–20.
[54] R.D. Araújo, C.S. Nascimento, C.C. Nascimento, J. Santos, Evaluation of the strength of edge-glued panels (EGP) manufactured with tropical woods from Brazil, Eur. J. Wood Prod. 83 (2025) 107, https://doi.org/10.1007/s00107-025-02264-6.
[55] Instituto Nacional de Pesquisas da Amazônia/Coordenação de Pesquisas Eem Produtos Florestais, Catálogo de Madeiras da Amazônia, INPA, 1991.
[56] D. Huang, J. Li, S. Li, J. Hu, Z. Cao, Z. Gou, Y. Ding, M. Zhu, Y. Chen, Self-densified super-strong wood, J. Bioresour. Bioprod. 10 (2025) 199–208, https://doi.org/10.1016/j.jobab.2025.03.001.
[57] M.S. Akhtar, M. Rafiullah, W.A. Shehata, A. Hossain, M. Ali, Comparative phytochemical, thin layer chromatographic profiling and antioxidant activity of extracts from some indian herbal drugs, J. Bioresour. Bioprod. 7 (2022) 128–134, https://doi.org/10.1016/j.jobab.2022.01.001.
[58] M.N. Acda, M.D. Rizare, A.P.G. Cantalejo, Decay, mold, and termite resistance of high density fiberboard from wood and chicken feather fibers, BioResources 20 (2025) 725–736, https://doi.org/10.15376/biores.20.1.725-736.
[59] M.N. Uddin, T. Ferdous, Z. Islam, M.S. Jahan, M.A. Quaiyyum, Development of chemometric model for characterization of non-wood by FT-NIR data, J. Bioresour. Bioprod. 5 (2020) 205–212, https://doi.org/10.1016/j.jobab.2020.07.005.
[60] L. Bianchini, A. Colantoni, R. Venanzi, L. Cozzolino, R. Picchio, Physicochemical properties of forest wood biomass for bioenergy application: a review, Forests 16 (2025) 702, https://doi.org/10.3390/f16040702.
[61] C.S. Nascimento, I.A. Cruz, NIR models for predicting chemical properties of Amazonian wood - INPA (version 1.0–10). http://portal.inpa.gov.br/cotei/2, 2020.
[62] C.S. Nascimento, C.C. Nascimento, M.M. Brasil, NIR models for predicting physico-mechanical properties of Amazonian wood - INPA (version 1.0–20). http://portal.inpa.gov.br/cotei/2, 2020.