

## RESEARCH ARTICLE

# Determination of hemicellulose, cellulose, holocellulose and lignin content using FTIR in *Calycophyllum spruceanum* (Benth.) K. Schum. and *Guazuma crinita* Lam.

Rosario Javier-Astete<sup>1</sup>, Jorge Jimenez-Davalos<sup>1</sup>, Gaston Zolla<sup>1,2\*</sup>

**1** Grupo de Investigacion en Mutaciones y Biotecnologia Vegetal, Facultad de Agronomia, Universidad Nacional Agraria La Molina - Lima, Lima, Peru, **2** Laboratorio de Fisiologia Vegetal, Facultad de Ciencias, Universidad Nacional Agraria La Molina - Lima, Lima, Peru

\* [gembz@yahoo.com](mailto:gembz@yahoo.com)

## OPEN ACCESS

**Citation:** Javier-Astete R, Jimenez-Davalos J, Zolla G (2021) Determination of hemicellulose, cellulose, holocellulose and lignin content using FTIR in *Calycophyllum spruceanum* (Benth.) K. Schum. and *Guazuma crinita* Lam.. PLoS ONE 16(10): e0256559. <https://doi.org/10.1371/journal.pone.0256559>

**Editor:** Daniel Cullen, USDA Forest Service, UNITED STATES

**Received:** August 7, 2021

**Accepted:** October 11, 2021

**Published:** October 27, 2021

**Copyright:** © 2021 Javier-Astete et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data are contained within the manuscript and [Supporting information files](#).

**Funding:** GZ: funded by Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica, grant number: 159-2018 FONDECYT-BM-IADT-AV The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Capirona (*Calycophyllum spruceanum* (Benth.) K. Schum.) and Bolaina (*Guazuma crinita* Lam.) are fast-growing Amazonian trees with increasing demand in timber industry. Therefore, it is necessary to determine the content of cellulose, hemicellulose, holocellulose and lignin in juvenile trees to accelerate forest breeding programs. The aim of this study was to identify chemical differences between apical and basal stem of Capirona and Bolaina to develop models for estimating the chemical composition using Fourier transform infrared (FTIR) spectra. FTIR-ATR spectra were obtained from 150 samples for each species that were 1.8 year-old. The results showed significant differences between the apical and basal stem for each species in terms of cellulose, hemicellulose, holocellulose and lignin content. This variability was useful to build partial least squares (PLS) models from the FTIR spectra and they were evaluated by root mean squared error of predictions (RMSEP) and ratio of performance to deviation (RPD). Lignin content was efficiently predicted in Capirona (RMSEP = 0.48, RPD > 2) and Bolaina (RMSEP = 0.81, RPD > 2). In Capirona, the predictive power of cellulose, hemicellulose and holocellulose models ( $0.68 < \text{RMSEP} < 2.06$ ,  $1.60 < \text{RPD} < 1.96$ ) were high enough to predict wood chemical composition. In Bolaina, model for cellulose attained an excellent predictive power (RMSEP = 1.82, RPD = 6.14) while models for hemicellulose and holocellulose attained a good predictive power (RPD > 2.0). This study showed that FTIR-ATR together with PLS is a reliable method to determine the wood chemical composition in juvenile trees of Capirona and Bolaina.

## Introduction

The Amazonian Forest plays an important role in human being and environmental welfare; however, the growing deforestation could reach a summit point thus this forest will turn to non-forest ecosystem [1, 2]. Just in the last 10 years, agriculture, mining, narcotraffic and

**Competing interests:** The authors have declared that no competing interests exist.

growing population [2] have contributed to the disappearance of 26 million Ha of Amazonian Forest [3], an area close to New Zealand surface. Thus, Amazonian deforestation has increased the intensity of global warming, loss of biodiversity and affects indigenous communities adversely (migration, prostitution, etc) [2]. Deforestation also increases risk of human contact with zoonotic pathogens [4].

In this context, tree breeding is a competitive and environmentally friendly opportunity to recover and increase woody biomass [5]. Tree breeding programs increase wood quality and forestry production; such is the case of *Scots pine*, *Norway spruce*, *Stika spruce*, *Eucalyptus pellita* and *Acacia mangium* [5, 6]. In the amazon forest there is interest in tree breeding programs of Capirona and Bolaina [7]. These fast-growing trees deliver wood products at an early stage [8] and can be coppiced for successive harvests [9]. Both species have a high wood market demand and are widely used for reforestation [10–12]. Furthermore, the production of both species is a sustainable activity for small farmers across Amazon and provides environmental services like carbon capture, conservation of biodiversity and soil fertility [8].

Selection of superior trees (plus-tree) is normally based on visual properties rather than quality parameters such as cellulose, hemicellulose and lignin content [5, 6]. The chemical composition is a key feature that determines wood quality and the end-use of wood in early stage [13, 14]. For instance, lignin is particularly appreciated in timber industry but not in pulp production, while cellulose is quite valued in pulp industry [13]. The information of chemical composition can help breeders in the selection process of superior trees [6]. However, determination of chemical composition using conventional methods is expensive (USD 1000 per tree), time consuming, destructive and requires chemical reagents [6, 15].

On the contrary to conventional techniques, FTIR-ATR spectroscopy is a practical and economical analytic method to determine wood chemical composition [15]. This method has been successfully used to identify cellulose, hemicellulose, lignin, monosaccharide, extractive compounds and proteins in forest species [15–18]. The complex data generated by FTIR-ATR (vibration of chemical bonds and functional groups) is processed by multivariate analysis [19] such as partial least squared (PLS). This multivariate method is one of the most used ones and can predict sample properties by a number of latent variables which compress spectral information [20]. Despite the successful application and numerous advantages of these methods, they are barely applied in native Amazonian trees.

To the best of our knowledge, there is no FTIR-ATR spectroscopy coupled to PLS in fast growing trees in juvenile stages. Therefore, the aim of this work was to develop efficient models to predict cellulose, hemicellulose, holocellulose and lignin content in juvenile trees of Capirona and Bolaina.

## Material and methods

### Plant material

100 Capirona trees and 150 Bolaina trees acquired from Ayahuasca Ecolodge E.I.R.L. and Est. Exp. Agraria Pucallpa–Ucayali–INIA, were kept in a net house in La Molina (12°05' S, 76°57' W and 243.7 masl). When trees reached 1.8 year-old, the samples were collected from apical and basal stem to expand range of wood chemical variability. We grouped 4 and 6 plants as a sample for Capirona and Bolaina, respectively, to reach enough dry matter (5 g) for chemical analysis. In both species, the third and fourth internode from the upper stem were taken as apical samples while the third and fourth internodes from the lower stem as basal samples.

## Wood chemical analysis

The harvested stems were cut into small pieces, dried at 60°C and milled to quantify cellulose, hemicellulose, and lignin content according to Van Soest and Robertson [21]. This method determines neutral detergent fiber (NDF), acid detergent fiber (ADF) and acid detergent lignin (ADL) after digesting samples with chemical reagents. NDF is first obtained after digestion in neutral detergent solution, and it mainly consists of cellulose, hemicellulose and lignin. After digesting NDF with acid detergent, ADF is obtained and contains cellulose and lignin predominantly. To obtain ADL, cellulose was removed from ADF with H<sub>2</sub>SO<sub>4</sub>. Cellulose content is calculated from the difference between ADF and ADL, hemicellulose from the difference between NDF and ADF. Lignin content is expressed as ADL and holocellulose was determined by addition of cellulose with hemicellulose [22]. We analysed 25 samples per apical and basal stem and all experiments were run twice for each sample. Analysis of variance (ANOVA) followed by Tukey comparison test ( $\alpha = 0.05$ ) between the chemical composition of apical and basal stem was performed using R studio.

## FTIR-ATR spectroscopy

Spectra from dried, milled and sieved (60 Mesh) samples of both species were collected. All samples were kept in sealed containers at air-dry moisture content [27]. Triplicated FTIR spectra of each sample were acquired by attenuated total reflection (ATR) in Spectrum 100 Perkin Elmer spectrometer equipped with a diamond ATR accessory, a Tantalato de Litio detector and a KBr beam splitter. Thirty-two scans were run per sample in the spectra range between 4000 cm<sup>-1</sup> and 400 cm<sup>-1</sup> at 4 cm<sup>-1</sup> spectral resolution. Before each measurement, background spectra with the same settings were collected.

## Chemometrics

**Data acquisition and pre-processing.** Spectral data were acquired through Spectrum 10 software in SP format. Prior to pre-processing, spectral data were mean-centered. To remove the effect of interference factors (light scattering, noise, and pathlength variations) from spectral data, the following pre-processing methods (Savitzky-Golay first derivative, Savitzky-Golay second derivative, Smoothing, standard normal variation, multiplicative scatter correction, straight line correction and min-max normalization) were evaluated. Pre-processing was applied in the full spectra region (3700–850 cm<sup>-1</sup>) and the fingerprint region (1800–850 cm<sup>-1</sup>).

**Partial Least Squares (PLS) analysis.** PLS regression was used to build models for predicting cellulose, hemicellulose, holocellulose and lignin content from FTIR spectra. A cross validation was performed in the Peak spectroscopy software (<https://www.essentialftir.com/>) to select the best pre-treatment for each compound in the full spectra and the fingerprint region. Once identified the best pre-treatment for each compound, the data were split into calibration set (100 spectra) and validation set (50 spectra). Models were built using calibration set and evaluated using validation set.

**Determination of number of latent variables.** Two methods were performed to determine the number of latent variables: Haaland and Thomas [23] and Gowen et al. [24]. This last method uses the appearance of noise in regression coefficients as a sign of over-fitting. The noise and sharp peaks in regression coefficients were quantified (Jaggedness) and paired to RMSE to estimate the optimal numbers of latent variables [24].

**Model evaluation.** Performance of models were evaluated using root mean squared error of calibration (RMSEC) and predictions (RMSEP), coefficient of determination of calibration ( $R_c^2$ ) and prediction ( $R_p^2$ ), and ratio of performance to deviation (RPD). To assess the predictive

ability of models (RPD), Santos et al. [25] and Li et al. [26] formula were used. According to Karlinasari et al [27], models with RPD values higher than 1.5 are satisfactory and useful for initial screenings and preliminary predictions, RPD values between 2.0 and 2.5 have very good prediction and RPD values higher than 3 are efficient prediction. In wood samples an RPD of 1.5–2.5 is enough to predict wood chemical composition [28]. Finally, the best model performance was selected based on  $R_c^2$  and  $R_p^2$  values close to 100%, RMSEC and RMSEP values close to zero and a high RPD value [17].

## Results and discussions

### Analysis of chemical composition

The chemical composition of the apical and basal stem for both species (S1 Table) was determined by Van Soest and Robertson [21] method and it is summarized in Table 1. In both species, the percentage of cellulose and hemicellulose was significantly higher ( $p < 0.001$ ) in the basal stem than apical stem. In contrast, a significant decrease in hemicellulose was observed in the basal stem with respect to apical stem in Capirona ( $p < 0.05$ ) and Bolaina ( $p < 0.001$ ). Like carbohydrates, lignin content increased in Capirona basal stem while in Bolaina the lignin content decreased in the apical stem. Therefore, the higher content of carbohydrates and lignin in the base of trees may be related to denser wood as an adaptative response to alleviate any bending stress related to wind and other forces [9, 29]. The mechanical benefit of denser wood is the increasing strength in supporting tissues at the base of tree stalk where cellulose provides strength and lignin, compression and shear resistance [30].

### FTIR spectra

In the FTIR spectra of Capirona and Bolaina wood, two regions of analysis were distinguished (3800–2800  $\text{cm}^{-1}$  and 1800–850  $\text{cm}^{-1}$ ), see Fig 1. The peak numbers and their assignment to chemical compounds are presented in Table 2. All samples show peaks 1, 2 and 3, which were assigned to O-H stretching, symmetric and asymmetric C-H stretching, respectively (Fig 1 and

**Table 1. Descriptive statistics of the chemical composition of *C. spruceanum* and *G. crinita*.**

Specie	<i>Calycophyllum spruceanum</i>				<i>Guazuma crinita</i>			
	% Cel	% Hem	% Hol	% L	% Cel	% Hem	% Hol	% L
<b>Apical stem</b>								
Average	29.81***	10.17*	39.98***	8.14***	20.44***	19.31***	39.75***	12.25***
Minimum	26.24	8.08	35.84	7.05	16.48	11.56	31.54	7.82
Maximum	36.13	12.13	46.19	9.88	25.83	35.27	53.77	15.56
SD	2.47	0.92	2.47	0.64	2.03	5.92	5.48	1.89
<b>Basal stem</b>								
Average	35.97***	9.64*	45.61***	9.65***	42.44***	14.25***	56.69***	11.22***
Minimum	30.91	5.49	38.99	8.14	38.83	12.42	52.04	7.26
Maximum	41.69	13.64	51.49	11.28	48.67	16.53	63.98	14.33
SD	2.94	1.39	3.16	0.67	2.04	0.86	2.37	1.35

Cel: Cellulose, Hem: Hemicellulose, Hol: Holocellulose and L: lignin.

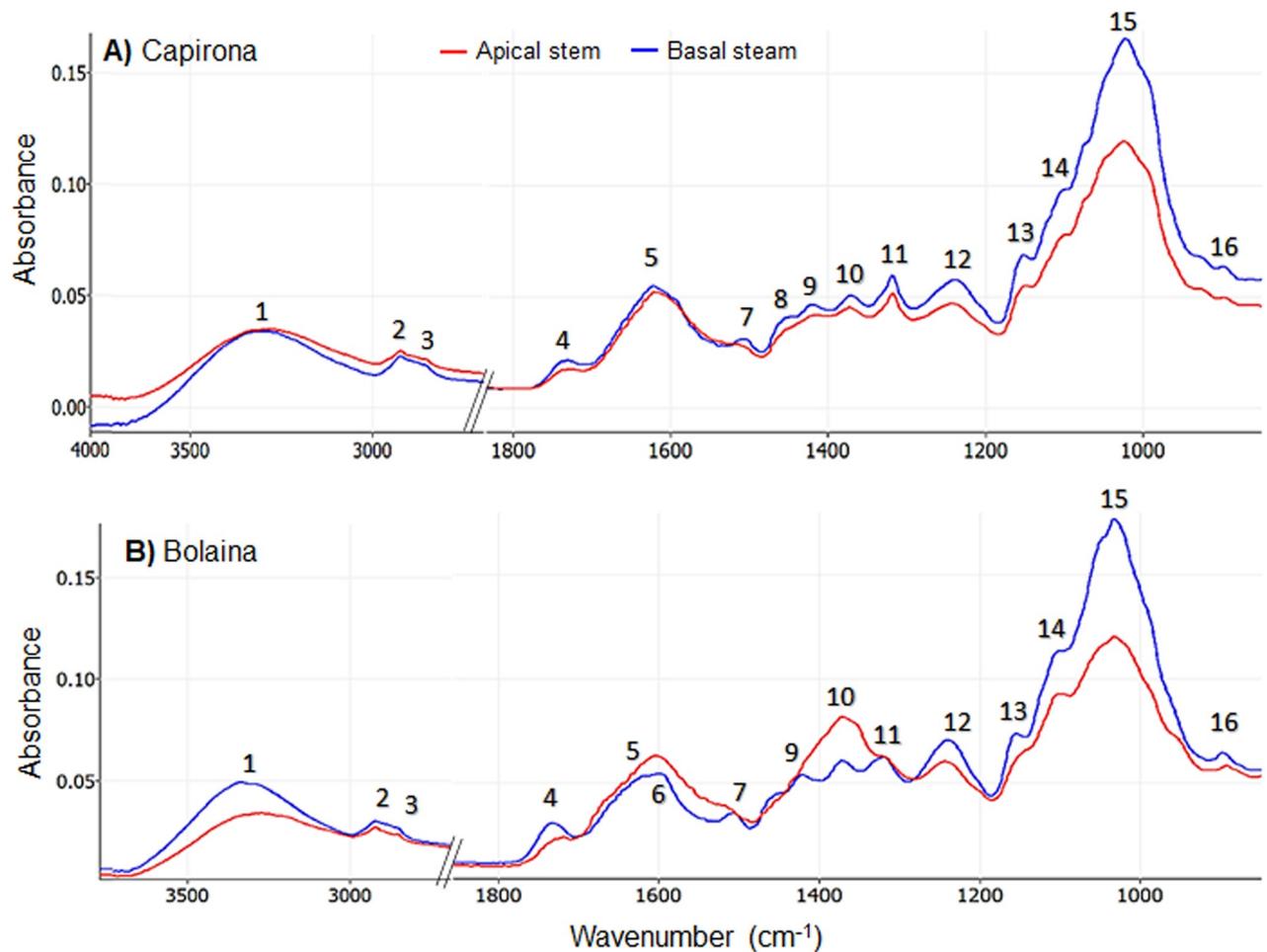
Significance:

\*  $p < .05$ ,

\*\*  $p < .01$  and

\*\*\*  $p < .001$ .

<https://doi.org/10.1371/journal.pone.0256559.t001>



**Fig 1.** A representative average raw ATR-FTIR spectra from juvenile trees of Capirona and Bolaina.

<https://doi.org/10.1371/journal.pone.0256559.g001>

Table 2) and they are presented in all components of wood [31]. In the fingerprint (1800–850  $\text{cm}^{-1}$ ), there were many peaks assigned to cellulose (peaks 11, 13 and 16), hemicellulose (peak 4), lignin (peaks 6 and 7), holocellulose (peaks 10, 12, 14 and 15) and common peaks to carbohydrates and lignin (peaks 8 and 9).

Peak 4 corresponds to vibration bonds of  $\text{C}=\text{O}$  in hemicellulose and extractive compounds such as esterified resin acids, waxes and fats [17, 33] and its absorbance was higher in the basal stem than apical stem in both the species, which contradicts the chemical analyses (Table 1). These differences could be explained because of the presence of extractive compounds which can be masking hemicellulose vibration [37]. In relation to lignin, peaks 6 and 7 indicate the vibration of the aromatic skeletons [17, 33] and there were more intense in the basal stem than apical stem in both the species. However, peak 6 was not well defined in Capirona due to the presence of a prominent peak 5, which is attributed to calcium oxalate and flavones [33]. While in Bolaina, peak 5 is poorly defined in apical stem, which could be a signal of a little presence of extractives [34] and it is close to peak 6. Peaks 5, 6 and 7 are in the region from 1750 to 1510  $\text{cm}^{-1}$ , which was used by Funda et al. [18] to report chemical differences between juvenile and mature wood in Scots pine because of vibrations in lignin and some carbohydrates. Similarly, Ozgenc et al. [38] were able to determine lignin content in alder and cedar

Table 2. Assignment of peaks in FTIR spectra of *C. spruceanum* and *G. crinita*.

Peak	Boliana		Capirona		Assignments	References
	Apical	Basal	Apical	Basal		
1	3266	3340	3281	3315	Lignin, cellulose and hemicellulose: OH intramolecular and intermolecular stretching.	[32]
2	2923	2922	2923	2923	Lignin, cellulose and hemicellulose: symmetric methyl and methylene stretching.	[32]
3	2853	2857	2854	2855	Lignin and hemicellulose: asymmetric methyl and methylene stretching.	[32]
4	1718	1733	1730	1731	Hemicellulose: C = O stretching in unconjugated ketone, carbonyl, and aliphatic groups. Extractives: C = O stretching in ester carbonyl.	[17, 33]
5	-	1622	1621	1623	Flavones and calcium oxalate: C = O stretching.	[33]
6	1603	1598	-	-	Lignin and extractives: aromatic ring vibration.	[17, 33]
7	1518	1508	1516	1511	Lignin and extractives: aromatic skeletal vibrations.	[17, 33]
8	-	-	-	1447	Cellulose, lignin, and hemicellulose: O-H in-plane bending.	[34]
9	-	1422	1417	1419	Lignin: C = O stretching and aromatic skeletal vibration. Carbohydrates: CH and OH bending.	[16–18, 32]
10	1371	1372	1373	1371	Cellulose and hemicellulose: CH bending.	[32, 35]
11	1320	1319	1318	1318	Cellulose: CH <sub>2</sub> wagging.	[32, 35]
12	1243	1240	1244	1239	Polysaccharides: C-O stretch and O-H in plane	[36]
13	-	1155	1150	1152	Cellulose: C–O–C asymmetric stretching	[32]
14	1104	-	-	1099	Cellulose and hemicellulose: C–O–C stretching. Lignin: Aromatic C–H in-plane deformation.	[32, 35]
15	1032	1032	1024	1021	Holocellulose and lignin: C–O stretching.	[33]
16	893	897	896	897	Cellulose: C1–H deformation	[33]

<https://doi.org/10.1371/journal.pone.0256559.t002>

based on skeleton of aromatics rings (1508–1510  $\text{cm}^{-1}$  and 1603–1609  $\text{cm}^{-1}$ ). Moreover, peaks 8, 9, 10 and 11 were assigned to stretching and bending bonds of carbohydrates and lignin [16–18, 32, 34, 35]; the absorbance of these peaks in Capirona was higher at basal than apical stem and it was in agreement with chemical analysis (Table 1). For Bolaina, the absorbance of peaks 9, 10 and 11 (Fig 1B) was higher in the apical section versus basal stem but peak 8 was not distinguished. The spectral region (1510–1265  $\text{cm}^{-1}$ ) for peaks 8, 9, 10 and 11 was also used in *Pinus koraiensis* to detect high lignin and carbohydrates concentrations in artificial inclined stem [39]. In the carbohydrate region (1200 to 890  $\text{cm}^{-1}$ ), peaks 12, 13, 14, 15 and 16 (Fig 1A and 1B) evidenced a higher concentration of carbohydrates in the basal stem than in the apical stem in both the species. Furthermore, within the fingerprint region, Acquah et al. [36] were able to report polysaccharides by the presence of peaks: 897  $\text{cm}^{-1}$  (C-H deformation in cellulose and C-O stretch in polysaccharides), 1157  $\text{cm}^{-1}$  (C-O-C vibrations) and 1239  $\text{cm}^{-1}$  (C-O stretch and O-H stretch in plane in polysaccharides) in forest logging residue. Thus, the FTIR spectra evidenced the compositional difference in different forest samples.

## PLS modeling

PLS regression was applied to develop predictive models that correlate FTIR spectra with chemical wood composition. First, cross validation was performed to select the best pre-treatment for each compound in the full spectra and fingerprint region (S2 Table). The number of latent variables (LV) were determined by Haaland & Thomas [23] method, however higher-order latent variables were observed (S2 Table). Therefore, Gowen et al. [24, 40] method was used to prevent overfitting. Then, external validation was performed using the best predictive model obtained by cross validation (Table 3). RMSEC, RMSEP,  $R_c^2$ ,  $R_p^2$  and RPD were used to evaluate the model performance.

**Table 3. Performance evaluation of PLS models to predict wood chemical composition.**

Species	Component	PP	LV	$R_c^2$	RMSEC	$R_p^2$	RMSEP	RPD <sub>a</sub>	RPD <sub>b</sub>
Capirona	Cellulose	2 <sup>nd</sup> Der	4	85.1	1.57	74.0	2.06	1.96	1.96
	Hemicellulose	1 <sup>st</sup> Der	5	77.3	0.52	61.1	0.68	1.60	1.60
	Holocellulose	2 <sup>nd</sup> Der	4	83.0	1.62	72.1	2.07	1.89	1.90
	Lignin	2 <sup>nd</sup> Der	4	90.2	0.31	75.4	0.48	2.02	2.06
Bolaina	Cellulose	MM-N	4	97.3	1.83	97.3	1.82	6.14	6.14
	Hemicellulose	1 <sup>st</sup> Der	5	85.1	1.98	81.3	2.12	2.31	2.32
	Holocellulose	V-N	6	89.3	3.09	88.8	3.16	2.99	2.99
	Lignin	1 <sup>st</sup> Der	6	82.6	0.72	77.7	0.81	2.12	2.12

PP: Pre-processing methods: (1<sup>st</sup> Der: First derivative, 2<sup>nd</sup> Der: Second derivative. MM-N: Min—Max normalization and V-N: Vector normalization).

LV: Latent variables according to Gowen et al. [21].

<sup>a</sup> RPD according to Santos et al. [25].

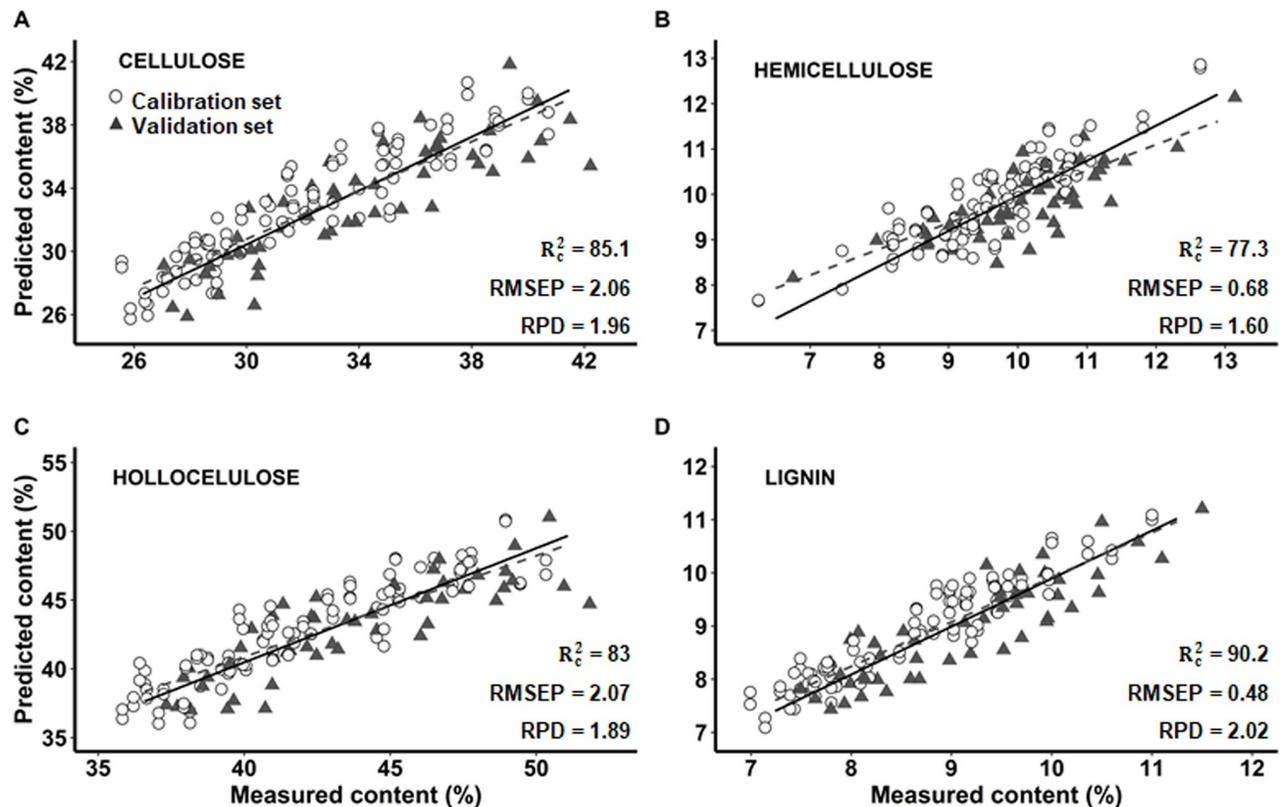
<sup>b</sup> RPD according to Li et al. [26].

<https://doi.org/10.1371/journal.pone.0256559.t003>

The lignin models of Capirona and Bolaina (Table 3) were highly accurate because of low RMSEP values (0.48 and 0.81, respectively), with a good prediction power (RPD > 2.0) and a good data fitting ( $R_c^2$  of 90.2% and 82.6%). Our results are in agreement with Acquah et al. [36], Zhou et al. [17], Zhou et al. [37], who reported values of  $R^2 \approx 90\text{--}80\%$  and RMSEP  $\approx 0.47\text{--}0.80$ . Despite these authors reported models with lower RPD (1.4 and 1.72) values than our models' (see Table 3), the author models were considered to have a good predictive power.

In Capirona, the predictive power of cellulose (Fig 2A), hemicellulose (Fig 2B) and holocellulose (Fig 2C) models were satisfactory to quantify these compounds. The RPD values of these three compounds were within 1.5 to 2.0 range (Table 3), they are suitable for initial screenings and preliminary predictions [26, 27]. However, these models would be proper for estimating wood properties according to Hein et al. [28]. Moreover, the  $R_c^2$  values for cellulose and holocellulose content were 85.1% and 83.0% with a RMSEP of 2.06 and 2.07, respectively. Meanwhile, the  $R_c^2$  value for hemicellulose was 77.3% with a RMSEP of 0.68. A good hemicellulose predictive model with similar values to our model was reported by Funda et al. [18] in juvenile wood of Scots pine. Furthermore, our cellulose model is consistent with Zhou et al. [17], who reported  $R^2$  and RPD values of 85.62% and 1.72 in hardwoods so this model was strong enough for screening and cellulose characterization. Finally, our holocellulose model was better than the one reported by Acquah et al. [36] in forest logging residue. On the other hand, Funda et al. [18], Zhou et al. [17] and Acquah et al. [36] reported more efficient lignin models based on a high RPD and a low RMSEP than carbohydrate models. We found this trend in Capirona models too. A possible explanation to this trend may be the presence of unique chemical structures in lignin unlike carbohydrates [36].

In Bolaina, the RMSEP values of cellulose, hemicellulose and holocellulose (1.82, 2.12 and 3.16, respectively) were higher compared to that of the other authors. For instance, Funda et al. [18], who reported RMSEP values from 0.69 to 0.71 in cellulose and 0.71 in hemicellulose, while Zhou et al. [17] found RMSEP values from 0.8 to 1.9 in cellulose and hemicellulose. Since  $R^2$  value depends on the distribution of the data and not only on the model performance [41], it is possible to use it as a secondary measure of model performance [18]. Thus, the performance models should achieve RMSEC and RMSEP values close to zero, a high RPD with  $R_c^2$  and  $R_p^2$  values close to 100%. In Table 3, models showed a high predictive power (RPD > 2)



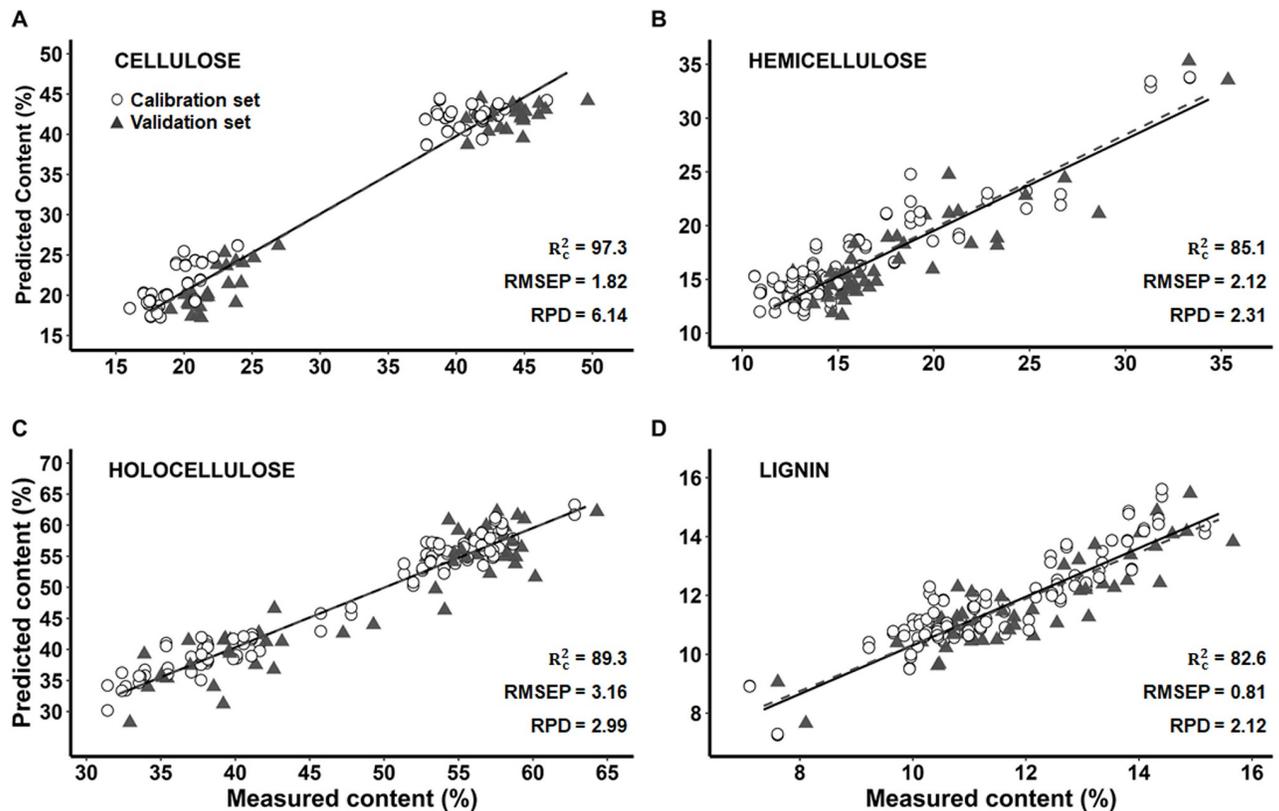
**Fig 2. Prediction of cellulose, hemicellulose, holocellulose and lignin content by multivariate models with FT-IR (PLS) in Capirona (*C. spruceanum*).** A: Cellulose, B: Hemicellulose, C: Holocellulose and D: Lignin.  $R_c^2$ : coefficient of determination of calibration. RMSEP: root mean squared error of predictions. RPD: ratio of performance to deviation. Solid line: regression line of calibration set. Dashed line: regression line of validation set.

<https://doi.org/10.1371/journal.pone.0256559.g002>

with RMSEP slightly high, so  $R^2$  was considered to evaluate model performance. Thus hemicellulose (Fig 3B) and holocellulose (Fig 3C) models had a high data correlation ( $R_c^2$  of 85.1 and 89.3, respectively) and  $RPD > 2.0$ , so they were considered as very good prediction models [26, 27]. Cellulose model (Fig 3A) showed exceptionally high values of  $R_c^2 = 97.3$  and  $RPD = 6.14$ , so it was classified as an excellent predictive model [26, 27]. In contrast, the cellulose, hemicellulose and holocellulose models reported by Acquah et al. [36] in forest logging residue and Zhou et al. [17] in hardwoods were considered satisfactory.

On the other hand, in Bolaina lignin model (Fig 3D) had a higher accuracy ( $RMSEP = 0.81$ ) than carbohydrates models ( $RMSEP$  from 1.82 to 3.16), which agrees with *Pinus sylvestris* [18], forest logging residue [36] and hardwoods [17] models. However, the lignin model presented lower predictive power ( $RPD = 2.12$ ) than the carbohydrates models ( $RPD$  from 2.32 to 6.14). In NIR, Karlinasari et al. [27] and Jiang et al. [22] reported this trend in *Acacia mangium* and *Pinus spp.*, respectively.

For Capirona and Bolaina, the accuracy of the models ( $RMSEP$ ) was higher in lignin than the carbohydrate models (Table 2). According to Hein et al. [25], an  $RPD > 1.5$  allows to quantify the chemical composition of the wood. However, the  $RPD$  scale established by Karlinasari et al. [27] allows a better classification of the models. Finally for each model, the values of  $R^2$  and  $RMSE$  were close between the set of calibration and validation, therefore models have good predictive capacity according to Li et al. [26].



**Fig 3. Prediction of cellulose, hemicellulose, holocellulose and lignin content by multivariate models with FT-IR (PLS) in Bolaina (*G. crinita*).** A: Cellulose, B: Hemicellulose, C: Holocellulose and D: Lignin.  $R_c^2$ : coefficient of determination of calibration. RMSEP: root mean squared error of predictions. RPD: ratio of performance to deviation.

<https://doi.org/10.1371/journal.pone.0256559.g003>

## Conclusions

A rapid method for wood composition determination in an integrated production system based on Capirona (international market) and Bolaina (national market) is required for genetic improvement and hence improve a forest-based bioeconomy in Amazon. This study confirmed significant differences in cellulose, hemicellulose, holocellulose and lignin content between apical and basal stem in Capirona and Bolaina. The variability found was used to build PLS models to estimate cellulose, hemicellulose, holocellulose and lignin content in juvenile wood from FTIR-ATR spectra. Thus, lignin models of both species achieved high accuracy (low RMSEP) and were considered as very good prediction models ( $RPD > 2$ ). In Bolaina, models of hemicellulose and holocellulose showed a very good prediction power ( $RPD > 2.0$ ), meanwhile cellulose model was excellent ( $RPD > 3.0$ ). In Capirona, cellulose, hemicellulose and holocellulose models attained sufficient predictive power ( $RPD > 1.5$ ) to estimate wood composition. Thus FTIR-ATR spectroscopy combined with PLS models offers a technologic tool that can be applied to early tree selection based on chemical wood composition.

## Supporting information

**S1 Table. Wood chemical composition.**  
(XLSX)

**S2 Table. Leave-one-out cross validation method and number of latent variables determination.**

(XLSX)

**Acknowledgments**

The authors would also like to thank Vladimir Camel for his critical reading.

**Author Contributions**

**Conceptualization:** Gaston Zolla.

**Data curation:** Rosario Javier-Astete, Gaston Zolla.

**Formal analysis:** Rosario Javier-Astete, Gaston Zolla.

**Funding acquisition:** Jorge Jimenez-Davalos, Gaston Zolla.

**Investigation:** Rosario Javier-Astete, Gaston Zolla.

**Methodology:** Rosario Javier-Astete, Jorge Jimenez-Davalos, Gaston Zolla.

**Project administration:** Gaston Zolla.

**Resources:** Jorge Jimenez-Davalos, Gaston Zolla.

**Software:** Gaston Zolla.

**Supervision:** Gaston Zolla.

**Validation:** Rosario Javier-Astete, Gaston Zolla.

**Visualization:** Rosario Javier-Astete, Gaston Zolla.

**Writing – original draft:** Rosario Javier-Astete, Gaston Zolla.

**Writing – review & editing:** Rosario Javier-Astete, Gaston Zolla.

**References**

1. Lovejoy TE, Nobre C. Amazon Tipping Point. *Science Advances*. 2018 feb; 4(2):1. <https://doi.org/10.1126/sciadv.aat2340> PMID: 29492460
2. Ellwanger JH, Kulmann-LEAL B, Kaminski VL, Valverde-Villegas JM, Veiga ABG DA, Spilki FR. Beyond diversity loss and climate change: Impacts of Amazon deforestation on infectious diseases and public health. *Anais da Academia Brasileira de Ciências*. 2020; 92(1):1–33. <https://doi.org/10.1590/0001-3765202020191375> PMID: 32321030
3. FAO, UNE. The State of the World's Forests 2020. Forests, biodiversity and people. 2020:139 p.
4. Gibb R, Redding DW, Chin KQ, Donnelly CA, Blackburn TM, Newbold T, et al. Zoonotic host diversity increases in human-dominated ecosystems. *Nature*. 2020; 584(7821):398–402. <https://doi.org/10.1038/s41586-020-2562-8> PMID: 32759999
5. Haapanen M, Jansson G, Nielsen UB, Steffenrem A, Stener L-G. The status of tree breeding and its potential for improving biomass production: A review of breeding activities and genetic gains in Scandinavia and Finland. *SkogForsk*, 2015. 53 p.
6. Meder R The Magnitude of Tree Breeding and the Role of near Infrared Spectroscopy. *NIR news*. 2015 May 1; 26(3):8–10.
7. Cornelius JP, Pinedo-Ramírez R, Montes CS, Ugarte-Guerra LJ, Weber JC. Efficiency of early selection in (*Calycophyllum spruceanum*) and (*Guazuma crinita*), two fast-growing timber species of the peruvian amazon. *Canadian Journal of Forest Research*. 2018; 48(5):517–523.
8. Sears R, Cronkleton P, Arco MP, Robiglio V, Putzel L, Cornelius JP. Producción de madera en sistemas agroforestales de pequeños productores. Programa de Investigación de CGIAR sobre Bosques, Arboles y Agroforestería (CRP-FTA). 2014:1–8.

9. Weber JC, Sotelo Montes C. Geographic variation in tree growth and wood density of (*Guazuma crinita*) Mart. in the Peruvian Amazon. *New Forest*. 2008; 36(1):29–52.
10. Hecht SB, Morrison KD, Padoch C. The Social Lives of Forests. *Journal of Physics A: Mathematical and Theoretical*. 2014; 387 p.
11. Ushiñahua D. Comportamiento fenológico preliminar de Capirona en la provincia de San Martín, región de San Martín. *HOJA DIVULGATIVA N° 002—Instituto Nacional de Innovación Agraria*. 2016; 1–2.
12. IIAP Evaluación económica de parcelas de regeneración natural y plantaciones de Bolaina Blanca, (*Guazuma crinita*), en el VRAEM—AYNA San Francisco. *Avances Económicos N° 11*. 2017 mar; 12:49–58.
13. Pereira H, Graca J, Rodrigues JC. Wood Chemistry in Relation to Quality. *ChemInform*. 2004 nov 16; 35(46):225p.
14. Fromm J. Cellular Aspects of Wood Formation. Berlin, Heidelberg: Springer Berlin Heidelberg. Editors: Fromm J. 2013:260p.
15. Alonso-Simón A, García-Angulo P, Mérida H, Encina A, Alvarez JM, Acebes JL. The use of FTIR spectroscopy to monitor modifications in plant cell wall architecture caused by cellulose biosynthesis inhibitors. *Plant Signal Behavior*. 2011 Aug 31; 6(8):1104–1110. <https://doi.org/10.4161/psb.6.8.15793> PMID: 21791979
16. Chen H, Ferrari C, Angiuli M, Yao J, Raspi C, Bramanti E. Qualitative and quantitative analysis of wood samples by Fourier transform infrared spectroscopy and multivariate analysis. *Carbohydr Polym*. 2010; 82(3):772–778.
17. Zhou C, Jiang W, Cheng Q, Via BK. Multivariate Calibration and Model Integrity for Wood Chemistry Using Fourier Transform Infrared Spectroscopy. *Journal of Analytical Methods in Chemistry*. 2015; 2015:1–9. <https://doi.org/10.1155/2015/429846> PMID: 26576321
18. Funda T, Fundova I, Gorzsás A, Fries A, Wu HX Predicting the chemical composition of juvenile and mature woods in Scots pine (*Pinus sylvestris* L.) using FTIR spectroscopy. *Wood Science Technology*. 2020 mar 8; 54(2):289–311.
19. Kelly JG, Trevisan J, Scott AD, Carmichael PL, Pollock HM, Martin-Hirsch PL, et al. Biospectroscopy to metabolically profile biomolecular structure: a multistage approach linking computational analysis with biomarkers. *Journal of Proteome Research*. 2011 Apr; 10(4):1437–1448. <https://doi.org/10.1021/pr101067u> PMID: 21210632
20. Héberger K. Chemoinformatics—multivariate mathematical—statistical methods for data evaluation. In: *Medical Applications of Mass Spectrometry*. Budapest: Elsevier; 2008. 141–169.
21. Van Soest P, Robertson J. Systems of Analysis for Evaluating Fibrous Feeds. *Standardization of Analytical Methodology for Feeds*. 1979; 4:49–60.
22. Jiang W, Han G, Via BK, Tu M, Liu W, Fasina O. Rapid assessment of coniferous biomass lignin-carbohydrates with near-infrared spectroscopy. *Wood Science Technology*. 2014; 48(1):109–122.
23. Haaland DM, Thomas EV. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Analytical Chemistry*. 1988 Jun 1; 60(11):1193–1202.
24. Gowen AA, Downey G, Esquerre C, O'Donnell CP. Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients. *Journal of Chemometrics*. 2011 jul; 25(7):375–381.
25. Santos D, Lima K, Marco P, Valderrama P. Vitamin C Determination by Ultraviolet Spectroscopy and Multiproduct Calibration. *TrAC Trends in Analytical Chemistry*. 2016 jan 12; 27(10):1912–1917.
26. Li X, Sun C, Zhou B, He Y. Determination of Hemicellulose, Cellulose and Lignin in Moso Bamboo by Near Infrared Spectroscopy. *Scientific Reports*. 2015 dec 25; 5(1):1–17. <https://doi.org/10.1038/srep17210> PMID: 26601657
27. Karlinasari L, Sabed M, Wistara INJ, Purwanto YA. Near infrared (NIR) spectroscopy for estimating the chemical composition of (*Acacia mangium*) Willd.) wood. *Front. Journal of the Indian Academy of Wood Science*. 2014 Dec 19; 11(2):162–167.
28. Hein PRG. Multivariate regression methods for estimating basic density in Eucalyptus wood from near infrared spectroscopic data. *Cerne, Lavras*. 2010 nov; 16:90–96.
29. Weber JC, Montes CS. Variation and correlations among stem growth and wood traits of (*Calycophyllum spruceanum*) Benth. from the Peruvian Amazon. *Silvae Genetica*. 2005 Dec 1; 54(1–6):31–41.
30. Voelker SL, Lachenbruch B, Meinzer FC, Strauss SH. Reduced wood stiffness and strength, and altered stem form, in young antisense 4CL transgenic poplars with reduced lignin contents. *New Phytologist*. 2011 Mar 15; 189(4):1096–1109.

31. Popescu C-M, Singurel G, Popescu M-C, Vasile C, Argyropoulos DS, Willför S. Vibrational spectroscopy and X-ray diffraction methods to establish the differences between hardwood and softwood. *Carbohydrate Polymers*. 2009 Jul 19; 77(4):851–857.
32. Popescu M, Popescu M-C, Singurel G, Vasile C, Argyropoulos DS, Willför S. Spectral Characterization of Eucalyptus Wood. *Applied Spectroscopy*. 2007 Nov; 61(11):1168–1177. <https://doi.org/10.1366/000370207782597076> PMID: 18028695
33. Zhang F-D, Xu C-H, Li M-Y, Chen X-D, Zhou Q, Huang A-M. Identification of (*Dalbergia cochinchinensis*) (CITES Appendix II) from other three (*Dalbergia*) species using FT-IR and 2D correlation IR spectroscopy. *Wood Science and Technology*. 2016 Jul 5; 50(4):693–704.
34. Xu F, Yu J, Tesso T, Dowell F, Wang D. Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review. *Applied Energy*. 2013 Apr; 104:801–809.
35. Traoré M, Kaal J, Martínez Cortizas A. Differentiation between pine woods according to species and growing location using FTIR-ATR. *Wood Science and Technology*. 2018 Mar 4; 52(2):487–504. <https://doi.org/10.1007/s00226-017-0967-9> PMID: 29497215
36. Acquah GE, Via BK, Fasina OO, Eckhardt LG. Rapid Quantitative Analysis of Forest Biomass Using Fourier Transform Infrared Spectroscopy and Partial Least Squares Regression. *Journal of Analytical Methods in Chemistry*. 2016; 2016:1–10. <https://doi.org/10.1155/2016/1839598> PMID: 28003929
37. Zhou G, Taylor G, Polle A. FTIR-ATR-based prediction and modelling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars. *Plant Methods*. 2011; 7(9):1–10. <https://doi.org/10.1186/1746-4811-7-9> PMID: 21477346
38. Ozgenc O, Durmaz S, Kustas S. Chemical analysis of tree barks using ATR-FTIR spectroscopy and conventional techniques. *BioResources*. 2017; 12(4):9143–9151.
39. Shi J, Xing D, Lia J. FTIR Studies of the Changes in Wood Chemistry from Wood Forming Tissue under Inclined Treatment. *Energy Procedia*. 2012; 16:758–762.
40. Mukherjee S, Martinez-Gonzalez JA, Gowen AA. Feasibility of attenuated total reflection-fourier transform infrared (ATR-FTIR) chemical imaging and partial least squares regression (PLSR) to predict protein adhesion on polymeric surfaces. *The Analyst*. 2019; 144(5):1535–1545. <https://doi.org/10.1039/c8an01768a> PMID: 30542682
41. Kjeldahl K, Bro R. Some common misunderstandings in chemometrics. *Journal of Chemometrics*. 2010 Jul; 24(7–8):558–564.