

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

FACULTAD DE CIENCIAS



**"VALIDACIÓN DE LA METODOLOGÍA PARA LA DETERMINACIÓN
DE LIGNINA CELULOSA Y HEMICELULOSA POR
ESPECTROSCOPIA INFRARROJA EN ÁRBOLES"**

Presentada por:

ROSARIO ELYZABETH JAVIER ASTETE

TESIS PARA OPTAR EL TÍTULO DE:

BIÓLOGA

Lima – Perú

2021

**La UNALM es titular de los derechos patrimoniales de la presente investigación
(Art. 24 Reglamento de Propiedad Intelectual)**

UNIVERSIDAD NACIONAL AGRARIA

LA MOLINA

FACULTAD DE CIENCIAS

**"VALIDACIÓN DE LA METODOLOGÍA PARA LA
DETERMINACIÓN DE LIGNINA CELULOSA Y HEMICELULOSA
POR ESPECTROSCOPIA INFRARROJA EN ÁRBOLES"**

Presentada por:

ROSARIO ELYZABETH JAVIER ASTETE

TESIS PARA OPTAR EL TÍTULO DE:

BIÓLOGA

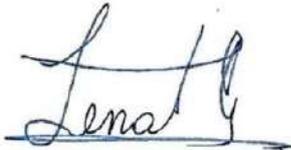
Sustentado y aprobado ante el siguiente Jurado



Mg. Sc. Epifanio Teófilo Chire Murillo
PRESIDENTE



Blga. Milagros Del Rosario Chang La Rosa
MIEMBRO



Mg. Sc. Lena Asunción Tellez Monzón
MIEMBRO



Dr. Jorge Eduardo Jimenez Dávalos
ASESOR



Dr. Gastón Enrique Martín Zolla Benites
Co Asesor

DEDICATORIA

A mis abuelitos Carmen Córdova,
Domitila León y Máximo Javier

AGRADECIMIENTO

- Al Vicerrectorado de Investigación de la UNALM por el financiamiento a través de “9° Concurso para subvención de tesis de pregrado UNALM 2019”.
- Al proyecto 159-2018-FONDECYT-BM-IADT-AV por el financiamiento.
- A la compañía Essential FTIR por brindarnos una licencia extra del software Peak Spectroscopy en estado de emergencia.
- Al Ph.D. in Chemistry Efstathiou Constantinos de la National and Kapodistrian University of Athens y al Ph.D in Chemometrics Åsmund Rinnan, de la University of Copenhagen por el apoyo en el análisis PLS
- Al Laboratorio de evaluación nutricional de alimentos LENA, en especial a Deivis Villalobos por su contribución en el análisis químico.
- A todos mis compañeros del laboratorio por su apoyo y contribución a enriquecer mi conocimiento y en especial a Vladimir Camel.
- A mi familia y en especial a mis abuelos por su apoyo incondicional.
- A mis amigos y en especial a Jessenia Melo por compartir la pasión por la química.

ÍNDICE GENERAL

RESUMEN	viii
ABSTRACT	ix
I. INTRODUCCIÓN	1
II. REVISION DE LITERATURA.....	3
2.1. SECTOR FORESTAL PERUANO	3
2.2. CAPIRONA	4
2.2.1. Descripción botánica y taxonomía	5
2.2.2. Distribución y ecología de capirona	5
2.2.3. Madera de capirona	6
2.3. BOLAINA	7
2.3.1. Descripción botánica y taxonomía	7
2.3.2. Distribución geográfica y ecología.....	8
2.3.3. Madera de bolaina	9
2.4. XILOGENESIS	10
2.5. COMPONENTES QUÍMICOS DE LA MADERA	11
2.5.1. Celulosa	11
2.5.2. Hemicelulosa	15
2.5.3. Lignina.....	18
2.6. ESPECTROSCOPIA INFRARROJA.....	26
2.6.1. Espectro electromagnético.....	26
2.6.1. Principio de espectroscopia infrarroja media	29
2.6.2. Ley de Beer.....	31
2.6.3. Movimientos vibracionales.....	31
2.6.4. Frecuencias de grupo	33
2.6.5. Huella dactilar.....	36
2.6.6. Espectogramas de polimeros de madera.....	36
2.6.7. Espectroscopía infrarroja por transformada de Fourier (FTIR).....	39
2.7. QUIMIOMETRÍA	43
2.7.1. Análisis quimiométrico en espectroscopía infrarroja	45
2.7.2. Pre-tratamiento de datos	47
2.7.3. Regresión de mínimos cuadrados parciales (PLS)	54
2.7.4. Evaluación del modelo de calibración.....	56

2.7.5.	Número de variables latentes en pls	58
III.	METODOLOGÍA	62
3.1.	MATERIALES	62
3.1.1.	Reactivos	62
3.1.2.	Equipos	62
3.1.3.	Material de vidrio y de invernadero	62
3.1.4.	Material vegetal	62
3.1.5.	Softwares	63
3.2.	PREPARACIÓN DE LAS MUESTRAS VEGETALES	63
3.2.1.	Capirona	63
3.2.2.	Bolaina.....	63
3.3.	ANÁLISIS QUÍMICO.....	63
3.3.1.	Determinación de fibra detergente neutra (FDN).....	65
3.3.2.	Determinación de fibra detergente neutra (fda).....	66
3.3.3.	Determinación de lignina.....	66
3.4.	COLECTA DE ESPECTRO FTIR	67
3.5.	PROCESAMIENTO DE DATOS Y ANÁLISIS	67
3.5.1.	Análisis estadístico de celulosa, hemicelulosa, holocelulosa y lignina.	67
3.5.2.	Análisis quimiométrico	67
IV.	RESULTADOS Y DISCUSIÓN	70
4.1.	COMPOSICIÓN QUÍMICA DE LA MADERA DE CAPIRONA.....	70
4.2.	COMPOSICIÓN QUÍMICA DE LA MADERA DE BOLAINA	72
4.3.	ASIGNANCIÓN DE BANDAS EN FTIR	74
4.4.	CALIBRACIÓN DEL MODELO DE PREDICCIÓN PLS EN CAPIRONA	78
4.5.	CALIBRACIÓN DEL MODELO DE PREDICCIÓN PLS EN BOLAINA.....	83
V.	CONCLUSIONES	87
VI.	RECOMENDACIONES	88
VII.	BIBLIOGRAFÍA	89
VIII.	ANEXOS	100

ÍNDICE DE TABLAS

Tabla 1: Clasificación taxonómica de Capirona.....	5
Tabla 2: Clasificación taxonómica de Bolaina	8
Tabla 3: Frecuencias calculadas de moléculas diatómicas	34
Tabla 4: Frecuencias de grupos funcionales de compuestos orgánicos.....	35
Tabla 5: Picos FTIR de madera y los enlaces correspondientes.....	38
Tabla 6: Composición de tallo de Capirona (<i>Calycophyllum spruceanum</i>).....	70
Tabla 7: Composición de tallo de Bolaina (<i>Guazuma crinita</i>).....	73
Tabla 8: Cuadro de frecuencias de Capirona y Bolaina	76
Tabla 9: Modelos PLS de celulosa, hemicelulosa, holocelulosa y lignina en Capirona	79
Tabla 10: Mejores modelos PLS de celulosa, hemicelulosa, holocelulosa y lignina de Capirona.....	80
Tabla 11: Modelos PLS de celulosa, hemicelulosa, holocelulosa y lignina.....	84
Tabla 12: Mejores modelos PLS de celulosa, hemicelulosa, holocelulosa y lignina en Bolaina	85

ÍNDICE DE FIGURAS

Figura 1: Plantones de Capirona (<i>Callycophillum spruceanum</i>).....	4
Figura 2: Plantones de Bolaina (<i>Guazuma crinita</i>)	7
Figura 3: Estructura molecular de la celulosa	12
Figura 4: Complejo terminal o Roseta.....	13
Figura 5: Complejo celulosa sintasa en la membrana celular	14
Figura 6: Estructura química de las hemicelulosas	17
Figura 7: Monolignoles, unidades monoméricas de lignina.....	20
Figura 8: Unidades estructurales del polímero lignina.....	21
Figura 9. Síntesis de monolignoles de Lignina	23
Figura 10: Transporte de monolignoles y polimerización de Lignina.....	25
Figura 11: Espectro electromagnético y región infrarroja.....	26
Figura 12: Naturaleza ondulatoria de la radiación electromagnética	27
Figura 13: Diagrama de niveles de energía	29
Figura 14: Región de frecuencias de grupo y huella dactilar	30
Figura 15: Tipos de vibraciones moleculares	32
Figura 16: Modos de vibración de una molécula de agua	35
Figura 17: Estructura del Interferómetro de Michelson	41
Figura 18: Esquema de un Interferograma	42
Figura 19: Diagrama del muestreo por ATR.....	43
Figura 20: Matriz de datos y su reducción a variables latentes por quimiometría	44
Figura 21: Flujograma de los procesos de análisis quimiométrico	46
Figura 22: Desplazamiento y curvatura de la línea base	48
Figura 23: Espectros procesados mediante: A. Línea base. B. MCS. C. SNV. D. Normalización.....	51
Figura 24: Suavizado de datos según Norris – Williams (A) y Savinsky – Golay (B)	52
Figura 25: Estructura del modelo PLS	54
Figura 26: Matrices de datos PLS.....	55
Figura 27: Gráfico de BM + RVM en función del número de factores.....	61
Figura 28: Análisis secuencial de fibra basado en la metodología de Van Soest y Robertson.....	65
Figura 29: Flujograma para la construcción de modelos PLS.....	68

Figura 30: Composición química en la sección apical y basal de Capirona.....	71
Figura 31: Composición química en la sección apical y basal de Bolaina.....	74
Figura 32: Espectro FTIR de Capirona y Bolaina	75
Figura 33: Modelos PLS de celulosa, hemicelulosa, holocelulosa y lignina en Capirona ..	82

ÍNDICE DE ANEXOS

Anexo 1: Contenido de celulosa, hemicelulosa, holocelulosa y lignina en Capirona.....	101
Anexo 2: Contenido de celulosa, hemicelulosa, holocelulosa y lignina en Bolaina	102
Anexo 3: Histograma de frecuencia de Capirona.....	103
Anexo 4: Prueba de normalidad de los componentes de Capirona	104
Anexo 5: Modelo lineal mixto en función de la sección apical y basal en tallos de Capirona.....	105
Anexo 6: Homogeneidad de varianza de los modelos generalizados mixtos en compuestos de Capirona	107
Anexo 7: Prueba de Tukey en los componentes de Capirona.	108
Anexo 8: Histograma de frecuencias en Bolaina.....	110
Anexo 9: Prueba de normalidad de los principales componentes de Bolaina.	111
Anexo 10: Modelo lineal mixto en función de la sección apical y basal en Bolaina	112
Anexo 11: Homogeneidad de varianza de los modelos mixtos lineales generalizados en Bolaina	114
Anexo 12: Contraste de medias en la sección basal y apical por teste de Tukey en Bolaina.....	115
Anexo 13: Dos regiones y dos pretratamientos de Capirona.....	117
Anexo 14: Dos regiones y los pretratamientos de Bolaina.....	118
Anexo 15: Capirona: determinación del número de factores en modelo PLS de celulosa con 2ºderivada.....	119
Anexo 16: Capirona: determinación del número de factores en modelo PLS de hemicelulosa con 1ºderivada.....	120
Anexo 17: Capirona: determinación del número de factores en modelo PLS de holocelulosa con 2ºderivada.....	121
Anexo 18: Capirona: determinación del número de factores en modelo PLS de lignina con 2ºderivada.....	122
Anexo 19: Bolaina: determinación del número de factores en modelo PLS de celulosa con mínimo máximo de normalización.....	123
Anexo 20: Bolaina: determinación del número de factores en modelo PLS de hemicelulosa con 1ºderivada.....	124

Anexo 21: Bolaina: determinación del número de factores en modelo PLS de holocelulosa con vector de normalización	125
Anexo 22: Bolaina: determinación del número de factores en modelo PLS de lignina con 1° derivada.....	126
Anexo 23: Registro fotográfico	127

RESUMEN

Las especies forestales amazónicas, Capirona (*Callycophilum spruceanum*) y Bolaina (*Guazuma crinita*) tienen una gran importancia económica por su crecimiento rápido y diversos usos de su madera. El desarrollo de plantaciones forestales eficientes y competitivas mediante programas de mejoramiento genético requiere herramientas tecnológicas aplicadas a la selección de material vegetal basadas en propiedades como la composición química de la madera. El objetivo de este trabajo fue identificar las diferencias químicas entre el tallo apical y basal asimismo desarrollar modelos cuantitativos para predecir la composición química de ambas especies a partir de espectros infrarrojos con transformada de Fourier (FTIR). Los espectros FTIR fueron obtenidos a partir de 150 muestras de cada especie de entre 1.8 años. Nuestros resultados mostraron diferencias significativas entre la sección apical y basal de cada especie. Esta variabilidad permitió la construcción de modelos PLS (mínimos cuadrados parciales) a partir de los espectros FTIR-ATR y se evaluaron con la raíz cuadrada de error medio de la predicción (RMSEP) y la relación de rendimiento y desviación (RPD). En Capirona, el modelo de lignina presentó buena capacidad predictiva por tener valores de $RMSEP = 0.48$ y $RPD > 2.0$. Por otro lado, los modelos de celulosa, hemicelulosa y holocelulosa presentaron valores de RMSEP de 2.06, 0.68 y 2.07 y valores de RPD de 1.96, 1.60 y 1.90, los cuales fueron suficientes para la predicción de propiedades químicas de la madera (> 1.5). En el caso de Bolaina, el modelo de celulosa tiene excelente capacidad predictiva ($RPD = 6.14$) y el modelo de holocelulosa fue eficiente ($RPD > 3.0$). Asimismo, los modelos de lignina y hemicelulosa presentaron buena capacidad predictiva ($RPD > 2.0$). Se concluye que, en ambas especies, la composición química entre el tallo apical y basal presenta diferencias significativas y que la espectroscopía FTIR – ATR es una técnica confiable para determinar la composición química de la madera de árboles jóvenes de ambas especies.

Palabras clave: *Callycophilum spruceanum*, *Guazuma crinita*, espectroscopía FTIR, PLS.

ABSTRACT

The Amazonian trees, Capirona (*Callycophilum spruceanum*) and Bolaina (*Guazuma crinita*), have a great economic importance because of fast-growing cycle and a variety end-use of their wood. The development of competitive and efficient trees with forest breeding programs requires technological tools to select vegetal material based on wood chemical composition. The aim of this study was to identify chemical differences between upper and lower stem of Capirona and Bolaina as well as to develop models for predicting the chemical composition of these two species from Fourier transform infrared spectra (FTIR). FTIR spectra were obtained from 150 samples for each species that were 1.8 year-old. The results showed significant differences between the upper and lower stem for each species in terms of cellulose, hemicellulose, holocellulose and lignin. This variability was useful to build the PLS (partial least squares) models from the FTIR spectra and they were evaluated by root mean squared error of predictions (RMSEP) and ratio of performance to deviation (RPD). In Capirona, lignin model achieved high efficiency because of RMSEP = 0.48 and RPD > 2. On the other hand, cellulose, hemicellulose and holocellulose models had RMSEP values of 2.06, 0.68 and 2.07 and RPD values of 1.96, 1.60 and 1.90 respectively, which were sufficient for the prediction of wood properties (RPD > 1.5). In Bolaina, the cellulose model had excellent predictive capacity (RPD = 6.14). Likewise, the lignin, hemicellulose and holocellulose models showed good predictive capacity (RPD > 2.0). To conclude, both species showed significant differences between upper and lower wood chemical composition, and that FTIR-ATR together with PLS is a reliable method to determine the wood chemical composition in juvenile trees of Capirona and Bolaina.

Keywords: *Callycophilum spruceanum*, *Guazuma crinita*, FTIR spectroscopy, PLS.

I. INTRODUCCIÓN

En Sudamérica, los bosques amazónicos cumplen un papel fundamental en el bienestar del ser humano y del ambiente, sin embargo su creciente deforestación podría alcanzar un punto sin retorno, en el cual se haría imposible sostener este ecosistema. Sólo en los últimos 10 años, actividades como la agricultura, minería, narcotráfico y crecimiento poblacional (Ellwanger *et al.*, 2020) han propiciado la pérdida de 26 millones de Ha de bosques amazónicos (FAO y UNEP, 2020), un área cercana a la superficie de Nueva Zelanda. Además de la pérdida de biodiversidad y hábitats naturales, la deforestación amazónica intensifica el calentamiento global, interfiere en el ciclo del agua, propicia la migración humana y la prostitución en áreas de excavación de oro (Ellwanger *et al.*, 2020). Asimismo, la deforestación expande el contacto del ser humano con vida salvaje portadora de patógenos zoonóticos de forma alarmante (Gibb *et al.*, 2020).

La siembra de especies forestales en sistemas agroforestales se promueven con la finalidad de mitigar los efectos de la deforestación (Sears *et al.*, 2014; Villa *et al.*, 2015). En estos sistemas se asocian árboles con especies agropecuarias de forma que se conserva la biodiversidad y la fertilidad de suelos a la vez que se reforesta y se incrementa la captura de carbono (Sears *et al.*, 2014). Las especies forestales de rápido crecimiento son muy comunes en los sistemas agroforestales, especialmente Capirona (*Callycophilum spruceanum*) y Bolaina (*Guazuma crinita*, por su tolerancia a suelos empobrecidos con alta precipitación y por su ciclo corto de producción de madera: 5-6 años en bolaina y 6-12 años en Capirona (Hecht *et al.*, 2014; IIAP, 2009; Sears *et al.*, 2014; Ushiñahua, 2016). La madera de estas especies tiene alta demanda en el mercado nacional debido a la obtención de tabillas y postes, muy usados en construcción (Sears *et al.*, 2014). En consecuencia, la siembra de estas especies ofrece una actividad económica viable para pequeños agricultores (Sears *et al.*, 2014).

Dada la importancia ecológica y económica de Capirona y Bolaina, la selección temprana de estos árboles es de especial interés para el mejoramiento genético, lo cual ahorra tiempo y dinero (Cornelius *et al.*, 2018). La composición química de la madera (celulosa, hemicelulosa y holocelulosa) es un parámetro muy importante para su selección ya que esta directamente relacionada a la calidad y uso final (Fromm, 2013; H. Pereira *et al.*, 2004). Así la lignina, asociada a la fuerza de tensión, es ideal para producción de tablillas pero influye negativamente en la producción de pulpa, mientras que celulosa da flexibilidad y es ideal para la industria papelera (H. Pereira *et al.*, 2004). Sin embargo, la determinación de la composición química por métodos analíticos convencionales requiere alta cantidad de muestra, uso de reactivos químicos y es costosa (Alonso-Simón *et al.*, 2011).

La espectroscopia FTIR-ATR (infrarrojo con transformada de Fourier – reflectancia total atenuada) es una herramienta analítica moderna caracterizada por su simplicidad; y contrario a las técnicas tradicionales es de corto tiempo de análisis, requiere poca cantidad de muestra y no es destructivo (Alonso-Simón *et al.*, 2011). En FTIR -ATR ha permitido la identificación de celulosa, hemicelulosa, lignina, monosacáridos, compuestos extractivos y proteínas en especies forestales (Alonso-Simón *et al.*, 2011; Chen *et al.*, 2010; Funda *et al.*, 2020; C. Zhou *et al.*, 2015); gracias a la información de la vibración de los enlaces químicos y grupos funcionales de estos compuestos a niveles específicos de energía (Kelly *et al.*, 2011). La complejidad de esta información espectral requiere el uso de métodos multivariados, como la regresión de mínimos cuadrados parciales (PLS) que es uno de los métodos más empleados (Kelly *et al.*, 2011). El PLS predice las propiedades de la muestra mediante un número de variables latentes que resumen la información espectral (Héberger, 2008).

El objetivo de este trabajo fue validar modelos predictivos de celulosa, hemicelulosa, holocelulosa y lignina en plantas jóvenes de Capirona y Bolaina en base a sus espectros FTIR. Como objetivos específicos se: (1) determinó la diferencia composicional entre la sección apical y basal del tallo de ambas especies y (2) se analizó los espectros FTIR para asignar los grupos funcionales de cada compuesto en cada especie.

II. REVISION DE LITERATURA

2.1. SECTOR FORESTAL PERUANO

El Perú es el segundo país con mayor extensión forestal en Sudamérica con 72 millones de hectáreas, según SERFOR (2017). Las áreas que puede ser explotadas ascienden a 24.5 millones de hectáreas bajo la denominación de bosques de producción permanente (bosques primarios aprovechables por concesiones) y 20 millones Ha como bosques de producción en reserva (Galarza y Serna, 2005). En contraste, las plantaciones forestales solo ascienden a 43.5 miles de hectáreas. Como resultado, la principal fuente de madera son los bosques naturales.

A pesar del gran potencial forestal peruano la producción interna no satisface la demanda nacional de productos maderables. Por tal motivo, existe la necesidad de importar pulpa de madera de los países vecinos (SERFOR, 2017). Además, el sector forestal se enfrenta a la deforestación de bosques naturales principalmente por la agricultura, alto grado de informalidad laboral y empresarial y la subvaloración de productos forestales por falta de valor agregado (MINAM, 2016).

En la actualidad el gobierno peruano impulsa la agroforestería en regiones amazónicas con la finalidad de mitigar la deforestación, pobreza e informalidad del sector forestal (Sears *et al.*, 2014). Según el Censo Nacional Agropecuario del Perú (CENAGRO) del 2012 más de 4.5 millones de hectáreas en la amazonia se maneja por agroforestería. La agroforestería es un sistema manejado principalmente por pequeños agricultores quienes cultivan diversas especies forestales y agropecuarias en un mismo espacio. Dentro de este sistema se encuentran los huertos familiares, parches de bosque y purmas (Sears *et al.*, 2014). Estos pequeños agricultores abastecen el mercado forestal local de la construcción de viviendas. Además, estos sistemas facilitan el manejo y conservación de madera, carne de monte y plantas medicinales, y brindan servicios ecosistémicos como captura de carbono, mitigación del cambio climático, conservación de la fertilidad del suelo, prevención de inundaciones

(Sears *et al.*, 2014) y cosecha del agua.

Las purmas conocidas como tierras de barbecho o bosque secundario son ideales para la producción de árboles maderables de crecimiento rápido (Sears *et al.*, 2014). Entre esas especies destacan Capirona (*Callycophyllum spruceanum*) y Bolaina (*Guazuma crinita*) debido a la creciente demanda nacional de su madera (Sears *et al.*, 2014). Ambas especies son usadas para la recuperación de suelos degradados debido a sus características ecológicas (Flores, 2007; Ushiñahua, 2016). El cultivo de estas especies ofrece una actividad sostenible a los pequeños agricultores (Putzel *et al.*, 2013; Sears *et al.*, 2014). Con el apoyo adecuado la producción de estas especies podría permitir pasar de la extracción de madera a la producción sostenible (Sears *et al.*, 2014) y recuperar áreas deforestadas por la minería.

2.2. CAPIRONA

La Capirona (*Callycophyllum spruceanum*) es una especie maderable de crecimiento rápido, ampliamente distribuida en Sudamérica. La importancia de esta especie radica en el aumento de la demanda de su madera, sobre todo en el mercado local. Además, presenta potencial para plantaciones forestales y programas de mejoramiento genético (Sears *et al.*, 2014; Sibille, 2006).



Figura 1: Plantones de Capirona (*Callycophyllum spruceanum*)

2.2.1. Descripción botánica y taxonomía

El árbol de Capirona tiene fuste recto que puede llegar a medir de 20 a 35 m de alto y de 0.5 a 1.2m de diámetro. En estadios temprano este árbol reporta una altura de 1.4 – 1.5 metros a los 6 meses y de 3.5 – 4.7 metros al año de crecimiento. La corteza externa es de color verde con textura lisa y recubierta de ritidoma papiráceo rojizo (Reynel *et al.*, 2003). Las hojas son simples de forma elíptica de 5 a 10 cm de longitud y de 3 a 5 cm de ancho, posee nervación pinnada y están distribuidas en el tallo de forma opuesta. Las flores son hermafroditas de color blanco y se encuentra organizadas en inflorescencias cimosas terminales (Sibille, 2006). La propagación de la especie se da por semillas, cuyo poder germinativo es de 80 a 90% y se propaga mediante almácigos generalmente (Sibille, 2006).

Según Aróstegui (1987), la clasificación taxonómica de la Capirona se muestra en la Tabla 1.

Tabla 1: Clasificación taxonómica de Capirona

Reino	Vegetal
División	Fanerógama
Sub división	Angiosperma
Clase	Dicotiledónea
Familia	Rubiaceae
Género	Calycophyllum
Especie	Calycophyllum spruceanum Benth
Nombre común	Capirona, capirona de bajo, capirona negra

FUENTE: Aróstegui (1987)

2.2.2. Distribución y ecología de capirona

La Capirona se distribuye en toda la amazonia y se considera pionera en la sucesión de la planicie aluvial de Brasil, Colombia, Ecuador y Perú.(Reynel *et al.*, 2003; Ushiñahua, 2016). Se encuentra comúnmente en bosques primarios, pero también en bosque secundarios (Reynel *et al.*, 2003). Esta especie crece en zonas aluviales y ribereñas, desde suelos limosos a arcillosos y por debajo de los 1200 msnm. Sin embargo, también puede crecer en zonas con estación seca marcada (Reynel *et al.*, 2003). En el Perú la densidad de capirona estimada es de 4.5 m³/ha y se distribuye en los departamentos de Amazonas, San Martín, Huánuco, Madre de Dios, Loreto y Ucayali (Sibille, 2006).

Los Capironales son comunidades de capironas donde crecen asociadas a otras especies como Quinilla, Shihuahuaco, Quina Quina, entre otros (Sibille, 2006). Al mismo tiempo, las Capirona crecen en pequeñas parcelas denominadas purmas o tierras agrícolas de barbecho (Sears *et al.*, 2014). Las purmas son la principal fuente de producción de Capirona, pero no abastecen la demanda local (Sears *et al.*, 2014; Ushiñahua, 2016).

La producción de estas especies en purmas brinda servicios ecológicos tales como el mantenimiento de la fertilidad del suelo, la conservación de la biodiversidad y de la carne de monte, la conservación de plantas medicinales y de árboles frutales que a su vez atraen fauna silvestre (Sears *et al.*, 2014). Además, la captura de carbono es otro servicio brindado por los capironales. Debido a su crecimiento rápido y alta densidad, la capirona posee una alta capacidad de almacenamiento de carbono, con contenido de 380 T-CO₂/Ha en bosques capironales no manejados y 690 T-CO₂/Ha en bosques manejados (Otarola & Martinez, 2007).

2.2.3. Madera de capirona

La madera de Capirona presenta características que la hacen un buen material para procesos industriales y otros usos. Entre las principales características tecnológicas de la madera de capirona resalta su alta densidad (0.76 g/cm³), alta durabilidad y resistencia mecánica de media a alta (Sibille, 2006). Asimismo, presenta propiedades químicas como: contenido de celulosa de 59.71% y hemicelulosa de 18.6%, bajo contenido de compuestos extractivos (2.2%) y no es abrasiva debido al bajo contenido de sílice (0.14%). Por su alto poder calórico, la madera de capirona es muy usada como leña.

Además, la madera de capirona es muy resistente al ataque biológico y no requiere preservación. En función del comportamiento de la madera frente al maquinado y propiedades como secado, pegado, rugosidad, trabajabilidad, acabado y estabilidad, la madera de capirona se clasifica como madera de buena trabajabilidad y buena para proceso industriales (Sibille, 2006).

Los principales usos de la madera de capirona son: soporte estructural en construcción (vigas, viguetas y columnas); se usa en ebanistería: muebles, puertas, ventanas y pisos

(machihembrado y parquet); fabricación de artículos deportivos como raquetas de tenis y ping pong; artesanía; herramientas de cocina como cucharones, agitadores y tablas de picar; y utensilios que tienen contacto con alimentos (Orrego Medina y Bustamante Guillén, 2017; Sibille, 2006).

2.3. BOLAINA

La Bolaina (*Guazuma crinita*) es una especie forestal de crecimiento rápido que presenta características ecológicas ideales para la regeneración de bosques. La demanda de madera se ha incrementado debido a sus propiedades y a su bajo costo, además de ofrecer una buena relación costo beneficio para los pequeños agricultores (IIAP, 2009).



Figura 2: Plantones de Bolaina (*Guazuma crinita*)

2.3.1. Descripción botánica y taxonomía

La Bolaina es un árbol de 15-30 m de altura con diámetro de 25-80 cm. Presenta fuste cilíndrico con ramificación en el tercio superior. Puede llegar a medir 2-2.3 m a los 6 meses y 4.9-5.7 m al año. La corteza externa es agrietada y su color varía de marrón claro a grisáceo mientras que la corteza interna es fibrosa de color amarillo fácilmente oxidable a marrón. Las hojas son simples, dísticas, con longitud de 10-18 cm y ancho de 5-7 cm, posee nervación palmeada y se distribuyen de forma alterna en el tallo. Las flores son hermafroditas, de color rosado, con 8-12mm de longitud y están organizadas en panículas

axilares (Reynel *et al.*, 2003). Los frutos son dehiscentes y secos, y su fructificación se da a partir del segundo año. Las semillas miden de 1 a 2 mm y tienen de 60 a 70 % de germinación (IIAP, 2009).

La clasificación taxonómica de Bolaina según Mostacero y Mejía (1993) (IIAP, 2009).

Tabla 2: Clasificación taxonómica de Bolaina

Reino	Vegetal
División	Angiosperma
Clase	Dicotiledoneae
Orden	Malvales
Familia	Sterculiaceae
Género	Guazuma
Especie	crinita
Nombre común	Bolaina blanca

FUENTE: IIAP (2009)

2.3.2. Distribución geográfica y ecología

La Bolaina se distribuye desde Centroamérica hasta la amazonia sur de Brasil y Bolivia. Generalmente se encuentra hasta los 1500 msnm. En el Perú, se encuentra en los departamentos de Amazonas, Cajamarca, Huánuco, Junín, Lambayeque, Loreto, San Martín y Ucayali (IIAP, 2009). Las zonas ecológicas donde crecen las bolainas son: bosques húmedos premontados y bosque muy húmedo subtropical, generalmente se encuentran en bosques secundarios y orillas de los ríos (IIAP, 2009).

La bolaina es una especie heliófila, de rápido crecimiento y alta capacidad regeneración, con requerimiento exigente de luz y es característica de la vegetación secundaria temprana. Esta especie tolera precipitación anual de 1800-2500 mm y 25°C de temperatura media. Crece en suelos ricos con alto drenaje, también tolera suelos pobres y con bajo drenaje (IIAP, 2009). La Bolaina presenta floración y fructificación anual, que alcanza su máxima dispersión de semillas en los meses de septiembre y octubre. En este periodo coincide con el inicio de la temporada lluviosa en la Amazonia peruana, por tal motivo su regeneración anual se ve favorecida. Además, los árboles de Bolaina tienen estrategias de crecimiento rápido que les ayuda a ocupar claros y alcanzar la madurez fisiológica en el primer año. Por otro lado, la poca cobertura de dosel y profuso enraizamiento de la Bolaina la hace una especie adecuada

para sistemas silvopastoriles (IIAP, 2009). La Bolaina es capaz de crecer de forma abundante en zonas con alteración antropogénica y cerca de caminos (Reynel *et al.*, 2003). En adición, después del aprovechamiento del fuste principal la Bolaina rebrota fácilmente por lo que no es necesario volver a plantarla en el terreno (IIAP, 2009).

En el Perú, los árboles de Bolaina se encuentran asociados con cultivos agrícolas como plátano, yuca, maíz, frejol, cítricos, cacao. También se encuentran en parcelas pequeñas de bosques secundarios regenerados, y son aprovechadas a los cuatro años (edad comercial). Asimismo, se cultiva Bolaina como monocultivo y sistemas agroforestales (IIAP, 2009).

Debido a las características ecológicas previamente mencionadas, la Bolaina es una de las especies con mayor potencial para la producción forestal, mejoramiento genético y regeneración de bosques (IIAP, 2009).

2.3.3. Madera de bolaina

La madera de Bolaina es liviana y blanda (Reynel *et al.*, 2003). Con respecto a las características organolépticas, esta madera es de color blanco en la albura y el duramen, no presenta olor o sabor (Flores, 2007). Su composición química es 22-24% de lignina y 72-75% de holocelulosa (Contreras *et al.*, 2020). La densidad es de 0.41 g/cm³ la clasifica como madera de densidad mediana y la relación tangencial T/R = 1.57 indica buen comportamiento al secado (Flores, 2007). Además de buena calidad y durabilidad, la madera presenta excelente comportamiento al cepillado y buena trabajabilidad (Flores, 2007; Reynel *et al.*, 2003). Sin embargo, esta madera es susceptible al ataque biológico (Flores, 2007).

La madera de esta especie se usa en la fabricación de mondadientes, palitos de chupete, baja lenguas y palitos de fosforo. Asimismo, se fabrican paneles de madera para la construcción de casa prefabricadas. La madera cepillada tiene alta demanda en la manufactura de muebles, molduras, embalajes, carpintería de puertas y ventanas (IIAP, 2009). La corteza viva de Bolaina es usada como sogá y la pulpa es apta para la industria papelera (Flores, 2007).

2.4. XILOGENESIS

La madera es el xilema secundario producido por el cambium vascular (Ye y Zhong, 2015). El cambium juega un papel principal en la producción de madera y crecimiento diametral en tallos y raíces. Asegura la vida perenne del árbol gracias a la renovación continua de xilema y floema. El cambium vascular es un meristemo secundario derivado del procambium. Está compuesto por una fila de células juveniles llamadas células iniciales, las cuales origina a las células madre del floema y xilema mediante divisiones. La zona del cambium está compuesta por dos tipos de células, las primeras son células radiales que dan lugar a rayos que son esenciales para la translocación transversal de nutrientes entre el floema y xilema. El segundo tipo de células son fusiformes y produce tejido vascular secundario a través de divisiones periclinales en dos sentidos, hacia el lado interior da lugar al xilema (principalmente traqueidas en gimnospermas, pero también elementos de vasos, células asociadas a la red, parénquima axial y fibras en dicotiledóneas); hacia el exterior, células del floema conformados por tubos de tamiz y, en dicotiledóneas, células compañeras, parénquima axial y fibras (Plomion *et al.* 2018, Ye y Zhong, 2015).

El desarrollo del xilema involucra: expansión celular, deposición de multicapas de la pared celular secundaria, lignificación y muerte celular (Plomion *et al.*, 2018). Las células derivadas de las células iniciales forman la lamela media después de la división celular. Es la primera lámina en formarse, está compuesta por pectina y su función principal es mantener a las células vecinas adheridas entre sí. La formación de la pared celular primaria, adherido a la lamela media, empieza conforme se inicia la diferenciación celular. Es la segunda lámina en formarse y es flexible, lo cual permite a la célula expandirse radial y longitudinalmente hasta alcanzar su tamaño final. Está compuesta por capas de microfibrillas de celulosa que pueden contener lignina, pectina y hemicelulosa. Al alcanzar su tamaño final la célula inicia la formación de una capa interna llamada pared celular secundaria, esto conlleva la activación de numerosos genes involucrados en la síntesis de celulosa, hemicelulosa, otros componentes (proteínas, estilbenos,) y lignina (Plomion *et al.*, 2018).

Una vez completada la lignificación, las células del xilema experimentan muerte celular programada a través de hidrolasas sintetizadas y almacenadas en de forma inactiva en la vacuola, hasta que una señal desencadena un flujo de calcio que provoca el colapso de la vacuola y se liberan las hidrolasas que degradan el contenido celular pero no la pared

secundaria (Plomion *et al.*, 2018).

2.5. COMPONENTES QUÍMICOS DE LA MADERA

2.5.1. Celulosa

La celulosa es un polímero fibrilar formado por residuos de glucosa unidas por enlaces β -1,4 (Saxena y Brown, 2005) organizado en cadenas paralelas en forma de micro fibrillas con ordenamiento cristalino (Festucci-Buselli *et al.*, 2007; Fromm, 2013; Somerville, 2006). Este polímero envuelve a las células vegetales, su deposición orienta la dirección de la elongación celular y es responsable de la forma y tamaño de las células vegetales (Brown, 2004; Fromm, 2013). En decir, la celulosa cumple el rol de esqueleto vegetal que mantiene y controla el proceso de crecimiento (Brown, 2004). Además, protege el contenido celular contra agentes externos (Brett, 2000; Saxena y Brown, 2005) y provee resistencia y rigidez para contener la presión osmótica dentro de la célula (Somerville, 2006).

2.5.1.1. Estructura de la celulosa

La unidad monomérica de la celulosa es glucosa en forma de anillo (glucopiranos) es que más estable que la forma lineal (Mcnamara *et al.*, 2016). La glucosa es un carbohidrato de 6 carbonos (hexosa) pero el anillo está formado por 5 carbonos y un oxígeno, el sexto carbono está fuera del anillo y tiene su propio grupo hidroxilo. En la Figura 3 se observan glucopiranosas en conformación de sillas donde el carbono 1 es el único que está unido a dos oxígenos (uno ubicado en el anillo y el otro de su grupo hidroxilo) y se le conoce como carbono anomérico. La orientación del grupo hidroxilo de este carbono determina si el anillo tiene conformación α (orientado hacia abajo) o β (orientado hacia arriba) (L. G. Wade, 2004). Cada cadena de glucosas presenta un extremo reductor conformado por un grupo carboxilo y una estructura alifática mientras que el extremo no reductor formado por un anillo cerrado y es donde se polimerizan las unidades de glucosa (Festucci-Buselli *et al.*, 2007).

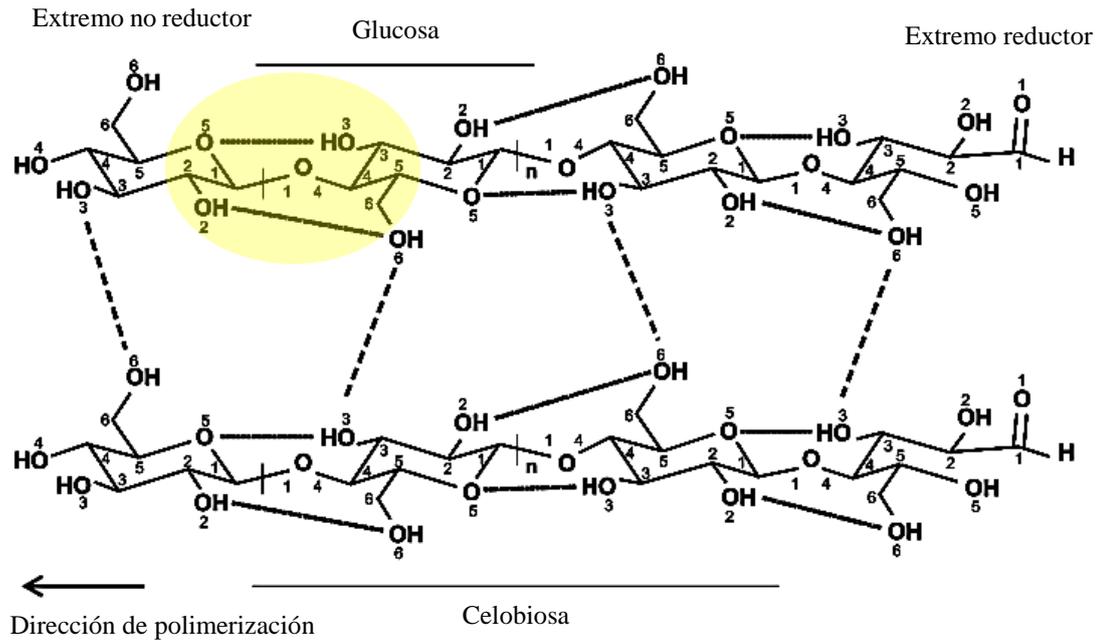


Figura 3: Estructura molecular de la celulosa

FUENTE: Festucci-Buselli *et al.* (2007)

Los enlaces intramoleculares estabilizan las cadenas de celulosa. El enlace β -1,4 se forma por la unión entre una glucopiranososa y otra que ha girado 180° grados, de modo que el carbono 1 del primer monosacárido (conformación beta) y el carbono 4 del segundo se enlazan. Otros enlaces como puente de hidrogeno O5---H-O3 y O2---H-O6 estabilizan la conformación del anillo de glucosa y forman un polisacárido plano y rígido, como se observa en el círculo amarillo de la Figura 3 (Festucci-Buselli *et al.*, 2007; Mcnamara *et al.*, 2016). Los enlaces intermoleculares promueven el empaquetamiento de las cadenas de celulosa en forma paralela y cristalina (S. Li *et al.*, 2014, Nishiyama *et al.*, 2003). Por ejemplo, los enlace de puente de hidrogeno O3-H---O5 unen las cadenas de celulosa y se muestran como líneas punteadas en la Figura 3 (Festucci-Buselli *et al.*, 2007). Las cadenas de glucosa también se mantienen unidas por fuerzas de van der Waals (Brett, 2000) que genera fuerzas de repulsión y atracción debido al carácter hidrofóbico e hidrófilo de los grupos hidroxilos y átomos de hidrógeno de los anillos (Mcnamara *et al.*, 2016).

Debido a la estructura de la celobiosa (2 glucosas unidas por β -1,4) y a las fuerzas que unen las cadenas de celulosa, ésta toma una conformación cristalina. En la naturaleza se encuentra, celulosa cristalina I y II (Brett, 2000). La forma I presenta dos alomorfos: $I\alpha$ presente en abundancia en algas y bacterias y $I\beta$, que es más estable y está presente en árboles, algodón

y tunicados (Festucci-Buselli *et al.*, 2007). Un individuo puede contener ambas aloformas y la relación entre las mismas determina las propiedades del polímero celulosa (Saxena y Brown, 2005). La diferencia entre las dos aloformas es su empaquetamiento cristalino, conformación molecular y puentes de hidrogeno, lo cual influye en sus propiedades físicas (Saxena y Brown, 2005). Con respecto al enlace de hidrogeno presente en las cadenas (O2--H-O6) es más corto en I α que en I β (Festucci-Buselli *et al.*, 2007).

2.5.1.2. Síntesis de celulosa

La celulosa es sintetizada por proteínas CESA a partir de residuos de UDP- α -glucosa y se elonga por reacciones de sustitución (Saxena y Brown, 2005; Mcnamara *et al.*, 2016). Las CESAs son proteínas integrales de membrana pertenecientes a la familia 2 glicosiltransferasa y presenta función invertasa, es decir cambian la configuración de carbonos anoméricos de α a β (Richmond, 2000; Saxena y Brown, 2005). El azúcar entrante UDP- α -glucosa, cuyo carbono 1 anomérico es α cuando está unido al UDP, cambia su configuración a β una vez que se une a la celulosa en elongación por acción de la proteína CESA (Mcnamara *et al.*, 2016). La elongación de las cadenas de celulosa se da por reacción de sustitución. El extremo no reductor del polímero en elongación es el nucleófilo que ataca a un residuo de glucosa-UDP, el grupo UDP es desplazado y se incorpora el residuo de glucosa a la cadena en crecimiento (Mcnamara *et al.*, 2016).

Existen 3 isoformas de las proteínas CESA conocidas como α_1 , α_2 y β , entre las cuales ocurren interacciones proteína-proteína de la forma: $\beta - \beta$, $\alpha_1 - \beta$, $\alpha_2 - \beta$ (Festucci-Buselli *et al.*, 2007). Estas interacciones están implicadas en la formación del complejo terminal (TC) ubicado en la membrana celular, como se muestra en la Figura 4 (Festucci-Buselli *et al.*, 2007).

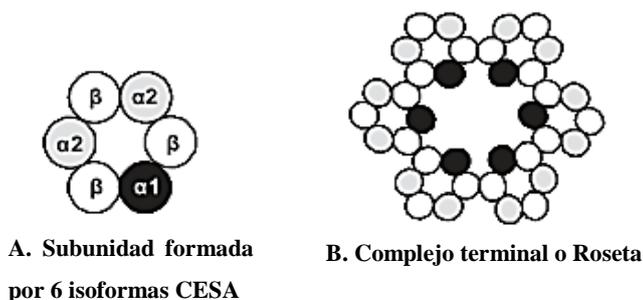


Figura 4: Complejo terminal o Roseta

FUENTE: Festucci-Buselli *et al.* (2007)

La celulosa es sintetizada dentro de un complejo terminal (TC) o roseta que usualmente se conoce como complejo celulosa sintasa (Festucci-Buselli *et al.*, 2007). Cada roseta está compuesta por 6 subunidades, dispuestas en simetría hexagonal, y cada subunidad está constituida por 6 proteínas CESA como se muestra en la Figura 4 (Festucci-Buselli *et al.*, 2007; Saxena y Brown, 2005). En ese sentido, cada roseta puede sintetizar 36 cadenas de glucosa (Festucci-Buselli *et al.*, 2007; Saxena y Brown, 2005). Además, se presume que los complejos terminales se ensamblan en el retículo endoplasmático y el aparato de Golgi, luego es transportada a la membrana celular donde se activa y comienza la síntesis de celulosa (Saxena y Brown, 2005). Debido a la distribución de las proteínas CESA, las cadenas de glucosa forman puentes de hidrogeno que constituyen las microfibrillas de celulosa como se muestra en la Figura 3 (Wightman y Turner, 2010). Los complejos terminales o rosetas ubicadas en la superficie de la membrana celular atraviesan la membrana y forman un canal que permite a los residuos de glucosa provenientes del citoplasma ser polimerizados en el complejo para posteriormente pasar a través de la membrana hacia el exterior (Festucci-Buselli *et al.*, 2007).

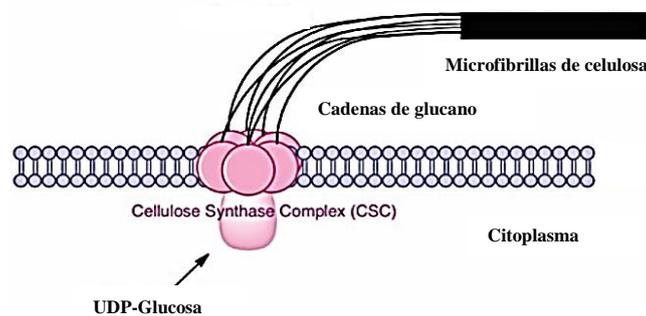


Figura 5: Complejo celulosa sintasa en la membrana celular

FUENTE: Wightman & Turner (2010)

Las microfibrillas de celulosa se depositan en dirección perpendicular al eje de crecimiento, esto permite orientar el crecimiento de las células mitóticas vegetales (S. Li *et al.*, 2014). Muchos autores propusieron teorías acerca la regulación de la dirección de la deposición de celulosa, la más extendida es la teoría del citoesqueleto que controla la dirección de síntesis de las microfibrillas de celulosa mediante asociación funcional de las proteínas CESA y microtúbulos corticales (S. Li *et al.*, 2014; Paredez *et al.*, 2006).

2.5.2. Hemicelulosa

La hemicelulosa es el componente más complejo de la pared celular (Ren y Sun, 2017), comprende los polisacáridos solubles en soluciones alcalinas (Scheller y Ulvskov, 2010) y constituye del 25-35% del peso seco en árboles (Fromm, 2013). Estructuralmente, la hemicelulosa es un heteropolisacárido compuesto por hexosas (glucosa, galactosa y manosa) y pentosas (xilosa y arabinosa) que forma cadenas ramificadas y de corta polimerización (50 a 200 monómeros). Las cadenas de hemicelulosa están ligados a otros compuestos de la pared celular mediante enlaces puente de hidrogeno con la celulosa y enlaces covalentes con lignina (Ren y Sun, 2017). La hemicelulosa es amorfa debido a estos enlaces y diversos sustituyentes asociados al esqueleto carbonado (Pauly et al., 2013).

Junto a celulosa y pectinas, la hemicelulosa pertenece a los componentes de construcción de la pared celular vegetal en plantas superiores (Ren y Sun, 2017) y su función principal es reforzar la pared celular (Scheller y Ulvskov, 2010). Además, la hemicelulosa cumple la función de almacenamiento de carbohidratos en semillas y son fuente de señales moleculares. Asimismo, el xilano (principal componente de la hemicelulosa) confiere fuerza a la pared secundaria (Scheller y Ulvskov, 2010). Esto determina el potencial uso de la hemicelulosa para la producción de etanol y la producción de papel (Zhao *et al.*, 2017).

Las hemicelulosas están divididas en 4 grupos de polisacáridos: xilanos, mananos, xiloglucanos y glucanos con enlaces mixto. El primero grupo, los xilanos, son las hemicelulosas más abundantes en la pared secundaria en monocotiledóneas (Grass y cereales) y dicotiledóneas (plantas herbáceas y arboles de madera dura). Pueden llegar a constituir del 20- 30% y 50% de la biomasa vegetal, respectivamente (Ren y Sun, 2017). Los xilanos son esenciales en el desarrollo de tejidos con pared secundaria, resistencia de tejidos portadores de carga y capacidad de almacenamiento en semillas (Pauly *et al.*, 2013). Los mananos, el segundo grupo, son las hemicelulosas más antiguas y son abundantes en la pared secundaria de gymnospermas. Su función es reforzar la rigidez de la pared celular, almacenamiento en semillas, señalización celular, diferenciación tisular y embriogénesis (Pauly *et al.*, 2013). Por otro lado, los xiloglucanos son abundantes en la pared primaria de dicotiledóneas y forman una red asociada a celulosa que constituye el componente de carga. Su presencia está asociada con la correcta elongación celular y almacenamiento en semillas. Finalmente, los glucanos con enlace mixto están restringidas a grasses, también se presenta

en algas y hepáticas. Se cree que su función principal es el almacenamiento de energía usada en la elongación celular (Pauly *et al.*, 2013).

2.5.2.1. Estructura de la Hemicelulosa

Las hemicelulosas se diferencian estructuralmente en base a su esqueleto carbonado formado por monómeros de xilosa, manosa y glucosa unidos mediante enlaces β -1,4 (Holtzapfle, 2003; Scheller y Ulvskov, 2010). Además, las cadenas laterales compuestas por otros azúcares (xilosa y arabinosa) pueden encontrarse acetiladas, metiladas o unidas a ácido glucurónico (Holtzapfle, 2003).

Los xilanos consisten en cadenas lineales de xilosa unidas por enlace β -1,4 que presentan grupos sustituyentes diversos en función de la especie y el tejido vegetal. En función de los grupos sustituyentes ligados a las cadenas de xilanos se encuentra: glucoronoxilanos y glucuronoarabinoxilanos. En los glucoronoxilanos, las cadenas de xilanos pueden estar acetiladas y ligadas a ácido glucurónico metilado y no metilado. En los glucuronoarabinoxilanos, el esqueleto carbonado está acetilado, unido a arabinofuranosil, ácido glucurónico metilado y no metilado. En los xilanos en general, los ácidos glucurónicos metilados y no metilados pueden unirse por enlaces α -1,2 a la cadena principal, mientras que los arabinofuranosil mediante enlaces α -1,2 y α -1,3 (Pauly *et al.*, 2013).

Los mananos se dividen en 4 grupos en función de la composición de su esqueleto carbonado y los grupos laterales asociados (Figura 6). Con respecto a los esqueletos carbonados compuestos por manosa unidas por enlace β -1,4, se tiene los grupos: mananos y galactomananos. Asimismo, los esqueletos de glucosa y manosa unidas por enlace β -1,4 constituyen los grupos: glucomananos y galactoglucomananos. En función de los grupos sustituyentes, los galactoglucomananos y galactomananos presentan galactosa en sus grupos sustituyentes ligados mediante enlaces α -1,6 (Pauly *et al.*, 2013). Los mananos se encuentran acetilados de forma frecuente (Scheller y Ulvskov, 2010).

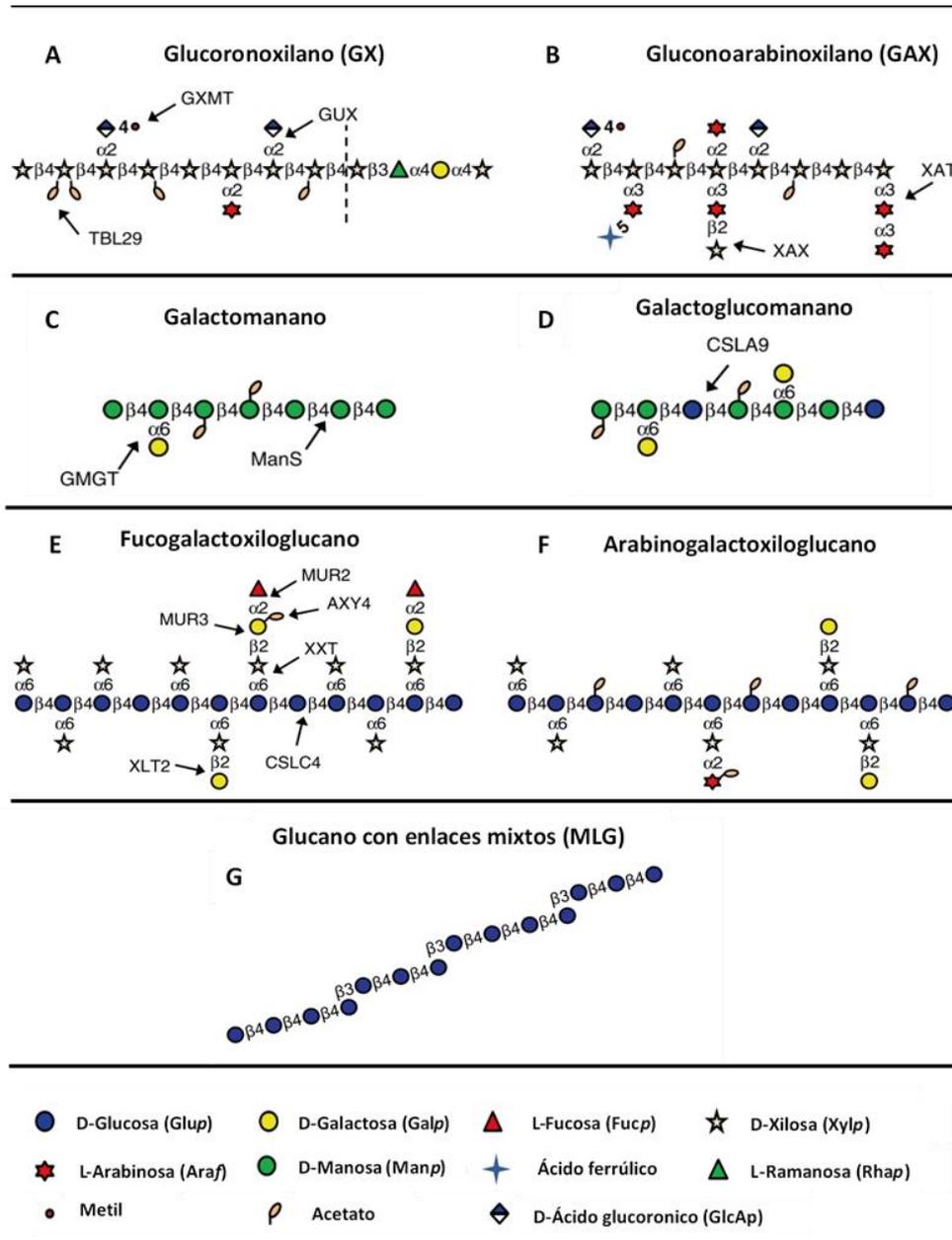


Figura 6: Estructura química de las hemicelulosas

FUENTE: Pauly *et al.* (2013).

Los xiloglucanos están conformados por cadenas de glucanos unidas por enlaces β -1,4 y sustituidas por grupos xilosa (mediante enlace α -1,6) que a su vez están sustituidos por galactosa, arabinosa, fucosa y otros residuos glucosil (Pauly *et al.*, 2013). Por otro lado, los glucanos de enlace mixto son polímeros no ramificados que no presentan grupos sustituyentes (Pauly *et al.*, 2013). Este grupo de hemicelulosa, está formado por la unión de glucosas mediante enlaces β -1,3 y β -1,4 (Scheller y Ulvskov, 2010). Su estructura está compuesta por cadenas de 3 y 4 glucosas unidas por enlaces β -1,4; estas cadenas se alternan y se unen entre ellas mediante enlaces β -1,3 (Pauly *et al.*, 2013).

2.5.2.2. Síntesis de Hemicelulosa

En general, la biosíntesis de hemicelulosa ocurre en el aparato de Golgi en dos pasos principales. La formación de los esqueletos de azúcares a través de polisacáridos sintasas y los sustituyentes son añadidos a través de glucosil transferasas una vez formados los esqueletos (Fromm, 2013). En el caso de la síntesis de los mananos se da a partir de los nucleótidos activados: manosa-GDP, glucosa-GDP y galactosa-GDP. Las enzimas β manano sintasas (ManS) son las encargadas de la polimerización de los monómeros de manosa, esta enzima pertenece a un grupo llamado Familia A y son análogos de celulosa sintasa (CSLA), las cuales pertenecen a la familia de las glucosil transferasas GT (Pauly *et al.*, 2013). En el grupo CSLA también se han estudiado enzimas con actividad glucomanosa sintasa, es decir presentan preferencia por GDP- glucosa y GDP-manosa como sustrato para formar cadenas de glucosa y manosa. Algunos autores señalan que la incorporación de estos azúcares se da en relación con la concentración relativa de los mismos en el momento de la síntesis. La polimerización se da en el aparato de Golgi donde se encuentran los nucleótidos, lo cual facilita la formación de enlaces específicos entre monómeros y luego la síntesis del monómero. Otra enzima implicada es galactosiltransferasa que incorpora galactosa como azúcar lateral en la síntesis de galactomananos. La concentración de UDP- galactosa y GDP-manosa determina el grado de sustitución de los galactomananos. Mientras más sustituciones de galactosa presente en los mananos más soluble será el polímero. (Scheller y Ulvskov, 2010).

2.5.3. Lignina

El término lignina proviene del latín *lignum* que significa madera (Barros *et al.*, 2015). La lignina es el segundo biopolímero más abundante en plantas, su contenido es del 20 a 40% en árboles y se deposita directamente en la pared celular de células especializadas. (Barros *et al.*, 2015; Fromm, 2013). Este polímero hace a las células más rígidas e impermeables y su síntesis se induce frente a estrés biótico o abiótico (Vanholme *et al.*, 2010). La lignina está estrechamente relacionada a la funciones de resistencia mecánica, protección de las células vegetales contra patógenos externos, crecimiento, y está asociada al sistema vascular que permite el transporte de sustancias (Baucher *et al.*, 1998). La presencia de este polímero es exclusivo de plantas vasculares como pteridofitas, gimnospermas y angiospermas (Baucher *et al.*, 1998).

Sin embargo, la función de la lignina dependerá del tejido u órgano vegetal al cual está

asociado. En otras palabras, la lignina exhibe propiedades diferentes para cada tipo de célula (Barros *et al.*, 2015). La resistencia del tallo a doblarse o quebrarse está en función de la altura de la planta, biomasa, diámetro de tallo y composición de la pared celular. La acumulación de lignina mejora la resistencia mecánica del tallo (Liu *et al.*, 2018). También se observa que si el tallo está inclinado la concentración de lignina se incrementa en esa sección de la planta (Shi *et al.*, 2012). En plantas como *Arabidopsis thaliana* la lignina está involucrada en los procesos de germinación y su ausencia reduce el porcentaje de germinación y crecimiento. Así mismo, la ausencia de lignina está relacionada a la esterilidad del gametófito masculino, es probable que la lignina tiene un rol importante en la estructura de las anteras (Liu *et al.*, 2018).

El contenido de lignina se incrementa en condiciones de estrés abiótico y biótico (Moura *et al.*, 2010). Por ejemplo, la infesta de insectos dispara la expresión de genes relacionados con la síntesis de lignina. Esto puede ser un mecanismo sinérgico de regulación que permite la resistencia de los tejidos vegetales a insectos (Liu *et al.*, 2018). La lignina es una barrera mecánica no degradable que bloquea a la mayoría de los microorganismos (Moura *et al.*, 2010). El incremento de lignina, por lo tanto, provee una barrera contra patógenos e impide la infiltración y propagación de toxinas y enzimas micóticas en tejidos vegetales (Liu *et al.*, 2018). Por otro lado, el estrés abiótico como la deficiencia de nutrientes, metales pesados o alta concentración de nutrientes afecta la acumulación de lignina. La luz es esencial para el crecimiento de las plantas, sin embargo la exposición prolongada dispara mecanismo de defensa frente al estrés oxidativo y también la incrementa la síntesis de lignina (Moura *et al.*, 2010).

2.5.3.1. Estructura de lignina

La lignina se forma por un proceso de acoplamiento de tres monómeros diferentes: p-hidroxifenol (**H**), Guaiacil (**G**) y siringil (**S**). Estos monómeros son originados a partir de alcoholes hidroxinamil o monolignoles: alcohol p-cumaril, coniferil, sinapil respectivamente (Fromm, 2013). En la Figura 7 se observa la estructura química de los monolignoles, los cuales presentan conformación C6-C3, es decir un anillo aromático de 6 carbonos unido a un radical de tres carbonos (Baucher *et al.*, 1998). La diferencia entre los monolignoles es su grado de metilación, donde p-coumaril no está metilado, Coniferil tiene el carbono 3 metilado y el carbono 5 libre para formar enlaces, Sinapil tiene metilado tanto el carbono 3

y 5. Por lo mencionado previamente, se puede decir que la lignina es un heteropolímero aromático complejo (Boerjan *et al.*, 2003).

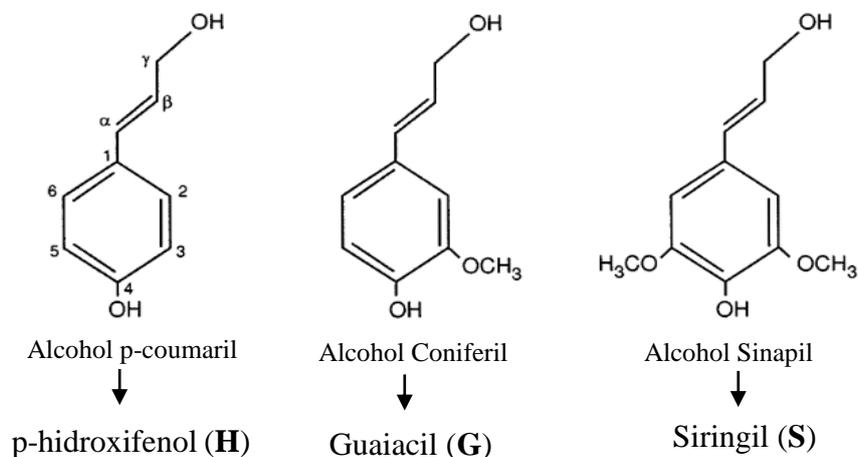


Figura 7: Monolignoles, unidades monoméricas de lignina

FUENTE: Baucher *et al.* (1998)

El contenido y la composición varía entre especies, individuos de una misma especie, tejidos celulares de un individuo, células de un mismo tejido, condiciones ambientales y estado de desarrollo del individuo (Fromm, 2013). Con relación al contenido de lignina, el contenido de la misma puede variar en secciones del tallo. Hennion *et al.*, 1992; citado por Bucher *et al.*, 1998, reportó que el contenido de lignina es 3 veces mayor en la base del tallo de plantas de espárrago que la parte apical. Así mismo, se observó variación en el contenido de lignina en lámina media, capa S2, células fibrilares y radiales de otras especies (Fromm, 2013). Además, se observa variación en la composición a nivel de la proporción de monómeros y los enlaces que los unen (Baucher *et al.*, 1998). Por ejemplo, en plantas gimnospermas la lignina está compuesta principalmente por Guaiacil (G) y bajas concentraciones de p-hidroxifenol (H). En angiospermas dicotiledóneas, la lignina consiste en unidades de Guaiacil (G) y Siringil (S) con trazos de H (Boerjan *et al.*, 2003).

En relación con los enlaces que unen los monómeros de lignina se conoce que el más abundante es β -O-4 (A) y se caracteriza por no ofrecer resistencia a la degradación química. Otros enlaces que unen los monómeros de lignina son β -5 (B), β - β (C), 5-5 (D2 y D), 5-O-4 (E) (Boerjan *et al.*, 2003; Fromm, 2013). Estos enlaces fueron denominados como

condensados porque ofrecen mayor resistencia a la degradación química (Baucher *et al.*, 1998). Los diferentes enlaces son posibles en los polímeros de lignina, pero no todos son igual de probables (Ralph *et al.*, 2004). Por ejemplo, en ligninas con alto contenido de unidades Guaiacil (**G**) se forman enlaces β -5 B, 5-5 y 5-O-4; debido al carbono 5 que se encuentra libre para formar enlaces. En contraste, Siringil (**S**) tiene el carbono 5 metilado.

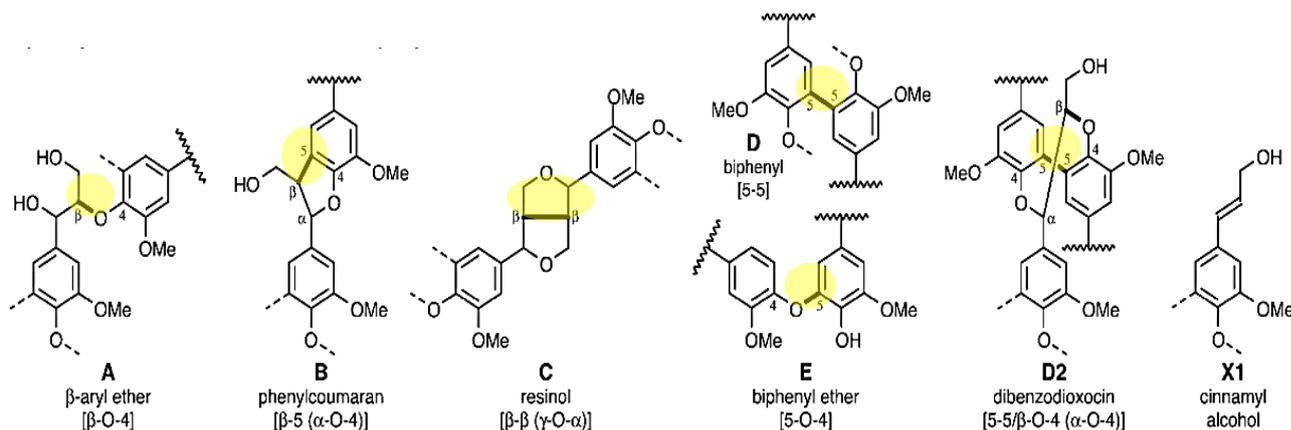


Figura 8: Unidades estructurales del polímero lignina

FUENTE: Ralph *et al.* (2004)

El tipo de enlace depende de los monómeros presentes en el proceso de formación de lignina (Boerjan *et al.*, 2003).

En la Figura 8 las líneas onduladas representan polímeros de lignina creciente u oligómeros, los enlaces entre monómeros están resaltados y etiquetados con letras mayúsculas y las líneas punteadas son enlaces químicos con compuestos que no se grafican por razones didácticas. El proceso de acoplamiento de monómeros a una cadena de lignina creciente se conoce como lignificación y genera enlaces **A**, **B** y **D2**, donde los monómeros se acoplan a una cadena en crecimiento (líneas ondeadas). La dimerización es un proceso diferente a la lignificación ya que se acoplan dos monómeros de lignina y forman enlace **C**. El acoplamiento de dos oligómeros forma enlaces **D** y **E** (Ralph *et al.*, 2004).

2.5.3.2. Síntesis de lignina

La síntesis de lignina es un proceso que se divide en tres etapas: síntesis de monómeros de lignina, transporte y polimerización. Los monómeros de lignina (monolignoles) son sintetizados en el citoplasma y transportados hacia el espacio celular exterior (apoplasto),

donde se da la polimerización de lignina a partir de los monolignoles (Liu *et al.*, 2018).

- **Síntesis de monómeros de lignina**

La síntesis de monómeros de lignina empieza con la fenilalanina, la cual sufre desaminación e hidroxilación y metilaciones a nivel del anillo aromático de forma sucesiva para dar lugar a los monómeros de lignina H, S y G (Boerjan *et al.*, 2003; Liu *et al.*, 2018).

La primera enzima que interviene en la síntesis es PAL (fenilalanina amonio liasa), esta enzima cataliza la desaminación de fenilalanina para obtener ácido cinámico. Este ácido es hidroxilado por C4H (cinamato 4 hidroxilasa) a ácido p-coumarico (Harakava, 2005). La enzima C3H (p-cumarato 3 hidroxilasa) desvía la síntesis de monómeros de lignina hacia la síntesis de G y S, mediante la hidroxilación del carbono 3 del anillo aromático, como se observa en figura 9 (flecha roja) (Barros *et al.*, 2019). El producto obtenido es ácido cafeico (con carbono 3 y 4 hidroxilado) que mediante la enzima COM (ácido cafeico O metil transferasa) sufre metilación a nivel del carbono 3 se obtiene ácido ferrúlico. Este ácido es hidroxilado en el carbono 5 mediante la enzima F5H (ferrulato 5 hidroxilasa), como producto se tiene ácido 5 hidroxiferrúlico (con carbono 3 metilado, carbono 4 y 5 hidroxilado). Para lograr la metilación del carbono 5, la enzima COM cataliza la metilación y se obtiene ácido sináptico (Harakava, 2005). En este punto, las moléculas ácido p-coumaril, cafeico, ferrulico, 5 hidroxiferulico y sinapico han sufrido hidroxilaciones y metilaciones, el siguiente paso es la incorporación de coenzima A(CoA) sobre estas moléculas.

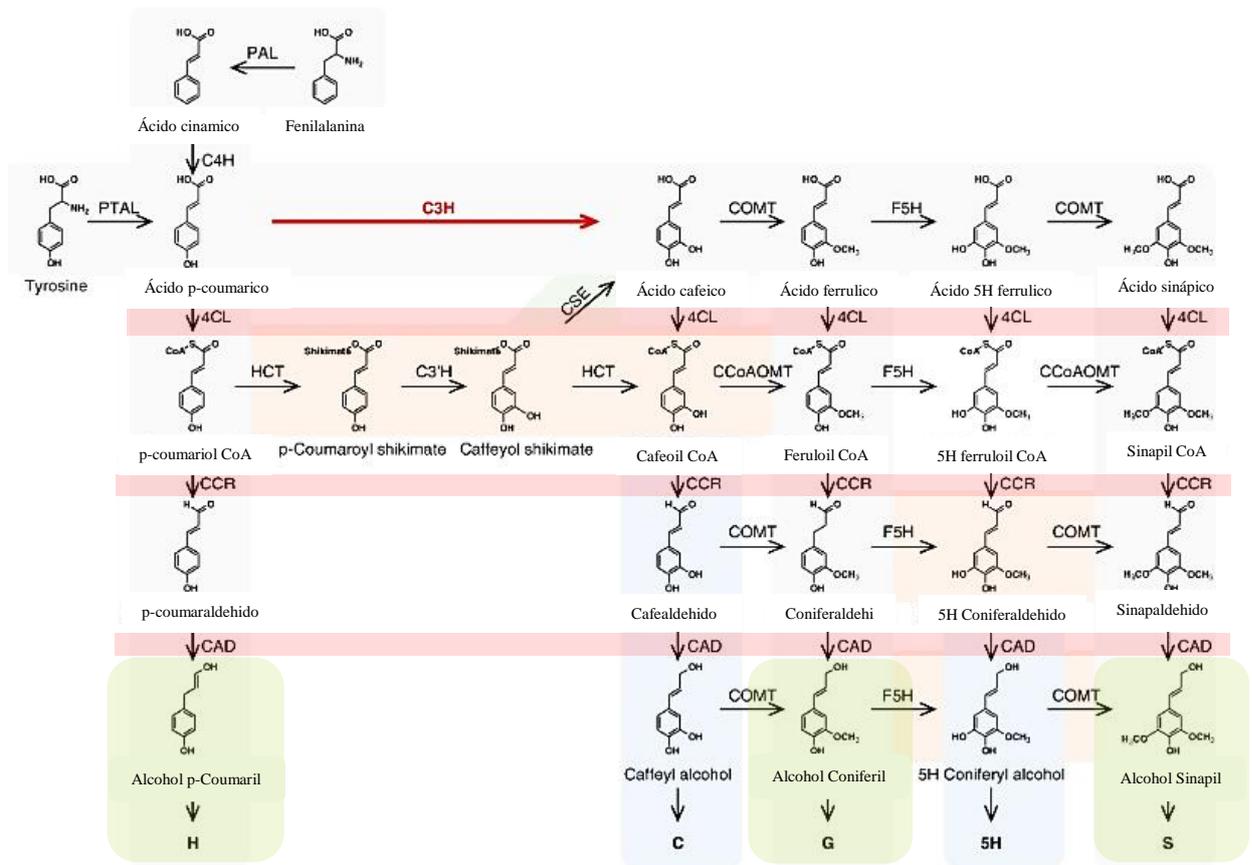


Figura 9. Síntesis de monolignoles de Lignina

FUENTE: Barros *et al.* (2019).

La enzima 4CL (4 coumarato CoA ligasa) incorpora la coenzima A sobre los ácidos antes mencionados, y se obtiene como productos p coumaril CoA, cafeoil CoA, ferrouil CoA, 5 hidroxiferrouil CoA y sinapoil CoA. Estas moléculas se convierten a sus respectivos aldehídos por la enzima CCR (cinamol CoA reductasa): p-coumaraldehído, cafealdehído, coniferaldehído, 5 hidroxiconiferaldehído y sinapaldehído. El último paso en la síntesis de monolignoles es la reducción de los aldehídos a sus alcoholes mediante CAD (alcohol cinamil deshidrogenasa) (Barros *et al.*, 2019; Baucher *et al.*, 1998). Los monolignoles alcohol cafeil e hidroxiconiferil alcohol, son unidades de lignina descubiertos recientemente. Los alcoholes p coumaril, coniferil y sinapil son los monolignoles más abundantes (Barros *et al.*, 2019).

- Transporte de monómeros de lignina

Una vez sintetizados, los monómeros de lignina se transportan para su acoplamiento

oxidativo en la pared celular. El transporte de monolignoles se puede dar por difusión pasiva a través de la bicapa lipídica, sin embargo, se considera que juega un rol menor en el transporte de monolignoles. Otro mecanismo propuesto es exocitosis, este mecanismo secreta vesículas que contienen los monómeros de lignina. Así mismo, se propuso transportadores ABC-G dependientes de ATP como mecanismo de transporte de monolignoles (Figura 10), pero se hizo pruebas con inhibidores de transportador ABC y la síntesis de lignina no se vio afectada. Se concluye que el mecanismo para la exportación de monolignoles requiere de mayor estudio (Barros *et al.*, 2015).

- **Polimerización de lignina**

La polimerización de lignina acopla un monolignol a la vez. El proceso empieza con la deshidrogenación de un monolignol, como resultado se tiene un monolignol radical. Este radical es altamente reactivo y es capaz de formar un enlace covalente con otro monolignol que también está en su forma oxidada (1 electrón desapareado). Así se forman un dímero de lignina, que vuelve a deshidrogenarse y acoplar un nuevo monolignol oxidado. Así mismo, el acoplamiento de dos oligómeros de lignina puede ocurrir pero su frecuencia es baja (Vanholme *et al.*, 2010).

En el proceso de deshidrogenación de monolignoles están implicadas lacasas y peroxidasas (Liu *et al.*, 2018; Vanholme *et al.*, 2010). Las lacasas consumen oxígeno (O_2) para oxidar a los monolignoles mientras que las peroxidasas usan peróxido de hidrógeno (H_2O_2) (Boerjan *et al.*, 2003). De ambas reacciones de oxidación se obtiene monolignoles con un electrón desapareado y moléculas de agua (Baucher *et al.*, 1998). Ambas enzimas pueden actuar de forma complementaria (ambas actúan en diferentes tejidos) o secuencial (las lacasas inician el proceso y las peroxidasas lo terminan) en el proceso de lignificación (Barros *et al.*, 2015).

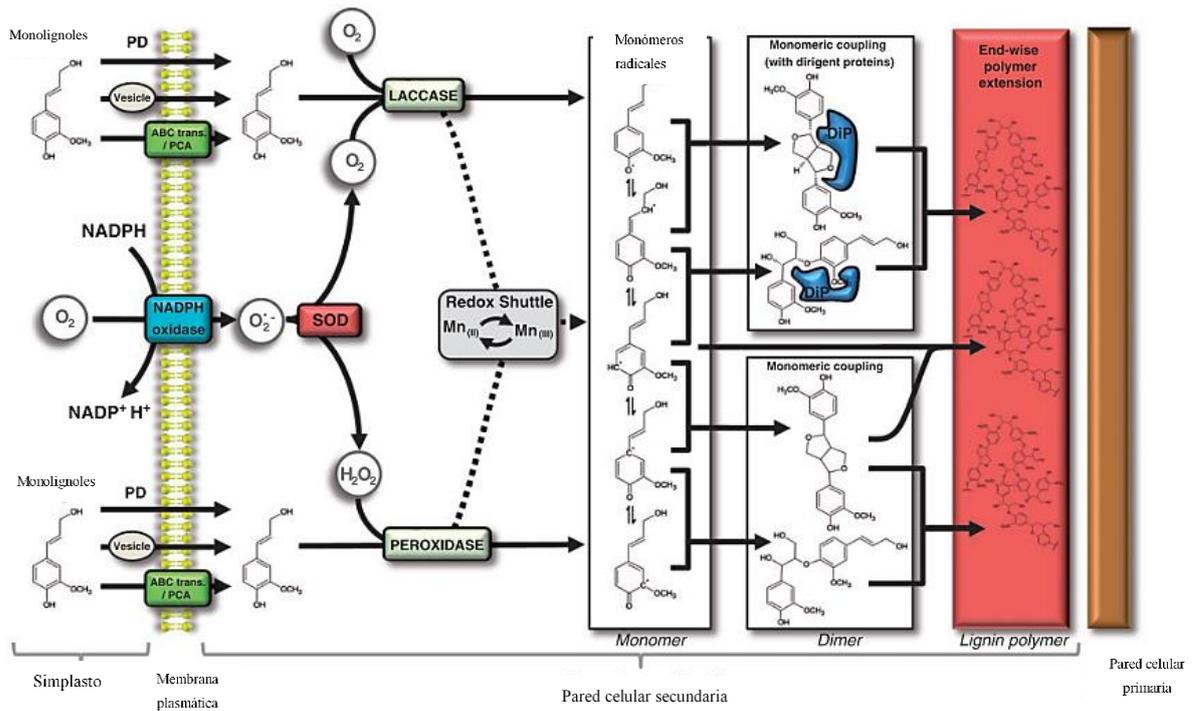


Figura 10: Transporte de monolignoles y polimerización de Lignina

FUENTE: Barros *et al.* (2015)

Para el proceso de oxidación se requieren los agentes oxidantes (O_2 y H_2O_2), cuya producción empieza con la formación de superóxido ($O_2^{\cdot-}$). NADPH oxidasa es una proteína de membrana que cataliza la producción de superóxido a partir de NADPH del citosol (figura 10). Las moléculas de superóxido pasan por un proceso de dismutación (reacción de óxido-reducción) realizada por enzimas superóxido dimutasa (SOD). Los productos de la reacción obtenidos son oxígeno (O_2) o peróxido de hidrógeno (H_2O_2), que son los sustratos oxidantes en el proceso de polimerización de lignina (Barros *et al.*, 2015).

Se cree que el mecanismo de acoplamiento de monolignoles es un proceso sin control bioquímico propiamente dicho, sino de un mecanismo de acoplamiento al azar (Boerjan *et al.*, 2003). Es decir, el tipo de enlace depende de los monolignoles presentes, condiciones físicas y químicas del medio, la estabilidad de los productos; en lugar de un mecanismo de control enzimático (Boerjan *et al.*, 2003; Ralph *et al.*, 2004). Otra teoría señala que este acoplamiento está controlado por proteínas específicas llamadas proteínas dirigidas (DiPs). Estas proteínas son capaces de

controlar la estereoquímica del acoplamiento oxidativo entre monolignoles y así controlar el tipo de enlace que se forma. Esto permite formar de enlaces β -O-4 en lugar β - β , que es termodinámicamente menos favorable (Barros *et al.*, 2015).

2.6. ESPECTROSCOPIA INFRARROJA

2.6.1. Espectro electromagnético

El espectro electromagnético representa el intervalo de todas las frecuencias de energía, como se muestra en la Figura 11 (Christian, 2009; Skoog *et al.*, 2008). En el espectro se representa radiación de baja frecuencia como ondas de radio hasta las de alta frecuencia como rayos X y gamma. El intervalo es tan vasto que se usa la escala logarítmica para representarla (Skoog *et al.*, 2008).

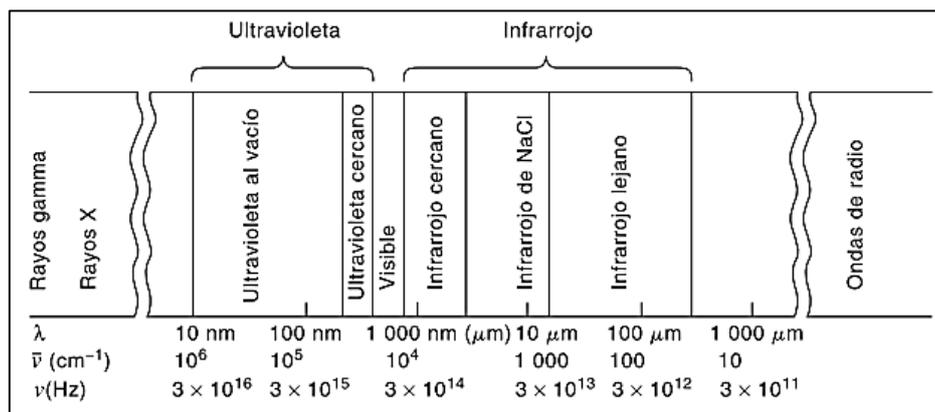


Figura 11: Espectro electromagnético y región infrarroja

FUENTE: Skoog *et al.* (2008)

La radiación electromagnética es una forma de energía de naturaleza dual, es decir viaja en forma de ondas y partículas (Christian, 2009; Skoog *et al.*, 2008). Según el modelo ondulatorio, la radiación electromagnética viaja en forma de ondas transversales (Stuart, 2004). Estas ondas oscilan en un plano perpendicular, donde el campo magnético se propaga en el eje X y el campo eléctrico en el eje Y (Figura 11). La intensidad del campo eléctrico se representa con flechas en el eje Y, la longitud es proporcional a la intensidad del campo (Skoog *et al.*, 2008).

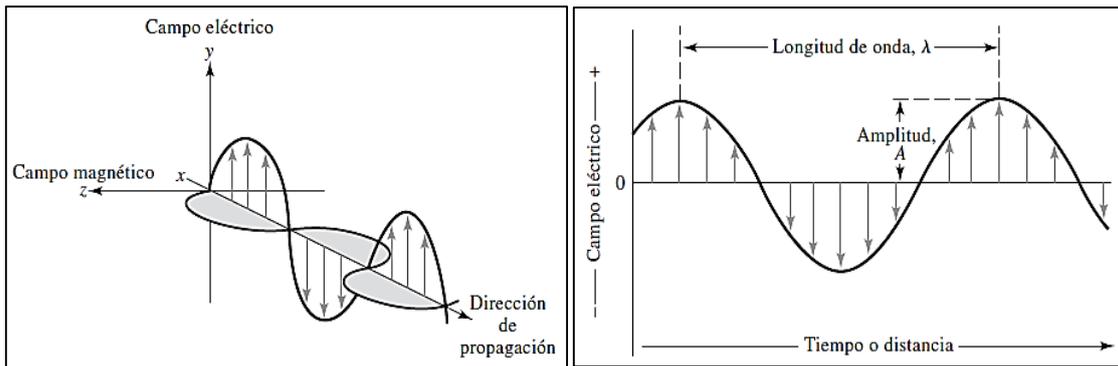


Figura 12: Naturaleza ondulatoria de la radiación electromagnética

El modelo ondulatorio explica varias propiedades de la radiación electromagnética. Por ejemplo, el modelo describe a la onda en función de su longitud, que es la distancia entre los puntos máximos o mínimos de una onda (Skoog *et al.*, 2008). El número de onda es la cantidad de ondas o ciclos que hay en un centímetro (cm^{-1}), se conoce también como la inversa de la longitud de onda, ver figura 12 (Christian, 2009; Skoog *et al.*, 2008). La frecuencia es la cantidad de ondas o ciclos que pasan por un punto fijo en un segundo, la unidad de medida es segundos recíprocos o Hertz (s^{-1} o Hz). Sin embargo, el modelo ondulatorio no describe los fenómenos de absorción o emisión de la radiación (Skoog *et al.*, 2008).

Gracias al desarrollo de la mecánica cuántica se plantea el modelo de partículas que permite explicar los fenómenos de transferencia de energía. Este modelo describe a la radiación electromagnética como una corriente de partículas o paquetes discretos de energía llamados fotones o cuantos. El modelo permitió explicar los fenómenos de absorción y emisión de radiación o transferencia de energía (Skoog *et al.*, 2008). Albert Einstein planteó la relación directa entre la frecuencia de la radiación (ν) y la energía contenida en dicha radiación (E). Como la frecuencia de la radiación y la longitud de onda son recíprocas, si se reemplaza la frecuencia en función de la longitud de onda (λ), se obtiene la siguiente ecuación:

$$E = h \cdot \nu = h \frac{c}{\lambda} \quad (1)$$

Donde E es la energía del fotón (ergios); h , la constante de Planck 6.62×10^{-34} joules-segundo (J-s), ν , la frecuencia de la radiación. Esta ecuación explica que la energía de la radiación electromagnética E está en función de la longitud de onda (λ). Se deduce que la energía es

mayor cuando la longitud de onda es menor o la frecuencia mayor (Christian, 2009; Skoog *et al.*, 2008).

La teoría cuántica también plantea que la absorción o emisión de radiación induce la transición energética de un estado basal (E_0) hacia un estado superior o excitado (E_1). La energía absorbida o emitida es igual a la diferencia entre estos estados y esta diferencia se relaciona con la frecuencia o longitud de onda de la radiación (Skoog *et al.*, 2008).

$$E_1 - E_0 = h \cdot \nu = h \frac{c}{\lambda} \quad (2)$$

Según la fórmula 2 se describe a cada transición energética como discreta o múltiplo de la constante de Planck h . En otras palabras, cada transición energética ocurrirá a un nivel de energía específico y está en función de la longitud de onda de la radiación incidente (Christian, 2009).

La energía absorbida puede causar el movimiento de los electrones alrededor del núcleo del átomo (transición electrónica), además, vibración entre átomos de una molécula (transición vibracional) o rotación de las moléculas alrededor de su centro de masa (transición rotacional) (Skoog *et al.*, 2008). Como se observa en la Figura 13, el requerimiento energético para alcanzar las transiciones rotacionales, vibracional y electrónico va de menor a mayor. Las transiciones rotacionales se alcanzan con radiación de energía baja (longitud de onda alta) en las regiones de microondas e infrarrojo lejano (Figura 13 - A). A mayor energía (menor longitud de onda) se alcanza niveles vibracionales además de los rotacionales, como sucede en la radiación infrarrojo medio y cercano (Figura 13 - B). Una transición electrónica tiene muchos niveles vibracionales de energía, a su vez, cada nivel vibracional tiene muchos niveles rotacionales. Cada transición energética esta cuantizada (discretas). Como resultado una molécula es capaz de absorber más de un haz de luz con diferente longitud de onda y el espectro arrojará numerosos picos que corresponden a la absorción de energía de la molécula a diferentes longitudes de onda (Christian, 2009).

2.6.1. Principio de espectroscopia infrarroja media

El espectro infrarrojo medio es una región del espectro infrarrojo (Stuart, 2004), el cual además comprende las regiones de infrarrojo cercano (NIR) y lejano (Figura 11). La región de infrarrojo lejano va de 200 a 10 cm^{-1} (50 000 – 1 000 000 nm) mientras que la región de infrarrojo cercano va de 12 500 – 4 000 cm^{-1} (780 – 2500 nm) (Sun, 2009, Skoog et al., 2008). Infrarrojo medio (MIR) es la principal zona de la espectroscopía vibracional y va de 4000 a 200 cm^{-1} (2500 – 50 000 nm) (Skoog *et al.*, 2008). MIR provee espectros con información abundante que permite caracterización estructural de moléculas. Esta región es robusta y reproducible, en consecuencia puede medir pequeñas diferencias en la composición de las muestras (Subramanian y Rodriguez-Saona, 2009).

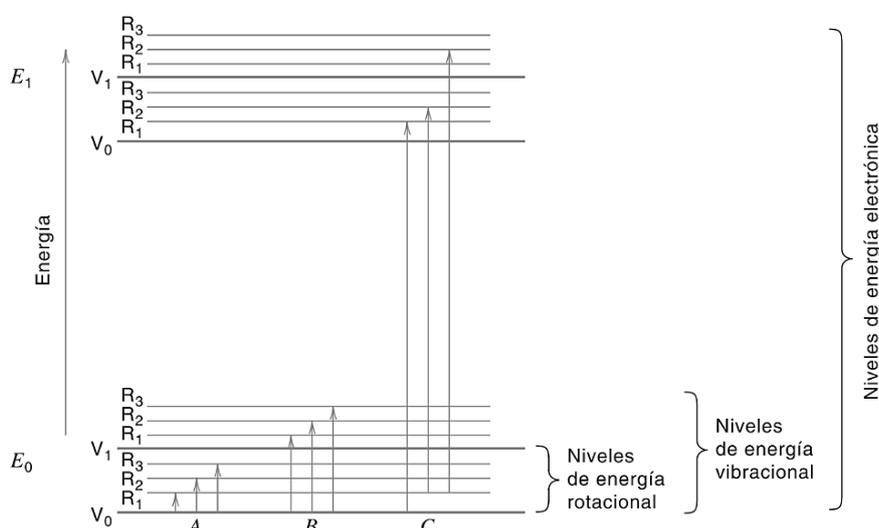


Figura 13: Diagrama de niveles de energía

FUENTE: Christian (2009).

La espectroscopía infrarroja media proporciona información de estructuras químicas por vibración de las moléculas, que a su vez producen un perfil químico de la muestra. Cada enlace o grupo funcional en la molécula absorbe radiación en una o más regiones del espectro infrarrojo (Subramanian y Rodriguez-Saona, 2009). En el espectro hay bandas de absorción a frecuencias (número de onda) que corresponden a grupos funcionales en la molécula, como se muestra en la Figura 14 (Coates, 2006). Toda la información del espectro es tan rica que ayuda al análisis de la composición y determina la estructura de las moléculas (Subramanian y Rodriguez-Saona, 2009). Así se obtiene información de la identidad de moléculas orgánicas como proteínas, polisacáridos y lípidos (Shurvell, 2006).

El momento dipolar es un requerimiento fundamental para la absorción de radiación infrarroja (Coates, 2006). El momento dipolar es la medida de polaridad molecular y es proporcional a la diferencia de cargas y la distancia entre ellas (Skoog et al., 2008, Mc Murry, 2012). Para que una molécula posea momento dipolar, los electrones compartidos por un enlace deben ser fuertemente atraídos por un átomo que por el otro (Mc Murry, 2012). Mientras más alto el momento dipolar, más intensa será la absorción y por lo tanto una banda fuerte se verá en el espectrograma (Skoog *et al.*, 2008). Debido a que la radiación infrarroja cambia el momento dipolar de las moléculas, solo las que poseen momento dipolar pueden absorber en la región infrarroja (Christian, 2009).

La identificación de compuestos orgánicos por espectro infrarrojo se logra en dos etapas. La primera analiza la región de frecuencias de grupo y abarca radiación entre $3\ 600\text{ cm}^{-1}$ y 1250 cm^{-1} aproximadamente (Skoog *et al.*, 2008) Esta región se aplica de forma general a la mayoría de los compuestos y caracteriza a los grupos funcionales (Coates, 2006). La segunda etapa analiza la huella dactilar o la vibración del esqueleto molecular. Abarca una región de 1200 a 600 cm^{-1} y es única para moléculas específicas. La combinación de bandas asociadas a frecuencias de grupo y las frecuencias del esqueleto son usadas para identificar un compuesto, figura 14 (Coates, 2006; Skoog *et al.*, 2008).

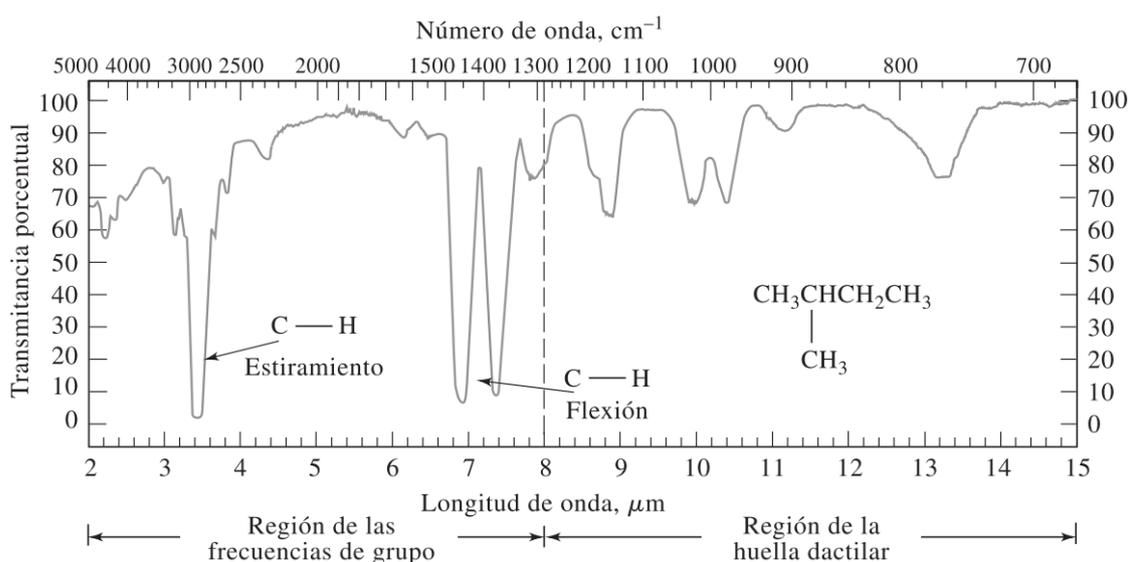


Figura 14: Región de frecuencias de grupo y huella dactilar

FUENTE: Skoog *et al.* (2008)

2.6.2. Ley de Beer

La ley de Beer explica una relación lineal entre la concentración de un analito y su absorbancia. La radiación atraviesa un medio que contiene una especie absorbente (c) en moles, a través de una trayectoria definida (l) que puede estar en centímetros. Se dice que la absorción (A) de la radiación es directamente proporcional a la longitud l y la concentración c de la especie absorbente (Skoog *et al.*, 2008). En la ecuación 3 se observa la absorptividad molar ϵ , que es una constante de proporcionalidad y es característico de cada elemento, además es independiente del estado físico y químico del elemento (Skoog *et al.*, 2008). Es decir, a mayor concentración de la especie en cuestión, mayor absorbancia de radiación. Así mismo, a mayor trayectoria que recorre el haz, mayor absorción de radiación. Si la longitud de la trayectoria l es constante, como sucede en espectroscopia, la absorbancia es directamente proporcional a la concentración de la especie absorbente.

$$A = \epsilon l c \quad (3)$$

Esta relación lineal se usa para cuantificar si la absorbancia es exclusiva de un analito específico. El análisis de muestras complejas mediante espectroscopia infrarroja es la excepción a esa exclusividad y se puede establecer relación lineal entre el espectro y la concentración de un compuesto (Blanco y Alcalà, 2009). Una banda de absorción asignada a un grupo de frecuencia se incrementa proporcionalmente al número de veces que el grupo funcional está presente en la molécula (Coates, 2006). La mayor parte de las desviaciones respecto a la ley de Beer se pueden corregir si se toma en cuenta los factores que causaron la no linealidad (desviación) y conservar la relación lineal de absorbancia y concentración (Christian, 2009).

2.6.3. Movimientos vibracionales

Las moléculas tienen cierta cantidad de energía y están en constante movimiento. Sus enlaces se contraen y estiran, los átomos se mueven adelante y hacia atrás. Estos movimientos se conocen como vibraciones. Cuando la molécula absorbe energía infrarroja los enlaces que conectan los átomos se estiran y contraen un poco más (McMurry, 2012). Los movimientos vibracionales son: estiramiento y flexión. En las vibraciones de estiramiento el enlace químico varía su longitud (Skoog *et al.*, 2008). Los enlaces pueden estirarse de forma simétrica y asimétrica (Figura 15 - a). Las vibraciones de flexión cambian el ángulo entre dos enlaces. Si los átomos cambian el ángulo de los enlaces sin salir de su

plano, se observa movimientos tijereteo y balanceo. Si, por el contrario, los átomos salen del plano se da: aleteo y torsión. En la Figura 15 – b se ilustra los movimientos de flexión, donde el signo + indica movimiento desde el plano hacia el lector y el signo – indica el movimiento del plano alejándose del lector. Las moléculas complejas pueden tener todos los tipos de vibración que se muestran en la Figura 15 (Skoog *et al.*, 2008).

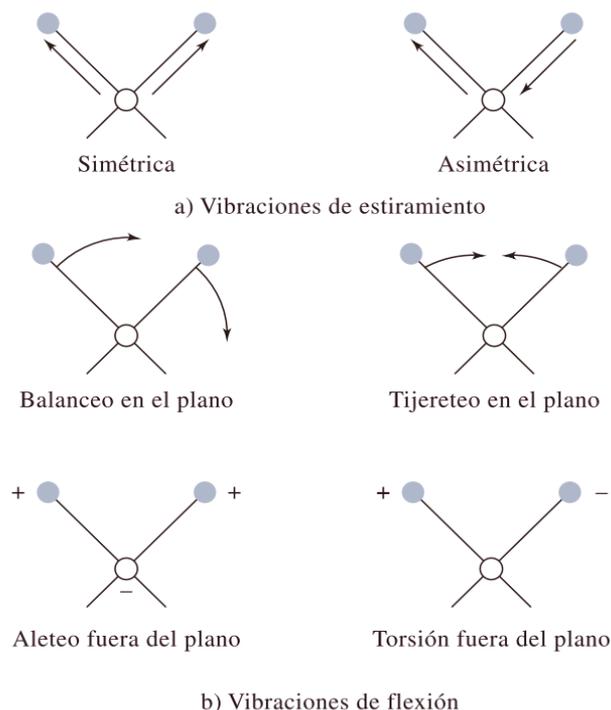


Figura 15: Tipos de vibraciones moleculares

FUENTE: Skoog *et al.* (2008).

Las moléculas complejas pueden tener todos de los movimientos vibracionales, lo cual dificulta la interpretación de los espectros. Se plantea un modelo que reduce al mínimo el número de vibraciones fundamentales y toma como referencia tres ejes coordenados (Coates, 2006). Una molécula tienen N átomos, cada átomo puede ubicarse por tres coordenadas: X, Y y Z, en consecuencia la molécula tiene 3N grados de libertad o vibraciones fundamentales o modos de vibración (Sun, 2009). Los modos normales de vibración pueden ser calculado de acuerdo con el número de moléculas N (ver ecuación 4).

$$\text{Numero normal de modos} = 3N - 6 \quad (4)$$

Por ejemplo, una molécula de metano (5 átomos), según la formula debería poseer 9 set de

vibraciones en el espectro, pero en la práctica se observan menos. Esto se debe a que muchas vibraciones son degeneradas, es decir poseen similar frecuencia de vibración (Coates, 2006).

2.6.4. Frecuencias de grupo

A temperatura ambiente la mayoría de las moléculas se encuentra en estado energético vibracional fundamental (Pasquini, 2003). Si la radiación incidente tiene la misma frecuencia que la vibración natural **de la molécula** se absorbe la radiación (nivel de energía superior) (Stuart, 2004). A esa frecuencia se le conoce como frecuencia fundamental de absorción o frecuencia de grupo y es específica de enlaces químicos y grupos funcionales (Coates, 2006). En el espectro hay bandas de absorción a frecuencias que corresponden a grupos funcionales en la molécula. Por lo que la frecuencia fundamental de absorción es un componente clave para interpretar las estructuras moleculares en espectroscopia infrarroja (Coates, 2006).

La frecuencia de la vibración fundamental depende de la fuerza de enlaces químicos y la masa de cada átomo (Pasquini, 2003).. Existe un modelo armónico que permite asignar la frecuencia vibracional a grupos de átomos o enlaces en función de esas variables (Coates, 2006). La ley de Hooke relaciona la frecuencia vibracional de una molécula con la fuerza del enlace covalente y las masas de los átomos que forman el enlace mediante un simple calculo (ecuación 5).

$$v(cm^{-1}) = 130.3 \frac{\sqrt{k}}{\mu} \quad (5)$$

La ecuación 5 toma en cuenta una constante numérica (k) constante de fuerza del enlace covalente y (u) masa reducida ($m_1m_2/(m_1+m_2)$). Donde m_1 y m_2 son las masas de los átomos que forman el enlace químico (Shurvell, 2006). Esta ecuación estima la frecuencia de vibración de enlace entre un par de átomos por lo tanto predice en qué región del espectro encontrarla (cm^{-1}) (Shurvell, 2006).

En la Tabla 3 se observan las frecuencias de grupo calculadas para algunos compuestos diatómicos mediante la fórmula 5. Las frecuencias de grupo predicen en que región del espectro se espera bandas de absorción. Las frecuencias de grupos funcionales se caracteriza porque la fuerza de estiramiento en los grupos funcionales es casi constante en diferentes

moléculas (Shurvell, 2006).

Tabla 3: Frecuencias calculadas de moléculas diatómicas

Group	Reduced mass (amu)	Force constant (N m ⁻¹)	Frequency (cm ⁻¹)
O-H	0.94	700	3600
N-H	0.93	600	3300
C-H	0.92	500	3000
C-C	6.00	425	1100
C=C	6.00	960	1650
C=O	6.86	1200	1725
C≡C	6.00	1600	2100
C≡N	6.46	2100	2350

FUENTE: Shurvell (2006)

Sin embargo, el modelo armónico es ideal y no se ajusta a la mecánica real de los átomos (Coates, 2006). Se toma un de modelo anarmónico para explicar las desviaciones del modelo ideal. En este modelo las moléculas se representan como dos esferas conectadas con un resorte (Pasquini, 2003). El modelo considera repulsión y atracción de las nubes electrónicas y disociación de enlaces a niveles altos de energía (Pasquini, 2003). Este modelo aproxima mejor los cálculos de vibraciones de alta frecuencia (Coates, 2006).

Además, en moléculas poliatómicas no solo se encuentran vibraciones de estiramiento que pueden ser calculadas por la ecuación 5, también se dan vibraciones de flexión (Shurvell, 2006). Los movimientos como estiramiento y flexión contribuyen al espectro de absorción ya sean simétricas o asimétricas. Debido a que la energía y la frecuencia están proporcionalmente relacionadas, se deduce que a frecuencias más altas ocurre absorción de estiramiento de un enlace químico y a frecuencias más bajas absorción de flexión. Por lo tanto, la energía necesaria para estirar un enlace es mayor que para flexionarlo (Coates, 2006). Así mismo, se requiere más energía para excitar una vibración asimétrica que una vibración simétrica. A frecuencias altas se alcanza vibraciones asimétricas; a bajas frecuencias, vibraciones simétricas (Coates, 2006). Un buen ejemplo es el set de vibraciones de una molécula de agua, como se observa en la Figura 16. La frecuencia de vibración de estiramiento asimétrico (3756 cm⁻¹) es mayor que la frecuencia de estiramiento simétrico (3657 cm⁻¹). La frecuencia de flexión (1595 cm⁻¹) es menor que las frecuencias de estiramientos (Sun, 2009).

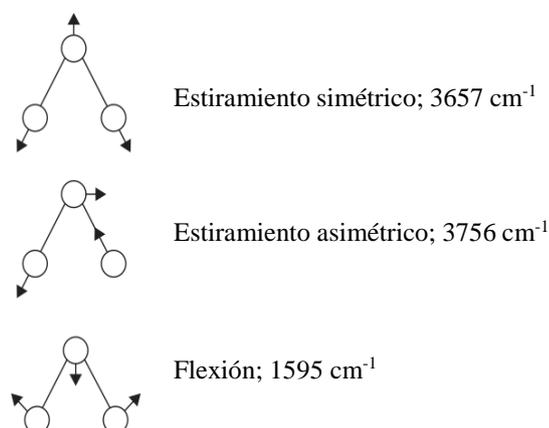


Figura 16: Modos de vibración de una molécula de agua

FUENTE: Sun (2009)

Las frecuencias de moléculas diatómicas mostradas en la Tabla 4 varían cuando estos átomos son parte de una molécula. Las vibraciones de un grupo diatómico causan el desplazamiento de otros átomos en la molécula. El desplazamiento de estos átomos afecta su vibración (Shurvell, 2006). Este fenómeno se conoce como acoplamiento de vibración. Por ejemplo, la frecuencia de estiramiento de C-O es 1034 cm^{-1} en metano. Al interactuar con otros enlaces como C-H y C-C las vibraciones se acoplan y resultan en variaciones de la frecuencia de C-O (1053 y 1105 cm^{-1} en etanol y butanol respectivamente) (Skoog *et al.*, 2008).

Tabla 4: Frecuencias de grupos funcionales de compuestos orgánicos

Enlace	Tipo de compuesto	Frecuencias, cm^{-1}	Intensidad
C—H	Alcanos	2850–2970	Fuerte
		1340–1470	Fuerte
C—H	Alquenos (>C=C<H)	3010–3095	Media
		675–995	Fuerte
C—H	Alquinos ($\text{—C}\equiv\text{C—H}$)	3300	Fuerte
C—H	Anillos aromáticos	3010–3100	Media
		690–900	Fuerte
O—H	Alcoholes monoméricos, fenoles	3590–3650	Variable
	Alcoholes con puentes de hidrógeno, fenoles	3200–3600	Variable, a veces amplia
	Ácidos carboxílicos monoméricos	3500–3650	Media
	Ácidos carboxílicos con puentes de hidrógeno	2500–2700	Amplia
N—H	Aminas, amidas	3300–3500	Media
C=C	Alquenos	1610–1680	Variable
	Anillos aromáticos	1500–1600	Variable
C≡C	Alquinos	2100–2260	Variable
C—N	Aminas, amidas	1180–1360	Fuerte
C≡N	Nitrilos	2210–2280	Fuerte
C—O	Alcoholes, éteres, ácidos carboxílicos, ésteres	1050–1300	Fuerte
C=O	Aldehídos, cetonas, ácidos carboxílicos, ésteres	1690–1760	Fuerte
NO ₂	Compuestos nitro	1500–1570	Fuerte
		1300–1370	Fuerte

FUENTE: Skoog *et al.* (2008).

Las frecuencias de grupo varían por factores como simetría, acoplamiento molecular y debido a las interacciones con vibraciones de los átomos que forman una molécula. Sin embargo, las variaciones suelen ser pequeñas, por eso se asigna un intervalo de frecuencias dentro del cual es probable encontrar bandas del grupo funcional determinado. En la Tabla 4 se señalan las frecuencias de grupo de los compuestos orgánicos más importantes (Skoog *et al.*, 2008).

2.6.5. Huella dactilar

La región comprendida entre 1800 y 600 cm^{-1} es conocida como huella dactilar. Esta región permite la identificación de compuesto por su carácter único para cada molécula. En una molécula la vibración de sus enlaces sencillos origina bandas de absorción y se produce interacciones fuertes entre enlaces vecinos (Kelly *et al.*, 2011; Skoog *et al.*, 2008). Las bandas de absorción que se observan conforma un patrón producto de la combinación de estas interacciones y depende de la estructura completa de la molécula (Skoog *et al.*, 2008; Stuart, 2004). Esta región permite identificar moléculas por comparación con el espectro de estándares conocidos o librerías (Peak, 2005). En esta zona es conocida por absorber diferentes compuestos bioquímicos que arrojan picos de absorción para: amidas I ($\sim 1650 \text{ cm}^{-1}$), amidas II ($\sim 1550 \text{ cm}^{-1}$), amida III ($\sim 1260 \text{ cm}^{-1}$), carbohidratos ($\sim 1155 \text{ cm}^{-1}$), glicógeno ($\sim 1030 \text{ cm}^{-1}$), lípidos ($\sim 1750 \text{ cm}^{-1}$), proteínas fosforiladas ($\sim 970 \text{ cm}^{-1}$) y otros (Kelly *et al.*, 2011).

2.6.6. Espectogramas de polimeros de madera

En el rango de 3800 a 2800 cm^{-1} se encuentran dos áreas características que corresponden a vibración de enlaces presentes en todos los componentes de madera. El primer pico se observa entre 3800 y 3000 cm^{-1} y se asigna a estiramiento del grupo -OH de celulosa intra e intermolecular, agua, y de O-H libres (Popescu *et al.*, 2007). El segundo pico aparece entre 3000 y 2700 cm^{-1} y puede presentar dos picos que corresponden al estiramiento simétrico (2920 cm^{-1}) y asimétrico (2850 cm^{-1}) de enlaces C-H de metil (CH_3), metileno (CH_2) y de grupos metoxilo de anillos aromáticos (Popescu *et al.*, 2007; Popescu *et al.*, 2009).

Por otro lado, la huella dactilar, comúnmente usada en análisis multivariado, es la región más rica y compleja en información. Esta región contiene picos característicos de los principales componentes de maderas: celulosa, hemicelulosa y lignina (Chen *et al.*, 2010;

Popescu *et al.*, 2009). Por ejemplo, el pico 1739 cm^{-1} corresponde al estiramiento de grupos carbonilo (C=O) en ketonas no conjugadas, carbonil y grupos alifáticos presentes en todos los componentes de maderas, principalmente en hemicelulosa (Chen *et al.*, 2010; Rana *et al.*, 2010; Zhang *et al.*, 2016). Los picos en 1610-1595 cm^{-1} (de anillo aromático unido a C=O), 1510-1505 cm^{-1} (enlaces C=C de anillos aromáticos) y 1270-1268 cm^{-1} (anillo guaiacil presentes) corresponden a vibraciones en lignina (Chen *et al.*, 2010; Kubo y Kadla, 2005; Popescu *et al.*, 2007). Los picos $\approx 1317 \text{ cm}^{-1}$ (flexión CH₂), $\approx 1154 \text{ cm}^{-1}$ (estiramiento asimétrico C-O-C en anillos de piranosa) y $\approx 896 \text{ cm}^{-1}$ (la vibración del anillo de glucosa y enlaces β) evidencian vibraciones en celulosa (Chen *et al.*, 2010; Karunakaran *et al.*, 2015; Pandey, 1999; Rana *et al.*, 2010).

La región entre 1200 cm^{-1} y 900 cm^{-1} presenta múltiples picos característicos de carbohidratos (C. Zhou *et al.*, 2015). Los picos 1035-1025 cm^{-1} y 1125-1105 cm^{-1} corresponde a estiramiento de enlace C-O y el pico 1380-1365 cm^{-1} aparece por la flexión C-H en celulosa y hemicelulosa (Popescu *et al.*, 2007; Zhang *et al.*, 2016). Sin embargo, estos picos tienen contribución de enlaces presentes en lignina, como se observa en la Tabla 5 (Rana *et al.*, 2010; Zhang *et al.*, 2016).

La región con los picos 1235-1230 cm^{-1} , 1470-1455 cm^{-1} y 1430-1422 cm^{-1} corresponde a vibración de enlaces en lignina y hemicelulosa. El pico en 1650-1640 cm^{-1} corresponde a vibración de estiramiento de C=O conjugado a anillos aromáticos en lignina y estiramiento de O-H y C=O conjugado en polisacáridos (Chen *et al.*, 2010; Traoré *et al.*, 2018). Por otro lado, la presencia de ceras, grasas y ácidos risínicos contribuye al pico de 1730 cm^{-1} y que corresponde al grupo C=O del grupo ester carbonil. El pico de 1600 cm^{-1} es de C=C estiramiento en la deformación del anillo aromático. (C. Zhou *et al.*, 2015).

Tabla 5: Picos FTIR de madera y los enlaces correspondientes

Número de onda (cm-1)	Vibración	Detalles	Referencia
3800- 3000	Estiramiento O-H	Lignina+celulosa+hemicelulosa: Estiramiento de enlaces OH en agua, en celulosa intra e intermolecular y O-H libres.	Popescu et al., 2007
2938-2920	Estiramiento C-H	Lignina+celulosa+hemicelulosa: estiramiento simétrico de CH en metocil de grupos aromaticos, grupos metil y metileno de cadenas laterales.	Popescu et al., 2007
2840-2835	Estiramiento C-H	Lignina y hemicelulosa: estiramiento asimétrico de CH en metocil de grupos aromaticos, grupos metil y metileno de cadenas laterales.	Popescu et al., 2007
1740-1720	Estiramiento C=O	Hemicelulosa: Estiramiento C=O de en ketonas no conjugadas, carbonil y grupos alifaticos Extractivos: Estiramiento C=O de ester carbonil.	Chen et al., 2010, Zhang et al., 2016
1650-1640	Estiramiento C=O y O-H	Lignina: Estiramiento de C=O conjugando a un anillo aromatico Celulosa: estiramiento de O-H y C=O conjugado en celulosa. Agua asociada a lignina y celulosa.	Chen et al., 2010, Zhou et al., 2015, Rana et al., 2010
1610-1590	vibración de esqueleto aromatico + estiramiento C=O	Lignina: Vibración del anillo aromático y estiramiento C=O, Estiramiento C=C de anillos en Lignina. Extractivos: Deformación C=C de anillo aromáticos	Popescu et al., 2007, Zhou et al., 2015
1515-1505	Vibración de anillo aromático	Lignina: Vibración del anillo aromatico, Vibración de C=C de anillos aromaticos. Extractivos: vibración de anillos benceno	Chen et al., 2010, Zhou et al., 2015, Zhang et al., 2016
1470-1455	Flexión C-H	Hemicelulosa: Tijereteo de CH2 en -CH3 y -CH2 de anillos de piranosa. Lignina: Flexión C-H de metil (-CH3) y metileno (CH2)	Chen et al., 2010, Zhang et al., 2016, Traoré et al., 2018
1430-1422	Flexión C-H	Celulosa: Deformación de C-H Lignina: Deformación en plano de C-H y vibración de anillo aromatico Hemicelulosa: Flexión de C-H y O-H.	Chen et al., 2010, Zhou et al., 2015
1380 - 1365	Flexión C-H	Celulosa y Hemicelulosa: Flexión de C-H	Popescu et al., 2007, Zhou et al., 2015
1335-1320	Flexión C-H y vibración C-O	Lignina: vibración de C-O del anillo siringil Hemicelulosa: Flexión de aleteo de CH2, Estiramiento C-O de grupo acetato Celulosa: Flexión en plano de C-H	Popescu et al., 2007, Zhou et al., 2015, Chen et al., 2010
1320-1310	Aleteo CH2	Celulosa: Flexión de aleteo CH2	Popescu et al., 2007
1270-1268	Estiramiento C=O y anillo aromático	Lignina: Vibración del grupo carbonilo (C=O) en anillo Guaiacil. Vibración de C-O del anillo Guaiacil.	Chen et al., 2010, Zhang et al., 2016
1235-1230	Estiramiento C-O	Lignina: Vibración de anillos Siringil Hemicelulosa: Estiramiento de C-O en Xilano	Popescu et al., 2007
1162-1125	Estiramiento asimétrico C-O-C Flexión C-H	Celulosa: Estiramiento asimétrico C-O-C en anillos de piranosa Lignina: Flexión en plano de enlaces C-H en anillos Guaiacil.	Popescu et al., 2007, Zhou et al., 2015
1125-1105	Flexión C-H	Lignina: Flexión en plano de C-H de anillo aromático Siringil Celulosa y Hemicelulosa: Vibración de anillos glucosa	Popescu et al., 2007, Rana et al., 2010
1035-1025	Estiramiento C-O flexión C-H	Celulosa y Hemicelulosa: estiramiento de C-O en anillos de carbohidratos Lignina: flexión en plano de C-H y C-O en alcoholes primarios.	Zhou et al., 2015, Rana et al., 2010, Chen et al., 2010
896	Flexión C-H	Celulosa: anillo de glucosa	Chen et al., 2010, Zhang et al., 2016

FUENTE: Elaboración propia.

2.6.7. Espectroscopía infrarroja por transformada de Fourier (FTIR)

La espectroscopia FTIR es considerada una de las técnicas más poderosas para el análisis químico y sus aplicaciones en el campo de la biología se ha incrementado debido a su versatilidad, sensibilidad, simplicidad y velocidad de análisis (Subramanian y Rodriguez-Saona, 2009). La identificación de compuestos por medios químicos tradicionales, donde se determina las propiedades físicas, y químicas son poco prácticos, ya que requieren gran cantidad de sustancia, y no son eficientes para compuestos nuevos o de gran complejidad (L. G. Wade, 2004b). En contraste con estas técnicas, FTIR requiere cantidades pequeñas de materia, no es destructiva y provee información abundante de grupos funcionales (Alonso-Simón *et al.*, 2011). Además, el tamaño de los equipos y el costo se han reducido en los años hasta convertirse en equipos de mesa de fácil mantenimiento y adquisición (Skoog *et al.*, 2008).

Los equipos FTIR presentan múltiples ventajas en comparación con los equipos IR dispersivos convencionales. La ventaja principal es la de Jacquinot y consiste en la alta capacidad de interacción entre la radiación y la materia. En un equipo FTIR la muestra interacciona con todas las longitudes de onda a la vez y no con pequeñas regiones. Esto mejora la relación señal/ruido e incrementa el rendimiento, como resultado se obtiene espectrogramas con alta resolución (Christian, 2009). Otra ventaja importante es la de Fellgett o múltiplex, la cual radica en que los equipos FTIR pueden medir múltiples longitudes de onda simultáneamente. Como resultados la colecta de datos requiere menos tiempo de colecta a diferencia de otros espectrómetros (Subramanian y Rodriguez-Saona, 2009).

Una innovación importante en los equipos FTIR (espectros infrarrojos por transformada de Fourier) es el uso de interferómetros para obtener espectros en lugar de los monocromadores de rejillas usados en los instrumentos dispersivos convencionales (Christian, 2009). Debido al uso de interferómetros se mejoró la relación señal-ruido y los límites de detección en comparación con instrumentos dispersivos. Así mismo, los interferómetros permitieron una alta resolución (0.1 cm^{-1}) de los espectros y frecuencia reproducible. Además de la cualificación de compuestos también permite su cuantificación (Skoog *et al.*, 2008).

Los equipos FTIR usan la transformada de Fourier para convertir los interferogramas (datos

obtenidos del interferómetro) a espectrogramas, los cuales son más sencillos de interpretar. La señal generada por los interferómetros está en función del desplazamiento de dos haces para dar patrones de interferencia (de todas las longitudes de onda en la radiación). La interferencia de radiación entre estos haces genera un interferograma (Stuart, 2004). La transformada de Fourier trabaja con los datos (Skoog *et al.*, 2008) y transforma la señal producida por la distancia entre estos haces a frecuencias (Stuart, 2004).

La capacidad de coleccionar múltiples componentes y la ley de Beer hace al FTIR una herramienta poderosa de cuantificación (van de Voort, 1992). De la mano la quimiometría, el FTIR es usado para cuantificación de material forestal (Acquah *et al.*, 2016). Con técnicas quimiométricas como PLS, el FTIR es capaz de cuantificar componentes químicos en madera (C. Zhou *et al.*, 2015; Chen *et al.*, 2010).

2.6.7.1. Interferómetro de Michelson

El interferómetro de Michelson es un dispositivo usado para modular la radiación óptica y generar de patrones de interferencia o interferogramas de forma controlada (Skoog *et al.*, 2008; Subramanian y Rodriguez-Saona, 2009). Este interferómetro es la base de la mayoría de instrumentos infrarrojo con transformada de Fourier FTIR (Skoog *et al.*, 2008). En la actualidad existen otros interferómetros, pero los diseños más modernos y usados están basados en el de Michelson. El interferómetro parte un haz de luz en dos rayos de intensidad similar, desvía su trayectoria y luego los recombina de tal forma que la intensidad del haz resultante varíe en función del desplazamiento de los dos rayos (Skoog *et al.*, 2008).

El interferómetro está compuesto por una fuente emisora de luz infrarroja, un divisor de rayos y dos espejos, uno fijo y uno móvil (Figura 17). El divisor de rayos divide en dos el haz infrarrojo proveniente de la fuente. El primer haz se refleja en el espejo fijo y regresa al divisor de haz. El segundo haz viaja hacia al espejo móvil el cual se desplaza (se aleja y se acerca del divisor de rayos), así se modifica la distancia de viaje del segundo haz y retorna al divisor de haz con un desfase (ligeramente desplazado). Los haces reflejados se encuentran y se combinan en el divisor de rayos, luego viajan hacia el detector a través de la muestra (Subramanian y Rodriguez-Saona, 2009).

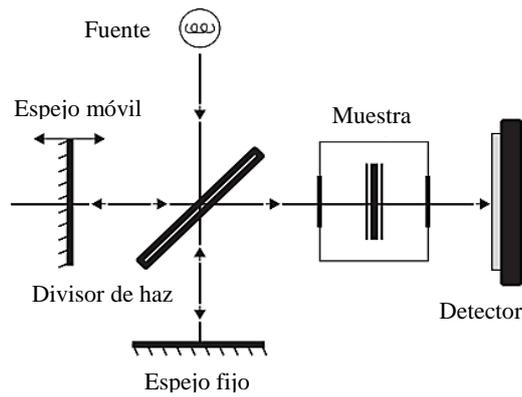


Figura 17: Estructura del Interferómetro de Michelson

FUENTE: Subramanian & Rodriguez-Saona (2009).

El retardo óptico es la diferencia en la distancia de viaje entre los dos haces de luz en centímetros, creada por el movimiento del espejo móvil (Subramanian y Rodriguez-Saona, 2009). Si los dos espejos están a la misma distancia del divisor de rayos el retardo óptico es cero y los haces están en fase por lo tanto ocurre interferencia constructiva. Durante la interferencia constructiva la intensidad del haz resultante es mayor y la potencia de la señal está en su máximo. En contraste, si los haces están fuera de fase la interferencia es destructiva por lo que la intensidad del haz resultante es menor. A otros valores del retardo óptico tanto interferencia positiva y negativa ocurren y resulta en variación de la intensidad de la luz en forma de una curva coseno (Subramanian y Rodriguez-Saona, 2009).

Un interferograma contiene la variación de intensidad de luz (voltios) generada por diferentes retardos ópticos versus el retardo óptico (cm), como se muestra en la Figura 18 (Subramanian y Rodriguez-Saona, 2009). Se interpreta al interferograma como el movimiento del espejo móvil hacia adelante hasta el punto donde el retardo óptico es cero (centro de exposición) y hacia atrás hasta su posición inicial. El centro de explosión es el de más alta intensidad y retardo óptico cero, donde las ondas tienen interferencia constructiva. Este punto no contiene información de la muestra, pero sí información del total de energía proveniente de la fuente. Las regiones laterales, donde se dan interferencias constructivas y destructivas en diversos niveles, se llaman alas del interferograma. Idealmente las alas contienen información de la muestra y deberían ser idénticas a ambos lados del punto de explosión (Subramanian y Rodriguez-Saona, 2009).

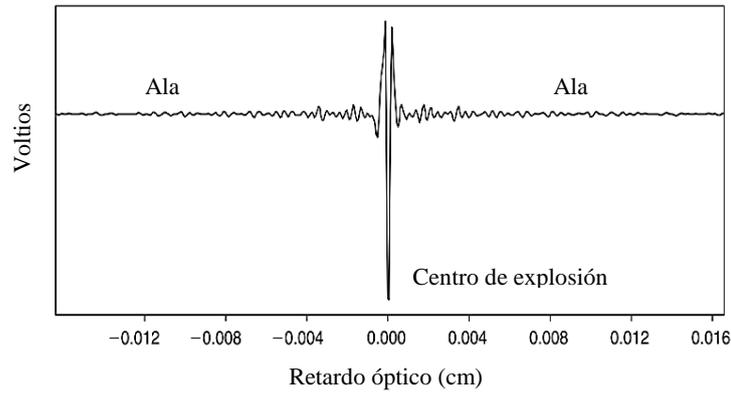


Figura 18: Esquema de un Interferograma

FUENTE: Subramanian & Rodriguez-Saona (2009)

2.6.7.2. Transformada de Fourier

La transformada de Fourier permite convertir un patrón de interferencia (Figura 18) a un espectrograma (Subramanian y Rodriguez-Saona, 2009). Los espectrogramas son más sencillos de interpretar visualmente y son los más usados. Específicamente, la transformada de Fourier convierte la intensidad que alcanza al detector ($I(\delta)$) en densidad de la potencia espectral a una longitud de onda específica ($B(\bar{\nu})$) (Stuart, 2004)..

$$I(\delta) = \int_0^{+\infty} B(\bar{\nu}) \cos(2\pi \bar{\nu} \delta) d\bar{\nu} \quad (6)$$

$$B(\bar{\nu}) = \int_{-\infty}^{+\infty} I(\delta) \cos(2\pi \bar{\nu} \delta) d\delta \quad (7)$$

La ecuación 6 representa a los interferogramas, donde la intensidad de la absorción varía en función del retardo óptico (δ). Por otro lado, la ecuación 7 muestra la variación de la intensidad en función de la longitud de onda ($\bar{\nu}$). Ambas ecuaciones son interconvertibles entre sí y se conocen como par de transformada de Fourier (Stuart, 2004).

2.6.7.3. ATR

La reflectancia total atenuada (ATR, attenuated total reflectance) es uno de los métodos de muestreo más comúnmente usado para obtener espectros IR (Subramanian y Rodriguez-Saona, 2009). Este método consiste en comprimir la muestra sobre un cristal de diamante y la radiación infrarroja penetra la muestra, reflejándose y se dirigiéndose al detector. Además del diamante, se pueden usar cristales de zinc selenoide, germanio y talio ioide. Gracias al

ATR se evita el uso de celdas, placas de sal y simplifica el manejo de muestras (Christian, 2009).

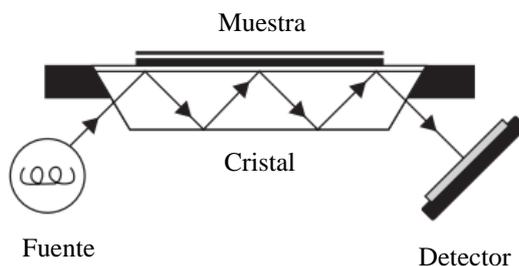


Figura 19: Diagrama del muestreo por ATR

FUENTE: Subramanian & Rodriguez-Saona (2009)

La reflexión interna total es el fenómeno que refleja toda la luz que incide dentro de una superficie del cristal (Figura 19). Para que ocurra este fenómeno el cristal debe poseer un alto índice de refracción y la luz debe incidir en un ángulo particular. Este fenómeno ocurre si un haz de luz incide sobre el cristal y la totalidad se refleja en forma de ondas evanescentes, las cuales se extienden a pequeñas distancias encima del cristal. En este punto, la intensidad de la luz reflejada se reduce y a este fenómeno se conoce como reflectancia total atenuada (Subramanian y Rodriguez-Saona, 2009). Cuando la muestra entra en contacto con el cristal, la luz total reflejada penetra la muestra y pierde energía en las longitudes de onda que la muestra absorbe (Stuart, 2004).

A pesar de la utilidad del ATR presenta algunos inconvenientes que pueden ser manejados por pretratamiento de los espectros. Por ejemplo, el estado físico de las muestras (sólido o líquido) y tamaño de partículas pueden afectar la intensidad del ruido. Otra desventaja de ATR es la cantidad de muestra analizada ya que la absorbancia de un pico esta proporcional a la cantidad de una sustancia en la muestra (Lee *et al.*, 2017).

2.7. QUIMIOMETRÍA

La definición de Quimiometría según “Chemometrics and Intelligent Laboratory System”, revista líder en el área, es la disciplina química que usa matemática, estadística y otros métodos para diseñar experimentos y procedimientos de medida óptimos, con la finalidad de proveer la mayor información relevante de los datos químicos (Héberger, 2008). Esta disciplina es particularmente útil en los datos obtenidos por instrumentos químicos como

espectrofotómetros, los cuales arrojan múltiples factores por muestra. La quimiometría usa métodos multivariados de tal manera que se consideran los múltiples factores al mismo tiempo (Héberger, 2008). Como resultado estos datos se explican mejor mediante quimiometría que con la estadística convencional, la cual analiza un factor a la vez (Geladi, 2003).

En la Figura 20 se observa una matriz de datos con múltiples variables (K) por cada muestra analizada (I). La idea de la quimiometría es reducir sustancialmente la información redundante obtenida por un instrumento analítico a información relevante o variables latentes (A), las cuales son más fáciles de entender y más estables. La información restante es irrelevante o contiene ruido y se conoce como residuales (Geladi, 2003). En particular, las variables arrojadas por los instrumentos analíticos pueden expresarse en: absorbancia, altura de pico, concentración, conteo de partículas, etc.

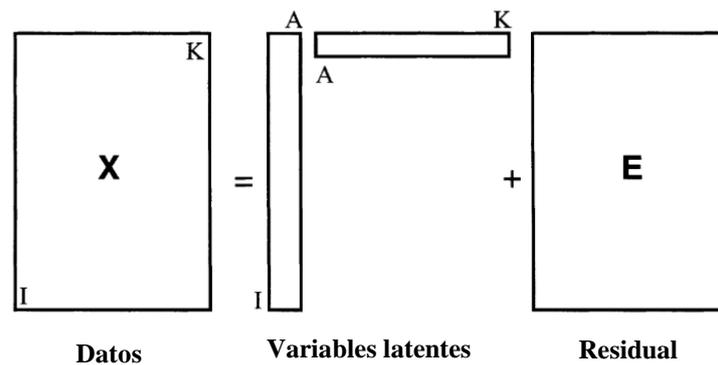


Figura 20: Matriz de datos y su reducción a variables latentes por quimiometría

FUENTE: Geladi (2003)

De forma general, los datos están organizados en una matriz con K factores de I muestras (Geladi, 2003). Ésta matriz se conoce como matriz **X** y contiene variables independientes. Además, en ciertos análisis se tiene una segunda matriz de datos con variables dependientes, conocida como matriz **Y**. Los datos de esta matriz pueden tener naturaleza categórica (estaciones, tipos de cáncer, etc.) o numérica (variables cuantitativas como temperatura) (Héberger, 2008).

En función de los datos que contengan los análisis, los métodos quimiométricos se clasifican en supervisados y no supervisados. Los métodos no supervisados se caracterizan por tener solo la matriz \mathbf{X} de datos (independientes). Estos métodos reconocen patrones en la matriz \mathbf{X} que permite clasificar o agrupar las muestras en subgrupos. El análisis de componentes principales o PCA y análisis de cluster son ejemplos de modelos no supervisados (Héberger, 2008). Por otro lado, los métodos supervisados se caracterizan por disponer de una matriz \mathbf{X} e \mathbf{Y} . Estos métodos encuentran patrones en la matriz \mathbf{X} que se relacionan con los datos de la matriz \mathbf{Y} (Kuhn y Johnson, 2013). Mediante los métodos supervisados se puede crear modelos de predicción para estimar las variables dependientes \mathbf{Y} de muestras nuevas. El método de regresión de mínimos cuadrados parciales (PLS) es usado cuando existe más de una variable \mathbf{Y} en escala numérica. Así mismo, se usan para reconocimiento de patrones y segregación de grupos si la variable \mathbf{Y} es de escala categórica. Los métodos supervisados comunes son: análisis discriminante lineal (LDA) y redes neuronales artificiales (ANN) (Héberger, 2008).

En el caso de espectroscopia infrarroja, la interpretación directa de la composición química de una muestra a través de su espectro infrarrojo es complicada debido a la densa información de los enlaces químicos. Los modelos multivariados, especialmente el modelo supervisado PLS, establece una relación entre las múltiples características espectrales (variables independientes) y alguna propiedad física o composición química de las muestras (variables dependientes). Esto se logra mediante la reducción de los datos espectrales a unas variables latentes, de forma que se pueda construir relación entre esas variables y la composición química (Xu *et al.*, 2013). Por ejemplo, el análisis quimiométrico ha permitido determinar la relación entre el espectro infrarrojo y el contenido de celulosa y hemicelulosa en *Acacia mangium* (Karlinasari *et al.*, 2014), componentes lignocelulósicos en *Miscanthus sinensis* (Jin *et al.*, 2017), celulosa, hemicelulosa y lignina en Bamboo (Li *et al.*, 2015).

2.7.1. Análisis quimiométrico en espectroscopía infrarroja

El análisis quimiométrico usualmente consiste en calibración y validación, como se muestra en la Figura 21 (Xu *et al.*, 2013). La calibración construye el modelo multivariado mediante una relación matemática entre las variables independientes \mathbf{X} (datos del instrumento analítico) y las variables dependientes \mathbf{Y} (datos obtenidos por un método químico de referencia) (Geladi, 2003). Mediante la calibración se calcula los parámetros de una ecuación

matemática (Blanco y Alcalà, 2009). Una vez obtenido la ecuación del modelo de calibración se preparan los datos de validación para evaluar la capacidad predictiva del modelo construido (Kuhn y Johnson, 2013). La validación compara los valores predichos por el modelo de calibración con los valores de referencia (Xu *et al.*, 2013). Los modelos construidos se evalúan por valores estadísticos, los cuales se mencionan en la sección 2.7.4.

La validación puede ser interna o externa (Faber y Rajkó, 2007). En el caso de la validación externa se requiere un grupo de datos para la calibración y los datos restantes, se conocen como datos de validación (Kuhn y Johnson, 2013). El grupo de muestras para validar el modelo debe ser independientes de las muestra usada para su calibración (Faber y Rajkó, 2007). Sin embargo, rara vez el número de datos es lo suficientemente grande para dividirlos en datos de calibración y de validación (Hubert y Branden, 2003).

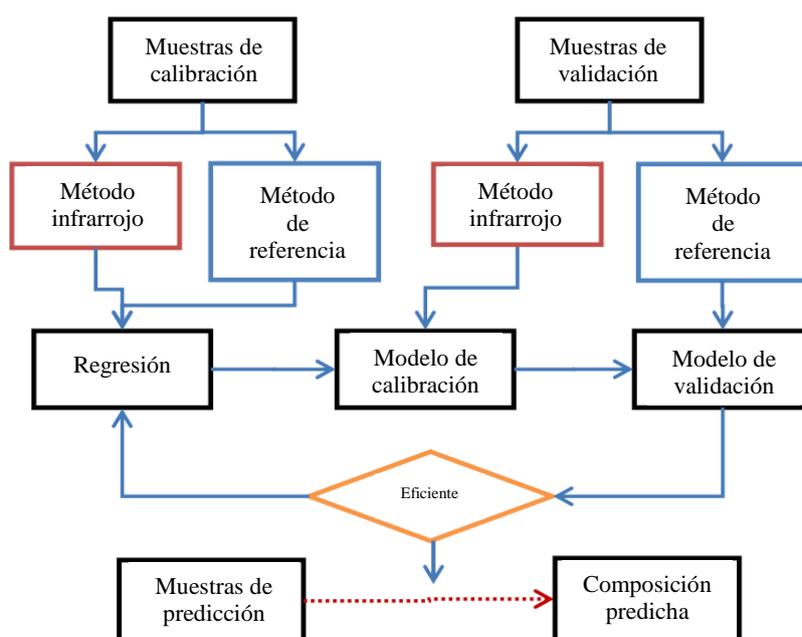


Figura 21: Flujograma de los procesos de análisis quimiométrico

FUENTE: Xu *et al.* (2013).

En casos con pocas muestras disponibles, se construye los modelos mediante validación cruzada (Chen *et al.*, 2010; Xu *et al.*, 2013). La validación cruzada es el modelo de calibración interna más usada y toma los datos para la validación de los datos de calibración, es decir los datos de calibración y validación no son independientes (Faber y Rajkó, 2007). La validación cruzada con un dato fuera (leave one out cross validation LOOCV) excluye

una observación en el momento de la calibración de los datos y el modelo resultante se evalúa (validación) con la observación excluida. Esto se repite para todas las observaciones y el desempeño promedio de todas las repeticiones se considera como el modelo multivariado. Otra técnica popular es la validación cruzada K- veces (K-fold cross validation KFCV). Esta técnica divide los datos originales en K subgrupos, cada grupo es excluido del grupo de datos de calibración y el modelo resultante se evalúa con el grupo excluido. Esto es repetido K veces hasta que todos los subgrupos sean validados. El promedio del desempeño a través de las K veces se considera el desempeño del modelo multivariado (Gautam *et al.*, 2015).

La validación cruzada determina la robustez del modelo estadístico (Gautam *et al.*, 2015) y evalúa la calidad de la predicción (Abdi, 2010). Además, esta validación ayuda a seleccionar el número de factores adecuados de un modelo para evitar sobreajuste (alto número de factores) o subajuste (pocos factores) y determina el error de predicción asociado para futuros trabajos (Gautam *et al.*, 2015).

Tanto en la validación interna y la externa se construyen los modelos predictivos con los datos de calibración y se evalúan con los datos de validación. Muchos autores emplean ambas validaciones, para confirmar los resultados de ambos modelos (Karlinasari *et al.*, 2014; Zhou *et al.*, 2015).

2.7.2. Pre-tratamiento de datos

El pretratamiento de datos constituye una parte integral y previa del modelamiento quimiométrico y su principal objetivo es remover interferencias o variaciones no deseadas antes del modelamiento (Rinnan, Berg, *et al.*, 2009). Las técnicas de preprocesamiento reducen la variabilidad no modelable en los espectros (Rinnan, Berg, *et al.*, 2009). Debe ser aplicado si se quiere obtener un modelo robusto de predicción por IR (Rinnan, Norgaard, *et al.*, 2009), sin embargo el preprocesamiento puede mejorar o perjudicar el éxito de un modelo (Kuhn y Johnson, 2013).

La mayor fuente de variación no deseada y comúnmente observado en espectros IR de muestras biológicas es el fenómeno de dispersión de luz o *scattering* (Butler *et al.*, 2018). La dispersión se produce con el esparcimiento de la radiación al contacto con partículas

pequeñas (Rinnan, 2014). Las condiciones ideales para crear este fenómeno sucede cuando las longitudes de onda en el espectro infrarrojo medio coincide con el tamaño de algunas muestras biológicas (2.5 - 25 μ m) (Butler *et al.*, 2018). El scattering se observa, en mayor medida, en muestras sólidas y es casi nulo en muestras líquidas, excepto en suspensiones donde se observa un ligero efecto de dispersión (Rinnan, 2014). La dispersión de luz causa aberraciones en la linealidad (Ley de Beer) y el desplazamiento en la línea base en los espectros (Rinnan, Berg, *et al.*, 2009).

La segunda fuente de variación no deseada es el ruido, es un problema inherente de FTIR y puede provenir de señales eléctricas, vibraciones mecánicas y parámetros ambientales que son inevitables. Los detectores con refrigerantes pueden ayudar a reducir el ruido pero no completamente (Butler *et al.*, 2018). Otra fuente de variación no deseada es la variación de la longitud del paso de luz que atraviesa la muestra o *path length*, fenómeno muy común en muestras solidas (Butler *et al.*, 2018; Rinnan, Berg, *et al.*, 2009). Este fenómeno es consecuencia de la heterogeneidad en el espesor de las muestras y cambios en la carga de la fuente infrarroja. La longitud del paso de la luz está directamente implicada en la ley de Lamber – Beer y puede reducir la calidad, precisión y exactitud del espectro. Con la aplicación de preprocesamiento estos parámetros puede mejorar dramáticamente (Butler *et al.*, 2018). En espectrofotómetros con ATR (reflexión total atenuada) el efecto de dispersión de la luz cobra mayor importancia en muestras sólidas así como las diferencias en la longitud del paso de luz (*path length*) (Butler *et al.*, 2018; Rinnan, Norgaard, *et al.*, 2009).

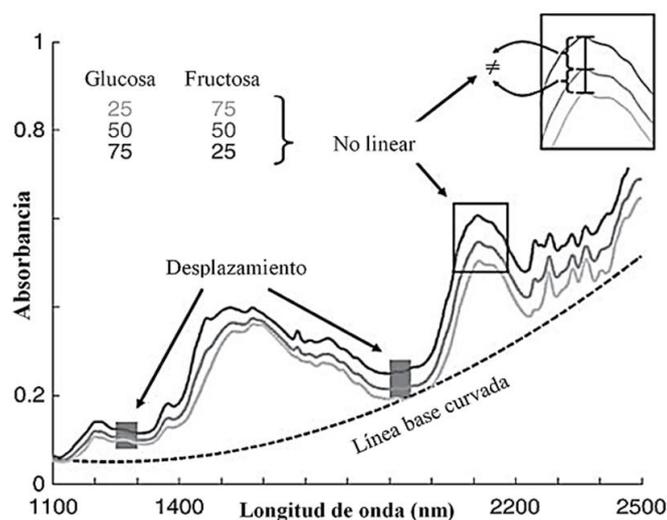


Figura 22: Desplazamiento y curvatura de la línea base

FUENTE: Rinnan, Norgaard, *et al.* (2009).

En los espectros, las interferencias no deseadas causan (i) pérdida de la linealidad entre la absorbancia de un analito y su concentración (ley de Beer), (ii) ruido ambiental, y (iii) desplazamiento y curvatura de la línea base (Rinnan, Norgaard, *et al.*, 2009).

(i) La relación lineal entre absorbancia y concentración del analito (Ley de Beer) se ve afectada por fenómenos físicos y químicos como dispersión de luz por partículas (*scattering*), interacciones moleculares, cambios en el tamaño de partícula o longitud de paso de luz (*path length*) (Rinnan, Norgaard, *et al.*, 2009). (ii) El ruido puede eclipsar características espectrales sutiles de una muestra (Butler *et al.*, 2018). En consecuencia, se combina información poco relevante (ruido) con información relevante (Rinnan, 2014). Asimismo, (iii) el desplazamiento y curvatura de la línea base, como se ilustra en la Figura 22, son causadas por la dispersión de luz (*scattering*) y afecta la linealidad entre la absorbancia y la concentración de un analito (Rinnan, Norgaard, *et al.*, 2009).

El objetivo principal del preprocesamiento es compensar la desviación de la relación lineal entre la señal espectral (absorbancia) y la concentración del analito, para que los datos espectrales se adhieran mejor a la ley de Beer (Rinnan, Norgaard, *et al.*, 2009). Del mismo modo, las técnicas de preprocesamiento corrigen las desviaciones de la línea base, ruido y variaciones en el espesor de las muestras (Trevisan *et al.*, 2012). Los métodos de preprocesamiento están enfocadas a NIR (infrarrojo cercano) ya que presenta mayor dispersión de luz (*scattering*), sin embargo pueden ser usados en otras técnicas espectroscópicas como Infrarrojo en general (IR) y Raman (Rinnan, 2014). Los métodos de pretratamiento están divididos en métodos de corrección de *scattering* y métodos derivativos (Rinnan, Norgaard, *et al.*, 2009).

Estos métodos de corrección de dispersión están diseñados para reducir el efecto físico de dispersión de la luz o *scattering* que causa variabilidad entre muestras. Esto incluye: corrección de línea base, corrección de señal multiplicativa (MSC), variante normal estándar (SNV) y normalización (Rinnan, Berg, *et al.*, 2009).

El método de la línea base se usa normalmente en análisis cuantitativo por infrarrojo (Christian, 2009). La aplicación de línea base obliga a una línea base curvada a formar una

línea base horizontal recta (FitzPatrick *et al.*, 2012). El primer paso es ajustar una línea base de un orden polinomial escogido al espectro de la muestra. Los puntos encima de esta línea base estimada se reemplazan por las predicciones de la línea base ajustada. Se asume que la línea base está formada por los puntos más bajos del espectro. La línea base artificial se ajusta nuevamente en el mismo orden polinomial hasta que ningún punto nuevo es reemplazado. Después del método de la línea base el espectro cambia a una nueva línea base cercana a cero (Rinnan, Norgaard, *et al.*, 2009).

Por otro lado, la corrección de señal multiplicativa o MSC (multiplicative signal correction) remueve tendencias no lineares en los datos espectrales causados por la dispersión de partículas y desplazamiento de espectros (FitzPatrick *et al.*, 2012; Rinnan, Norgaard, *et al.*, 2009). La operación se da en dos pasos: cálculo de los coeficientes de corrección y corrección de los espectros (Rinnan, Norgaard, *et al.*, 2009). Se calcula un espectro promedio o de referencia (X_{ref}) de las muestras analizadas, este espectro promedio se supone que está libre de ruido. Se calcula los coeficientes de corrección (b_0 y b_{ref}) a partir del espectro de referencia (fórmula 8). El ajuste de cada espectro original (X_{org}) al espectro de referencia se hace mediante mínimos cuadrados, que arroja parámetros o valores escalares para cada espectro. Con estos valores se obtiene los espectros corregidos (X_{corr}) como se muestra en la fórmula 9, así los datos están pre-procesados. (FitzPatrick *et al.*, 2012; Gautam *et al.*, 2015; Rinnan, Norgaard, *et al.*, 2009).

$$X_{org} = b_0 + b_{ref}X_{ref} + \epsilon \quad (8)$$

$$X_{corr} = \frac{X_{org} - b_0}{b_{ref}} \quad (9)$$

El método variante normal estándar (SNV) no requiere un espectro de referencia y la transformación de cada espectro está basada en la fórmula 10. Donde el espectro original (X_{org}) se resta el valor medio del espectro (a_0) y se divide entre la desviación estándar (a_1) calculada en todos los puntos del espectro (Gautam *et al.*, 2015; Rinnan, Norgaard, *et al.*, 2009). Los métodos SNV y normalización no involucran ajustes de mínimos cuadrados para calcular los parámetros de estimación, por lo tanto son más propensos a presentar ruido en el espectro procesado (Rinnan, Berg, *et al.*, 2009).

$$X_{corr} = \frac{X_{org} - a_0}{a_1} \quad (10)$$

El método de normalización está enfocado en minimizar los efectos de la fluctuación de la fuente de poder, scattering y variación en el espesor de la muestra. Este método corrige los efectos de escalamiento y desplazamiento causado por los fenómenos mencionados (Gautam *et al.*, 2015). Los métodos de normalización describen una serie de transformaciones y cálculos de cada espectro para convertir todos los datos a la misma escala (FitzPatrick *et al.*, 2012). Se usa la fórmula 10 pero los valores de a_0 y a_1 varían en función del tipo de normalización (Rinnan, Norgaard, *et al.*, 2009). El vector de normalización y min-max de normalización son las técnicas más utilizadas (Gautam *et al.*, 2015). En el caso de vector de normalización: a_0 es cero y a_1 se calcula como la raíz cuadrada de la suma de los cuadrados de las intensidades del espectro. Cada punto del espectro es dividido por la raíz calculada, así se obtienen los espectros normalizados (Gautam *et al.*, 2015). Por otro lado, Min-max de normalización: se calculan los máximos y los mínimos de la intensidad de un espectro. Cada punto se recalcula al restar el valor mínimo y luego dividir entre el rango (máximo - mínimo) (Gautam *et al.*, 2015).

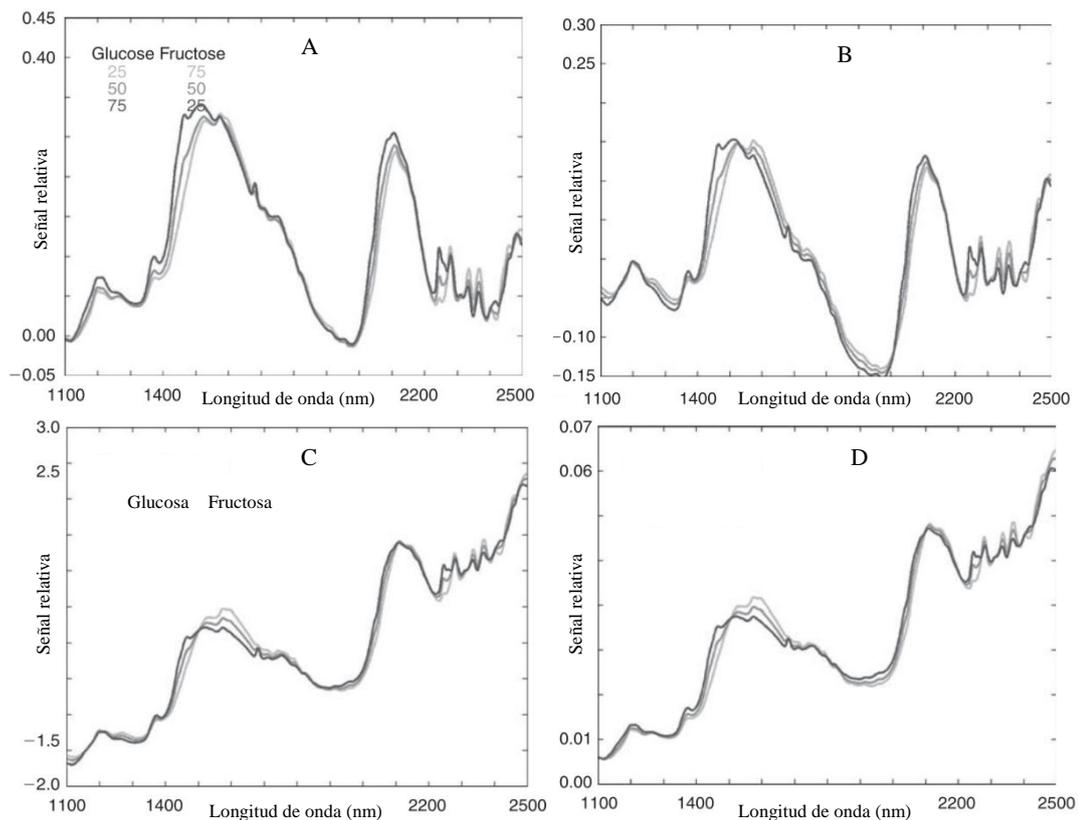


Figura 23: Espectros procesados mediante: A. Línea base. B. MCS. C. SNV. D. Normalización.

FUENTE: Rinnan et al. (2009).

Los métodos derivativos pueden ser el método de Savinsky - Golay y el método de Norris - Williams, ambos métodos usan suavizado (smoothing) antes de derivar los datos espectrales y tienen la finalidad de no minimizar demasiado la relación señal-ruido (Rinnan, Berg, *et al.*, 2009). El método de Norris - Williams es más práctico y fácil de aplicar con respecto al método de Savinsky - Golay. La diferencia radica en la forma de suavizar los datos (Rinnan, 2014).

Previo a la derivación de Norris – Williams, se realiza smoothing o suavizado de la siguiente forma: se realiza un promedio entre un número de puntos dados (m = número de puntos de smoothing) mediante la fórmula 11 y se obtiene un nuevo punto promedio. La primera derivada se calcula al restar dos puntos suavizados con intervalos entre ellos (en la Figura 24 A se observan 3 intervalos entre los puntos) y la segunda derivada suma los puntos externos del intervalo y resta dos veces el punto suavizado X_i (Rinnan, Berg, *et al.*, 2009).

$$X_{smooth,i} = \frac{\sum_{j=-m}^m X_{org,i+j}}{2m + 1} \quad (11)$$

Por otro lado, Savinsky – Golay plantean el ajuste de los datos mediante un polinomio. Los puntos de un espectro pasan a través de una ventana simétrica con $2g + 1$ puntos, donde se tiene un punto central con g puntos a cada lado. Los puntos de la ventana se ajustan a un polinomio de un orden n , y los coeficientes encontrados son usados para la estimación del smoothing y derivada del punto central. Este proceso se repite hasta que todos los puntos de espectro pasen a través de la venta simétrica para su transformación. En la Figura 24, se ilustra los dos métodos de suavizado discutidos (Rinnan et al., 2009).

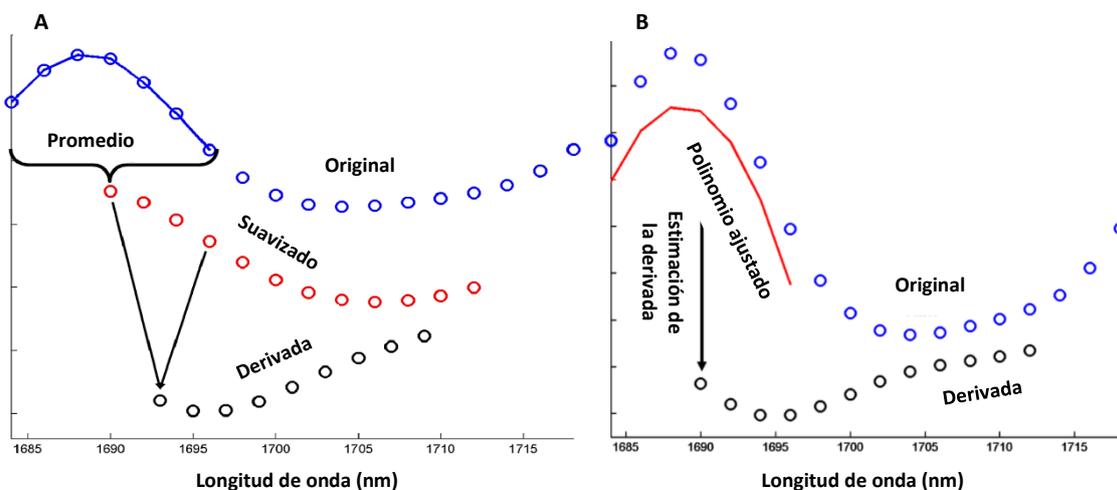


Figura 24: Suavizado de datos según Norris – Williams (A) y Savinsky – Golay (B)

FUENTE: Rinnan et al. (2009).

La derivación puede ser de primer y segundo orden. La primera derivada se estima como la diferencia entre dos puntos sucesivos en el espectro (absorbancia), la segunda derivada es la diferencia entre dos puntos sucesivos del espectro de la primera derivada (Rinnan, Berg, *et al.*, 2009). El desplazamiento de la línea base es removido por la primera y segunda derivada. La segunda derivada también remueve la curvatura de la línea base (Rinnan, Norgaard, *et al.*, 2009) y no desplaza el número de onda lo que hace la interpretación del espectro más sencillo porque se conservan las posiciones de los picos característicos. Esto no ocurre con la primera derivada, donde si hay desplazamiento (FitzPatrick *et al.*, 2012).

A pesar de la importancia del pretratamiento en análisis quimiométrico, no existe un consenso para su selección y aplicación en muestras biológicas (Butler *et al.*, 2018). Existen pocos trabajos que reportan los mejores tratamientos para un grupo de muestras basado en el impacto de los pretratamientos en sus datos y otros trabajos publicados no discuten acerca de estrategias para seleccionar el pretratamiento. Es común que los investigadores escojan los pretratamientos que funcionaron en otras publicaciones (Lee *et al.*, 2017). Además, debido a la alta variabilidad en muestras biológicas se debe tener un conocimiento *a priori* de los espectros para seleccionar el preprocesamiento adecuado (Butler *et al.*, 2018).

La combinación de más de dos pretratamientos es común en literatura y toda combinación es posible (Rinnan, Berg, *et al.*, 2009). Si se usa más de un pretratamiento, no existe guía para escoger la combinación o el orden de los pretratamientos (Lee *et al.*, 2017). Se recomienda que el primer método de pretratamiento usado remueva la mayor fuente de variación espectral para que no influya en los siguientes pasos. Generalmente los métodos de línea base cumplen con esta característica y se aplican previo a los métodos de normalización (Butler *et al.*, 2018). La evaluación del mejor pretratamiento o las mejores combinaciones se realizan por inspección visual del espectro, proyección de clustering por PCA o por el error calculado por el modelo de predicción (Lee *et al.*, 2017). El enfoque más efectivo de preprocesamiento debe ser el más sencillo y el número de métodos de preprocesamiento debe ser el mínimo (Butler *et al.*, 2018).

2.7.3. Regresión de mínimos cuadrados parciales (PLS)

Regresión de mínimos cuadrados parciales o PLS (Partial least square) es un método estadístico de regresión cuyo objetivo es predecir un conjunto de variables dependientes o respuesta (Y) a partir de un conjunto de variables independientes (X) o predictores (Abdi, 2010). La regresión PLS analiza predictores (X) altamente correlacionados (información similar), abundantes (mayor que el número de muestras) y con ruido, los cuales son comunes en instrumentación química espectrofotómetros, cromatógrafos, etc (Kuhn y Johnson, 2013; Wold *et al.*, 2001). Para predecir la variable respuesta (Y) a partir de los predictores (X) se extrae un conjunto de factores ortogonales llamados variables latentes, las cuales tienen el mejor poder de predicción (Abdi, 2010; Wold *et al.*, 2001). Es decir, PLS permite expresar el alto número de información de las variables predictores (X) en un número reducido de variables latentes que son las más relevantes para predecir las variables respuesta (Y) (Héberger, 2008).

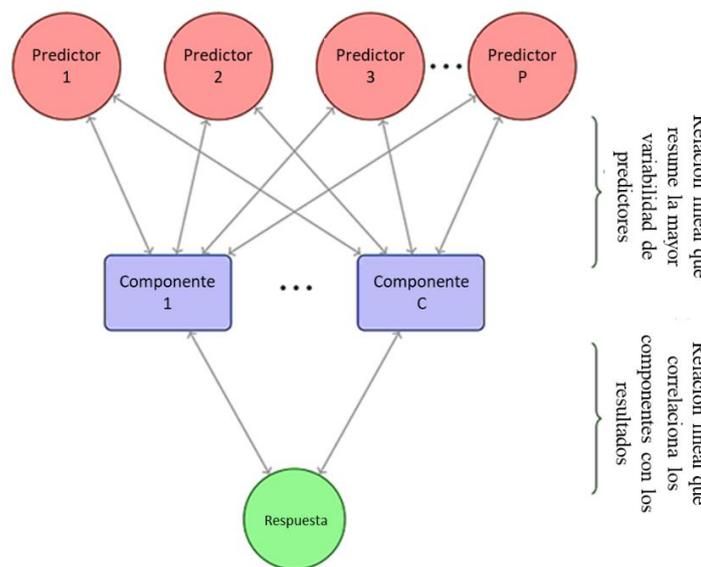


Figura 25: Estructura del modelo PLS

FUENTE: Kuhn & Johnson (2013)

Los datos deben ser preparados antes de construir el modelo, la preparación principalmente consiste en eliminar, adicionar o transformar los datos. Esto permite mejorar la eficiencia del modelo. La reducción de predictores genera variables latentes que captura la mayor información de los datos originales y contiene la mayor varianza posible. Asimismo, estas variables tienen la máxima correlación con la variable respuesta, como se observa en la Figura 25.

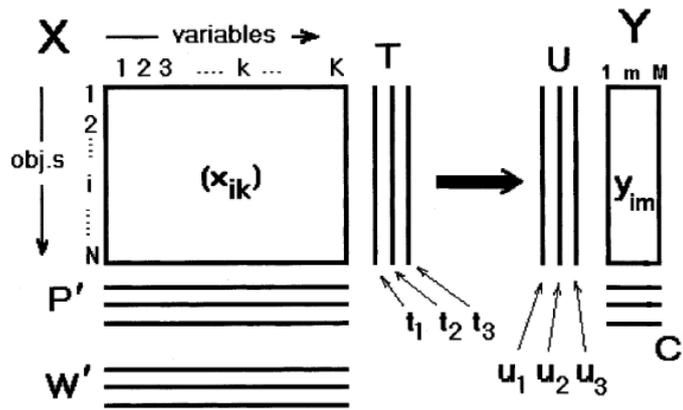


Figura 26: Matrices de datos PLS

FUENTE: Wold *et al.* (2001)

El modelo PLS calcula las variables latentes en una matriz de anotación (T). Esta matriz se estima por la combinación lineal de las variables X originales y la matriz de cargas (W'), como se muestra en la fórmula 12. En la Figura 26 se muestra la matriz X con k factores de i muestras y la matriz W' con a variables latentes por cada factor k (Wold *et al.*, 2001).

$$T = XW' \quad (12)$$

Se asume que la matriz X y Y puede ser, en parte, explicadas por la matriz de anotación (T), ver fórmula 13 y 14 (Wold *et al.*, 2001). En estas últimas fórmulas, P' y C' son matrices de coeficientes que relacionan T con los predictores (X) y las variables respuesta (Y), respectivamente. Además, E y F representa la información residual remanente (Gowen *et al.*, 2011).

$$X = TP' + E \quad (13)$$

$$Y = TC' + F \quad (14)$$

Si se reemplaza T de la fórmula 12 en la fórmula 14, se obtiene la fórmula 15. Donde los valores de Y se estiman por la multiplicación de los datos predictores (X) y la regresión de coeficientes representado por la matriz B . Finalmente, F representa la desviación entre las variables respuesta observada y predicha.

$$Y = XB + F \quad (15)$$

La fórmula 16 representa el modelo PLS y otros modelos de regresión lineal en general:

$$y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_jx_{ij} + e_i \quad (16)$$

Donde y_i representa la variable respuesta numérica i , b_0 el intercepto estimado; b_1 es el coeficiente estimado del primer predictor; x_{i1} , el valor del primer predictor y e_i , el error aleatorio que no puede ser explicado por el modelo (Kuhn y Johnson, 2013).

Además de la cuantificación, el modelo PLS se usa para estudiar todo el espectro y mediante técnicas de correlación determinar que partes del espectro están relacionadas a las propiedades de interés (van de Voort, 1992). Por ejemplo, se usó PLS para determinar la regiones del espectro infrarrojo que ayudan a predecir la concentración de lignina, celulosa, hemicelulosa y extractivos en madera (Chen et al., 2010; Zhou et al., 2015)

2.7.4. Evaluación del modelo de calibración

Existen parámetros que permiten evaluar la exactitud del modelo PLS. En particular, la raíz del error cuadrático medio (RMSE por sus siglas en inglés) que indica que tanto se alejan los valores calculados por el modelo con respecto a los valores de referencia (Kuhn y Johnson, 2013). Este parámetro es ideal para comparar la capacidad predictiva entre modelos (Kjeldahl y Bro, 2010; Kuhn y Johnson, 2013). Además, el coeficiente de determinación R^2 mide que tan exitoso es el ajuste del modelo para explicar la variación de los datos (Li et al., 2015). Sin embargo, este parámetro es específico del grupo de datos, por lo que no es ideal para comparar modelos predictivos (Kjeldahl y Bro, 2010). Otro parámetro que indica la capacidad predictiva de un modelo es la razón de rendimiento y desviación o RPD (Karlinasari *et al.*, 2014). El valor de RPD clasifica a los modelos desde excelentes hasta solo aptos para tamizado (Hein, 2010; Karlinasari *et al.*, 2014).

El RMSE es la distancia promedio entre los valores de referencia (y_i) y las predicciones del modelo (\hat{y}_i) (Kuhn y Johnson, 2013). Mientras menor sea su valor, mejor es el modelo predictivo (Zhou et al., 2015). La diferencia entre $y_i - \hat{y}_i$ se conoce como residuales del modelo y ayuda a calcular el RMSE; la suma de los residuales se eleva al cuadrado para finalmente sacar la raíz cuadrada (Fórmula 17). Por este motivo el RMSE está en las mismas unidades que los datos originales lo que facilita su interpretación.

El RMSEP (error cuadrático medio de la predicción) y RMSECV (error cuadrático medio de la validación cruzada) se calculan con la fórmula 17. Donde n es el número de muestras de validación y los residuales $(y_i - \hat{y}_i)$ provienen del grupo de validación para RMSEP (Faber y Rajkó, 2007). En el caso de RMSECV se toma en cuenta la diferencia entre el valor predicho \hat{y}_i y el valor de referencia de las muestras excluidas y_i . El modelo PLS se corre una vez para cada número de variable latente y se calcula el RMSECV correspondiente (Hubert y Branden, 2003).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (17)$$

Otro parámetro importante es el coeficiente de determinación R^2 . Este parámetro representa el porcentaje de la información que es explicado por el modelo de regresión (Kuhn y Johnson, 2013) y es una media de correlación mas no de exactitud (Kuhn y Johnson, 2013). Li et al, (2015) señala que R^2 se calcula por la fórmula 18. Donde n es el número de muestras, y_i es el valor de referencia para la muestras i , \hat{y}_i es el valor predicho para la muestra i , \bar{y}_i es la media de y_i para todas las muestras (X. Li *et al.*, 2015).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (18)$$

Si el valor de R^2 es mayor a 0.91, el modelo se considera excelente. Un modelo con R^2 entre 0.82 y 0.9 es bueno para hacer predicciones mientras que el rango 0.66 a 0.81 arroja predicciones cuantitativas aproximadas. Finalmente, un R^2 entre 0.5 y 0.65 construye modelos que discriminan entre altas y bajas concentraciones (He y Hu, 2013).

La razón de rendimiento y desviación o RPD (ratio of performance to deviation) es un indicador de la capacidad predictiva de un modelo (Karlinasari *et al.*, 2014). Es decir, RPD es calculado para estimar la habilidad predictiva de un modelo, mientras más alto es el valor de RPD, mayor habilidad predictiva obtiene el modelo. Se calcula mediante la fórmula 19. Donde n_p es el número de muestras en el set de predicción, $y_{i,p}$ es el valor químico de referencia para la muestra i , $\bar{y}_{i,p}$ es la media de todas las muestras del grupo de predicción y SEP es el error estándar de la predicción (Li et al., 2015).

$$RPD = \sqrt{\frac{\sum_{i=1}^{n_p} (y_{i,p} - \bar{y}_{i,p})^2}{(n_p - 1)}} / SEP \quad (19)$$

Santos et al. (2016) propone la fórmula 20 para calcular el RPD. Donde DP_{val} es la desviación estándar de los datos de validación.

$$RPD_{val} = DP_{val} / RMSEP \quad (20)$$

En muestras agrícolas, si el RPD es mayor a 1.5 se considera ideal para tamizado y predicciones iniciales, de 2.0 a 2.5 indica una muy buena predicción, y mayor a 3 indica que el modelo predice de forma eficiente (Karlinasari *et al.*, 2014). En muestras forestales un RPD de 1.5 a 2.5 es suficiente para estimar propiedades de la madera (Hein, 2010).

En general, los modelos con bajo RMSEP, alto R^2 y alto RPD son considerados como los mejores modelos (Zhou et al., 2015). Sin embargo, un valor alto de correlación (R^2) no siempre significa un mejor modelo. Además de la eficiencia del modelo PLS, este parámetro depende de la distribución de los datos. Como resultado el R^2 no es significativo para comparar modelos ya que es específico de un grupo de datos. En particular el R^2 se usa con precaución en la optimización de modelos y como medida final de la calidad de del modelo final. Por otro lado el RMSEP es mejor para hacer comparaciones (Kjeldahl y Bro, 2010).

2.7.5. Número de variables latentes en pls

Los factores o variables latentes son aquellos componentes que resumen la mayor variabilidad de los predictores (X) y al mismo tiempo poseen la máxima correlación con las variables respuesta (Y) (Kuhn y Johnson, 2013). Escoger el número óptimo de factores es análogo a determinar el grado del polinomio que se ajusta mejor a los datos (x, y) en modelos univariados. En el caso de análisis multivariado la selección del número de variables latentes determina la dimensionalidad óptima del modelo (Faber y Rajkó, 2007).

Un modelo que incluya muchas variables latentes se considera sobre ajustado mientras que un modelo con muy pocas variables latentes esta sub ajustado (Gowen *et al.*, 2011). Los modelos multivariados como el PLS son propensos al sobre ajustado debido a que los datos

que ingresan al modelo son abundantes, tienen alta correlación e interferencia de ruido (Faber y Rajkó, 2007; Wold *et al.*, 2001). Las variables latentes adicionales en PLS pueden compensar la tendencia no lineal si los datos no se apegan a la ley de Beer (Rinnan, Berg, *et al.*, 2009). Sin embargo, incluir variables latentes adicionales incrementa la complejidad del modelo, lo que conlleva a incorporar ruido al modelo y reducir la robustez para futuras predicciones (Faber y Rajkó, 2007; Rinnan, Berg, *et al.*, 2009).

Por lo tanto, la selección del número de variables latentes es un paso crítico en PLS (Faber y Rajkó, 2007). Para este fin es común usar la raíz de error cuadrado medio o RMSE; así como la suma de cuadrados de los errores predichos o PRESS (prediction error sum of squares) (Acquah *et al.*, 2016; FitzPatrick *et al.*, 2012; Haaland y Thomas, 1988). El número óptimo de variables latentes es aquel que arroja el menor valor RMSE o PRESS (Haaland y Thomas, 1988; Hubert y Branden, 2003). Sin embargo, el RMSE no siempre exhibe un valor mínimo de forma clara. Esto conlleva a escoger el primer valor mínimo de RMSE o el valor donde empiece una meseta (Faber y Rajkó, 2007).

El criterio ampliamente usado para la selección del número óptimo de factores es propuesto por Haaland y Thomas (1988). Estos autores proponen comparar el modelo con el menor valor de PRESS que tiene h^* factores con los modelos que presenten menor número de factores ($h < h^*$). El objetivo de la comparación es escoger el modelo con el menor número de factores (h) que no presente un PRESS significativamente mayor que el modelo con el mínimo valor de PRESS (h^*). El estadístico F es usado para determinar la significancia de los modelos. La fórmula 21 muestra el cálculo del estadístico F.

$$F(h) = \frac{\text{PRESS}(h \text{ factores})}{\text{PRESS}(h^* \text{ factores})} \quad (21)$$

Se calcula el estadístico F para todos los modelos con $h = 1, 2, 3, \dots, h^*$ factores. Se escoge el menor número de factores que cumpla con la condición de la fórmula 22. Donde m es el número de muestras de la calibración y α es el percentil de la distribución F de Snedecor y su valor arbitrario recomendado es 0.25 (Haaland y Thomas, 1988).

$$F(h) < F_{\alpha, m, m} \quad (22)$$

Si $F(h)$ sobrepasa $F_{\alpha, m, m}$ quiere decir que el modelo tiene PRESS significativamente mayor

que el PRESS mínimo. Por lo tanto, el modelo con el menor número de factores que no sea significativamente mayor que el PRESS mínimo es el modelo escogido (Haaland y Thomas, 1988).

Gowen et al. (2011) propone otro método, el cual considera un indicador visual de sobreajuste la alta oscilación de picos en el gráfico de coeficientes de regresión a medida que el número de factores se incrementa. El método plantea la cuantificación de esta oscilación de picos mediante el estadístico Jaggedness (J) (Gowen *et al.*, 2011). La fórmula 23 muestra el cálculo del estadístico J:

$$J_j = \sum_{i=2}^m \sqrt{(b_{ji} - b_{ji-1})^2} \quad (23)$$

Donde j es el número de factores, i es el número de onda y b son los puntos sucesivos del vector de regresión. Los valores de J calculado para cada factor $j=1, 2, 3, \dots, m$ se reescalan al rango 0 -1, el parámetro reescalado se conoce como Medida del vector de regresión o RVM (regression vector measures) y se calcula con la fórmula 24. Donde r hacer referencia al reescalado, min es mínimo y max es máximo.

$$RVM_r = \frac{RVM - RVM_{\min}}{RVM_{\max} - RVM_{\min}} \quad (24)$$

Otro parámetro es la medida de la desviación o BM (bias measure) que reescala el valor de RMSE de 0 a 1 mediante la fórmula 25.

$$BM_r = \frac{BM - BM_{\min}}{BM_{\max} - BM_{\min}} \quad (25)$$

Los parámetros RVM y BM de cada factor se suman y se elige el menor valor de la suma. La Figura 27 muestra una curva en forma de U, donde el menor valor determina el número de factores. En los casos donde el centro de la U es aplanada se escoge el menor número de factores en lugar del mínimo absoluto para lograr un modelo con mayor parsimonia (Gowen *et al.*, 2011).

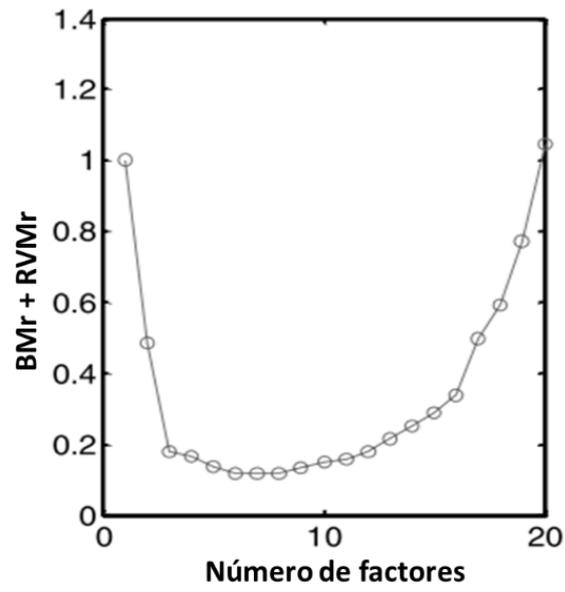


Figura 27: Gráfico de BM + RVM en función del número de factores

FUENTE: Gowen *et al.* (2011)

III. METODOLOGÍA

3.1. MATERIALES

3.1.1. Reactivos

Ácido sulfúrico (puro), Sulfito de sodio ACS ANKOM, alfa amilasa bacteriana estable a altas temperaturas AKOM (17 400 unidades liquefon/mL), detergente neutro AKOM (60g/L), detergente ácido AKOM (20g/L de ácido sulfúrico 1N), trietilen glicol AKOM, tiras de tornasol, agua destilada y acetona.

3.1.2. Equipos

Estufa de secado marca Binder, balanza analítica marca Sartorius, molino de bolas marca Tecnal, campana de extracción, analizador de fibra AKOM 200, incubadora Daisy, sellador de calor, agitador magnético Thermo scientific, espectrofotómetro infrarrojo Spectrum 100 con UATR marca Perkin Elmer.

3.1.3. Material de vidrio y de invernadero

Beaker 2 y 3L, fiolas de 1L, vasos de precipitado 100, 250 y 500 mL, tubos de ensayo, jarra de 1L, cámara de desecación, bolsas de filtro (retiene partículas mayores a 25 micrones), marcador permanente resistente a solventes y ácidos. Tijera de podar, bolsas de papel y bandejas de plástico.

3.1.4. Material vegetal

Plantas de Capirona y Bolaina fueron adquiridas del Ayahuasca Ecolodge E.I.R.L. y Est. Exp. Agraria Pucallpa – Ucayali – INIA. Estos árboles fueron cultivados en casas malla en el PIP de cereales y granos nativos, con control de horas de luz, riego constante y control de plagas.

3.1.5. Softwares

Peak Spectroscopy Software (<https://www.essentialftir.com/>) y R studio (<https://rstudio.com/products/rstudio/download/>)

3.2. PREPARACIÓN DE LAS MUESTRAS VEGETALES

3.2.1. Capirona

El número de muestras fue 10 veces mayor al número de variables latentes para armar los modelos predictivos (Lohmöller, 1989). Se tomaron 100 plantas de Capirona que alcanzaron 1.5m de altura aproximadamente y se cosecharon a finales de septiembre 2019. Cada muestra se obtuvo al agrupar 4 plantas de Capirona. De cada muestra se tomaron las secciones apicales y basales de los tallos. Los dos entrenudos adyacentes a la yema (excepto la yema apical) se cosecharon como la sección apical. El tercer y cuarto entrenudo del tallo basal se cosechó como la sección basal, esto se observa en el registro fotográfico (ANEXO 23). Se obtuvieron 25 muestras de la sección apical (A) y 25 de la sección basal (B).

3.2.2. Bolaina

El número de muestras fue 10 veces mayor al número de variables latentes para armar los modelos predictivos (Lohmöller, 1989). Se emplearon 150 plantas de Bolaina en esta etapa. Se formaron grupos de 6 plantas de altura similar. De los 25 grupos obtenidos se tomó la sección apical y basal de cada planta. Para la sección apical se descartó la yema apical y los dos primeros entrenudos, seguidamente se cosechó los 20 cm de tallo contiguos. Para la sección basal, se cosechó los 20cm inferiores del tallo desde la base. Se obtuvieron 25 muestras de la sección apical (A) y 25 de la sección basal (B).

Las muestras se trozaron en piezas pequeñas (≈ 1 cm), se pesaron en bolsas de papel y secaron en estufa a 60°C hasta que el peso fue constante. Se almacenaron a temperatura ambiente en un desecador para conservar las muestras.

3.3. ANÁLISIS QUÍMICO

La determinación de celulosa, lignina y hemicelulosa se realizó mediante el sistema de análisis de fibra con detergentes propuesto por Van Soest y Robertson, 1979 adaptada a la incubadora Daisy y al analizador de fibra AKOM 2000 (ver figura 28). El análisis secuencial

con detergentes permitió obtener el porcentaje de fibra detergente neutra (FDN), seguido de fibra detergente ácida (FDA) y lignina (Van Soest y Robertson, 1979). El detergente neutro (DN) solubiliza pectinas, taninos, sílica y proteínas. A este detergente se añadió alfa amilasa bacteriana estable a altas temperaturas para extraer almidón a 100°C, y etilenglicol que ayudó a remover el almidón degradado y mejoró la eficiencia de la amilasa (Van Soest et al., 1991). Además, se añadió sulfito de sodio que se une a puentes de sulfuros de proteínas y reduce el contenido proteico de la fibra analizada. La fibra detergente neutra es la fracción no soluble retenida luego de la digestión con detergente neutro, y está compuesta principalmente por celulosa, hemicelulosa y lignina. (Van Soest et al., 1991). El detergente ácido (DA) solubilizó hemicelulosa y proteínas remanentes; a su vez retuvo, como fracción no soluble, lignina y celulosa. Por último, la digestión con ácido sulfúrico 24N permitió la remoción de celulosa y retención de lignina Klason (no soluble en ácido sulfúrico), como se muestra en la siguiente imagen (Van Soest et al., 1991).

Una vez obtenidos los valores de %FDN, %FDA y %Lignina se procedió con las ecuaciones 16, 17 y 18 para hallar el porcentaje de hemicelulosa y celulosa.

$$\text{Fibra detergente neutra (FDN)} = \text{Hemicelulosa \%} + \text{Celulosa\%} + \text{Lignina\%} \quad (26)$$

$$\text{Hemicelulosa (\%)} = \text{FDN} - \text{Fibra detergente ácida (FDA)} \quad (27)$$

$$\text{Celulosa (\%)} = \text{FDA} - \text{Lignina (\%)} \quad (28)$$

$$\text{Celulosa (\%)} + \text{Hemicelulosa (\%)} = \text{Holocelulosa (\%)} \quad (29)$$

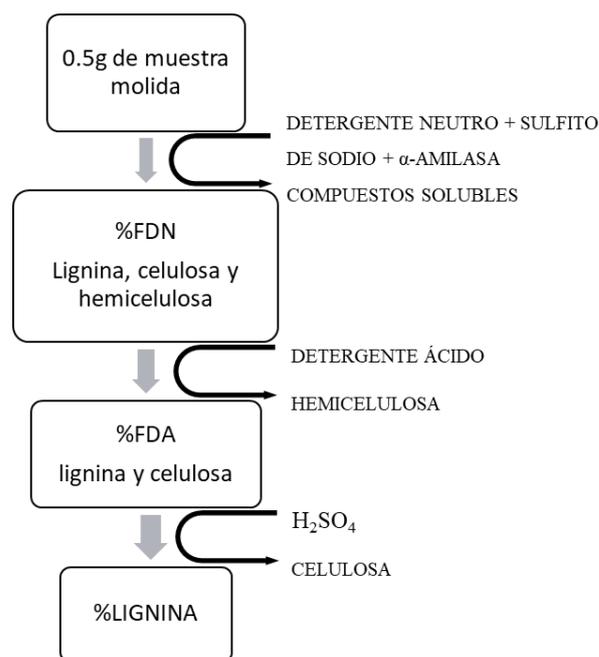


Figura 28: Análisis secuencial de fibra basado en la metodología de Van Soest y Robertson

FUENTE: Van Soest y Robertson (1979).

3.3.1. Determinación de fibra detergente neutra (FDN)

Las muestras molidas fueron pesadas dentro de bolsas de filtro y el peso de la bolsa se registró como W_1 . Se pesó 500 mg de cada muestra y se registró como W_2 . Las bolsas fueron selladas herméticamente. Se colocaron las bolsas en canastillas (3 por canastilla), las cuales se inmovilizaron con un soporte. Las canastillas se sumergieron en el contenedor del analizador de fibra AKOM, con 2 L de solución de detergente neutro, 20 g de sulfito de sodio, 20 mL de trietilenglicol y 4 ml de alfa amilasa. Las muestras fueron incubadas por 75 minutos en agitación constante a 100°C. Al término se drenó la solución y las bolsas fueron enjuagadas con 2L de agua a 70°C. Además, se añadió 4 mL alfa amilasa y se dejó en agitación por 5 minutos. Seguidamente, se drenó la solución y el enjuague se repitió tres veces. Se extrajeron las bolsas de filtro y se quitó el exceso de agua mediante ligera presión. Las bolsas se sumergieron en un beaker con acetona durante 5 minutos. Seguidamente, las bolsas de filtro se dejaron secar a temperatura ambiente para eliminar los restos de acetona por evaporación y se terminó el secado en estufa a 102°C. Una vez secas, las bolsas se dejan enfriar en un desecador hasta llegar a temperatura ambiente y su peso se registró como W_3 . Se incluyeron dos bolsas vacías en el análisis para determinar la pérdida de partículas de las muestras (Akon-Techonology, 2017), y la corrección de bolsa blanco (C1).

$$\% \text{FDN} = \frac{100 \times (W_3 - (W_1 \times C_1))}{W_2} \quad (30)$$

3.3.2. Determinación de fibra detergente neutra (fda)

Las muestras del ítem 3.3.1 se colocaron en canastillas y se inmovilizaron en un soporte universal. Se introdujeron las canastillas en el analizador de fibra con 2L de detergente ácido (40g/2l de ácido sulfúrico 1N) y se las incubó a 100°C por 60 minutos. Es crítico respetar el tiempo de digestión. Al término, se drenó la solución y se enjuagó con agua a 70°C durante 5 minutos en agitación constante. Este proceso se repitió dos veces más. Después del tercer enjuague se midió el pH de la solución, si es ácido se continúan los enjuagues hasta alcanzar un pH neutro. Se extrajeron las bolsas y el exceso de agua se quitó mediante una ligera presión. Para luego ser sumergidas en un beaker con acetona por 5 minutos. Luego se extrajeron las bolsas y se dejó evaporar la acetona en una campana de extracción. El secado se completó en una estufa a 102°C de 2 a 4 horas. Finalmente se dejaron enfriar en un desecador para tomar su peso (W3) (Akon-Technology, 2017).

$$\% \text{FDA} = \frac{100 \times (W_3 - (W_1 \times C_1))}{W_2} \quad (31)$$

3.3.3. Determinación de lignina

La determinación de lignina se llevó a cabo con las bolsas del ítem 3.3.2, las cuales fueron en ácido sulfúrico al 72%. Se colocaron 96 muestras en la incubadora Daisy, en donde se agitaron gentilmente durante 3 horas a temperatura ambiente. Al término se eliminó el ácido sulfúrico y se realizaron 7 lavados sucesivos con agua hasta llegar a pH neutro, el cual se controló con tiras de tornasol. Luego, las bolsas se sumergieron en acetona durante 5 minutos y se extrajeron las bolsas para dejarlas secar a temperatura ambiente. Finalmente se secaron las bolsas en estufa a 105°C y se pesaron como W₃ (Akon-Technology, 2016). W₂ es el peso de la muestra al inicio de los análisis químicos previos 0.5g (Akon-Technology, 2016).

$$\% \text{LIGNINA} = \frac{100 \times (W_3 - (W_1 \times C_1))}{W_2} \quad (32)$$

El contenido de celulosa, hemicelulosa, holocelulosa y lignina de cada muestra se reporta en el Anexo 1 de Capirona y Anexo 2 de Bolaina.

3.4. COLECTA DE ESPECTRO FTIR

Las muestras molidas se tamizaron a través de filtros de 60mesh para uniformizar las partículas a 0.25mm (Li et al., 2015). Se colectaron 3 espectros por muestra en el equipo Spectrum 100 con UATR marca Perkin Elmer, en el rango de 4000 a 400 cm⁻¹ y resolución de 4 cm⁻¹ (Traoré *et al.*, 2018). Sin embargo, se tomaron 32 scans por muestra en lugar de 100 scans como reporta Traoré et al., 2018. Esta etapa se realizó en los laboratorios de Científica Andina S.A.

3.5. PROCESAMIENTO DE DATOS Y ANÁLISIS

3.5.1. Análisis estadístico de celulosa, hemicelulosa, holocelulosa y lignina.

En cada especie se evaluó las diferencias en la composición de celulosa, hemicelulosa, holocelulosa y lignina, entre la sección apical y basal. Para lo cual se empleó el criterio de análisis reportado por Camel et al. (2019). Donde se emplearon histogramas de frecuencia para visualizar la distribución de datos y la prueba de Shapiro-Wilk fue utilizada para determinar la normalidad de los datos. Se empleó los modelos lineales generalizados mixtos en los datos que no cumplieron con la normalidad. Se evaluó la homogeneidad de varianza de los modelos mixtos lineales generalizados mediante el gráfico de residuales para cada componente Finalmente, para comparar la composición química entre la sección apical y basal de los tallos de Capirona y Bolaina se usó la prueba de Tukey. Los datos cumplen los supuestos de normalidad, homocedasticidad e independencia en las observaciones. Para los análisis estadísticos de Capirona (ver anexos 3, 4, 5, 6 y 7); y de Bolaina (ver anexos 8, 9, 10, 11 y 12).

3.5.2. Análisis quimiométrico

Mediante la regresión de mínimos cuadrados parciales o PLS (partial least square) se construyeron modelos para predecir la composición de madera (celulosa, hemicelulosa, holocelulosa y lignina) a partir de los espectros FTIR en Bolaina y Capirona. Los modelos se desarrollaron en el programa Peak Spectroscopy de la compañía Essential FTIR. El software combina los espectros de las muestras (etiquetados como muestras independientes) en una matriz X (matriz de variables predictoras). Mientras que el contenido de los componentes de madera determinado por el método químico de Van Soest, 1984 se combina en una matriz Y (matriz de variables respuesta).

Los modelos de calibración PLS se construyeron mediante validación externa y validación cruzada sin un valor (*cross validation leave one out*). La validación externa divide los datos en un grupo de calibración, con el que se construye el modelo PLS y otro grupo de predicción, el cual valida el modelo construido. El primer grupo se formó con 2/3 de los datos (n=100) y el segundo grupo, con los datos restantes (n=50) (Karlinasari *et al.*, 2014). En tanto, la validación cruzada excluye una muestra de los datos y con los datos restantes se construyó el modelo PLS para predecir el valor de la muestra excluida. Este proceso se repitió hasta que todas las muestras son excluidas (Zhou *et al.*, 2011).

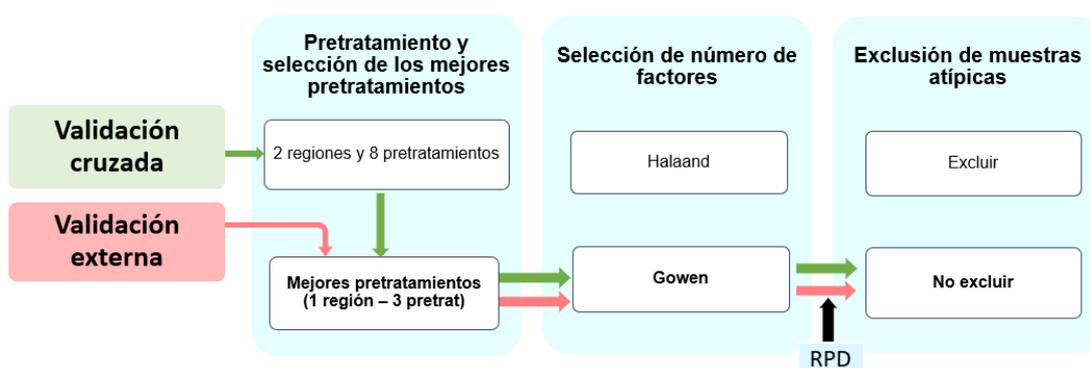


Figura 29: Flujograma para la construcción de modelos PLS

FUENTE: Elaboración propia.

La selección de pretratamientos y regiones de análisis óptimos fue el primer paso en la construcción de modelos predictivos (Figura 29). Los pretratamientos reducen las interferencias causadas por ruido y efecto de dispersión en los espectros. Los métodos de pretratamiento considerados fueron: 1° derivada S-G, 2° derivada S-G, Smoothing, standard normal variation, MSC, straight line correction y min-max normalization. Previo a la aplicación de los pretratamientos, los espectros fueron homogenizados a través del método de la media centrada (Acquah *et al.*, 2016). Asimismo, se evaluaron las siguientes regiones: a. 3700 cm^{-1} a 850 cm^{-1} (espectro completo) y b. 1800 cm^{-1} a 850 cm^{-1} (huella dactilar) (Acquah *et al.*, 2016). En los anexos 13 y 14 se muestran los resultados de los pretratamientos en las dos regiones de análisis para cada compuesto de Capirona y Bolaina, respectivamente.

La validación cruzada permitió seleccionar los mejores 2 pretratamientos de cada componente y la región óptima de análisis. Estos modelos se evaluaron en base a indicadores estadísticos como: la raíz cuadrada de error medio de la validación cruzada (RMSECV) y el coeficiente de determinación (R^2) (He y Hu, 2013). Con los 2 pretratamientos y región seleccionados se construyeron los modelos por validación externa. Estos últimos modelos se evaluaron mediante: la raíz cuadrada de error medio de la calibración y validación (RMSEC y RMSEP) y el coeficiente de determinación de calibración y validación (R_c^2 y R_p^2) (Zhou et al., 2015). En general, los modelos con bajo RMSEP y alto R^2 son considerados como los mejores modelos (Zhou et al., 2015, Li et al., 2015).

El número de factores se determinó por el método de Haaland y Thomas (1988) y para evitar modelos sobreajustados se empleó el método propuesto por Gowen et al. (2011). Este último método cuantifica en ruido de la regresión de coeficientes (estadístico J) y el margen de error del modelo (RMSEP) para determinar el número óptimo de factores. En los anexos 15, 16, 17 Y 18, se muestra la determinación del número de factores para Capirona según el método Gowen et al. (2011). En el caso de Bolaina, los cálculos se muestran en los anexos 19, 20, 21 y 22.

Finalmente, se incluyó la desviación predictiva residual o RPD (residual predictive deviation) como indicador de la capacidad predictiva de los modelos (Karlinasari *et al.*, 2014). Se usaron las fórmulas propuestas por Santos et al. (2016) y Li et al. (2015).

IV. RESULTADOS Y DISCUSIÓN

4.1. COMPOSICIÓN QUÍMICA DE LA MADERA DE CAPIRONA

Mediante el análisis de fibra propuesto por Van Soest y Robertson (1979) se analizó el contenido de celulosa, hemicelulosa y lignina en dos secciones de los tallos de Capirona de un año (Tabla 6). Adicionalmente se registró el contenido de holocelulosa, suma del contenido de celulosa y hemicelulosa (Jiang *et al.*, 2014). En el presente estudio se reportó a celulosa como el componente más abundante en la sección apical y basal (29.81% y 35.97%), seguido de hemicelulosa (10.17% y 9.64%), mientras que lignina fue el menos abundante (8.14 % y 9.65%). Por otro lado, el contenido de celulosa fue el que presentó mayor variabilidad en ambas secciones del tallo (D. E =2.47 y 2.94). En contraste, el contenido de lignina presenta menor variabilidad, tanto en la sección apical como en la basal (D.E=0.64 y 0.67).

Tabla 6: Composición de tallo de Capirona (*Calycophyllum spruceanum*)

	Celulosa (%)	Hemicelulosa (%)	Lignina (%)	Holocelulosa (%)
Sección Apical				
Promedio	29.81	10.17	8.14	39.98
Mínimo	26.24	8.08	7.05	35.84
Máximo	36.13	12.13	9.88	46.19
Desviación estandar	2.47	0.92	0.64	2.47
Coefficiente de variación	0.08	0.09	0.08	0.06
Sección Basal				
Promedio	35.97	9.64	9.65	45.61
Mínimo	30.91	5.49	8.14	38.99
Máximo	41.69	13.64	11.28	51.49
Desviación estandar	2.94	1.39	0.67	3.16
Coefficiente de variación	0.08	0.14	0.07	0.07

El presente trabajo (Tabla 6) es el primero en reportar la composición química en tallos juveniles de *C. spruceanum*, mientras que la literatura reporta la composición química de

árboles maduros. Por ejemplo, la concentración de celulosa y hemicelulosa (Tabla 6) es la mitad de la concentración publicada por Sibille (2016), quien reportó el contenido de celulosa de 59.71% y hemicelulosa de 18.6% en madera de *C. spruceanum*. Estos resultados concuerdan con el trabajo de Santana y Okino (2007) en 36 especies de árboles amazónicos, quienes reportaron el contenido de celulosa (40 a 60%), hemicelulosa (18 a 25%) y lignina (24 a 36%).

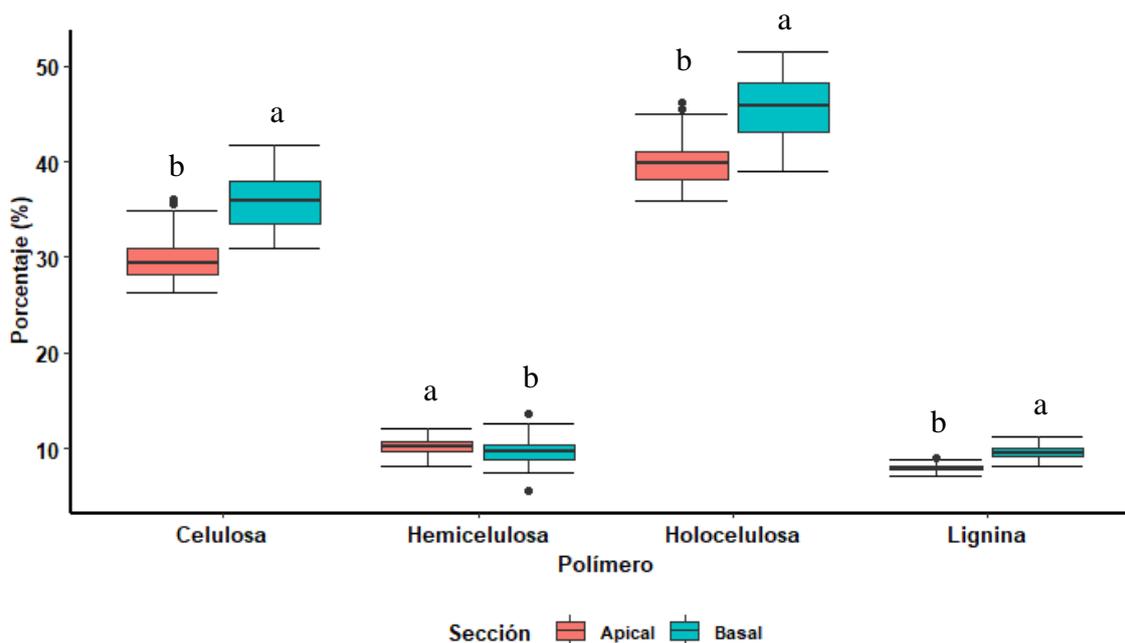


Figura 30: Composición química en la sección apical y basal de Capirona

La Figura 30 muestra las diferencias significativas en la composición de la sección apical y basal del tallo de Capirona. El contenido de lignina, celulosa y holocelulosa fue significativamente superior en la sección basal respecto a la sección apical ($p < 0.05$). En contraste, el contenido de hemicelulosa es significativamente menor en la sección basal en relación con la sección apical ($p < 0.05$).

El mayor contenido de celulosa y lignina en la sección basal del tallo se explica por la mayor densidad en los tejidos basales en comparación con los apicales. Weber y Montes (2005) señalan que la base del tallo de *C. spruceanum* produce madera más densa para contrarrestar el alto estrés de flexión y dar fuerza de soporte a los tejidos. Al contrario, la zona apical del tallo es menos densa, lo que le permite crecer (Weber y Montes, 2005). En *Coffea arabica*

L., de la familia Rubiaceas, se reportó mayor densidad en la base del árbol que en la zona apical (M. Pereira *et al.*, 2014). La densidad está directamente relacionada al peso seco del material vegetal, el cual está conformado principalmente por el peso de celulosa, hemicelulosa y lignina (Maiti *et al.*, 2016). La densidad de la madera se relaciona principalmente con el contenido de celulosa, lignina y sulfuro, que permite formar tejidos más densos (Maiti *et al.*, 2016).

Asimismo, los tejidos conductores pueden estar relacionados a la composición química de los tallos. Se conoce que en el eje de los árboles existe un gradiente de auxinas que actúa como señal morfológica y contribuye a la diferenciación celular. La decreciente concentración de auxinas hacia la base del tallo desencadena el incremento del diámetro de vasos y traqueidas. Es decir, en la base de los tallos se forman traqueidas y vasos con paredes más engrosadas (Fromm, 2013). Estas células contienen mayor concentración de lignina y celulosa en comparación a los tejidos apicales, que presentan vasos y traqueidas con paredes celulares delgadas y en expansión (Fromm, 2013).

4.2. COMPOSICIÓN QUÍMICA DE LA MADERA DE BOLAINA

El contenido de los principales componentes de madera: celulosa, hemicelulosa y lignina en árboles de Bolaina se reporta en la Tabla 7. Los tallos de Bolaina de 1.8 años muestran a celulosa como el componente más abundante tanto en la sección apical y basal (20.44 % y 42.44%), seguido de hemicelulosa (19.31% y 14.25%), mientras que la lignina fue el componente menos abundante (12.25% y 11.22%). Cabe resaltar que el contenido de celulosa en ambas secciones del tallo presenta una variabilidad superior (D.E >2.0) en comparación con lignina y hemicelulosa. La excepción a esa tendencia se presenta en la sección apical de hemicelulosa, la cual tiene una alta variabilidad (D.E = 5.92). En contraste, el contenido de lignina presenta la menor variabilidad, tanto en la sección apical como en la basal. Aunque no se reporte como componente principal de la madera, el contenido de holocelulosa es el más alto tanto en la región apical y basal.

El presente trabajo es el primer estudio en reportar los principales componentes de la madera en árboles juveniles de Bolaina (1.8 años). Las publicaciones científicas reportan la composición química de la madera en árboles maduros de *Guazuma crinita*. Contreras *et al.* (2020) reporta el contenido de lignina de 22-24% y 72-75% para holocelulosa. Estos valores

corresponden a madera extraída y contienen valores superiores de lignina y holocelulosa que los reportados por este trabajo. Por otro lado, Santana y Okino (2007) reportaron el contenido de celulosa (40 a 60%), hemicelulosa (18 a 25%) y lignina (24 a 36%), en 36 especies de árboles amazónicos. Estos resultados coinciden con celulosa y hemicelulosa, mientras que el contenido de lignina es mayor al reportado en este trabajo. En familia Malvaceae se reportó la composición química de la madera de *Gossypium hirsutum* con 18% de lignina y 45.5% de celulosa (Tutus *et al.*, 2010).

Tabla 7: Composición de tallo de Bolaina (*Guazuma crinita*)

	Celulosa (%)	Hemicelulosa (%)	Lignina (%)	Holocelulosa (%)
Sección Apical				
Promedio	20.44	19.31	12.25	39.75
Mínimo	16.48	11.56	7.82	31.54
Máximo	25.83	35.27	15.56	53.77
Desviación estandar	2.03	5.92	1.89	5.48
Coefficiente de variación	0.10	0.31	0.15	0.14
Sección Basal				
Promedio	42.44	14.25	11.22	56.69
Mínimo	38.83	12.42	7.26	52.04
Máximo	48.67	16.53	14.33	63.98
Desviación estandar	2.04	0.86	1.35	2.37
Coefficiente de variación	0.05	0.06	0.12	0.04

La Figura 31 muestra las diferencias composicionales entre la sección apical y basal del tallo de Bolaina, donde **a** es significativamente superior a **b**. El contenido de celulosa y holocelulosa es significativamente mayor en la sección basal comparado a la sección apical ($p < 0.05$). En contraste, el contenido de hemicelulosa y lignina es significativamente menor en la sección basal en relación con la sección apical ($p < 0.05$).

El mayor contenido de celulosa en la sección basal del tallo se explica por la mayor densidad en los tejidos basales en comparación con los apicales. En estadios juveniles *G. crinita* produce madera más densa para contrarrestar el alto estrés de flexión y dar fuerza de soporte a los tejidos (Tuisima-Coral *et al.*, 2017; Weber y Sotelo Montes, 2008). Al contrario, la zona apical del tallo es menos densa, lo que le da flexibilidad para crecer (Weber y Montes, 2005). La densidad se relaciona con la composición química de la madera como lo demostró

Maiti et al. (2016) en 10 especies arbóreas. Especialmente, los componentes como celulosa, lignina y sulfuro contribuyen a formar tejidos más densos (Maiti *et al.*, 2016). Esto coincide con el alto contenido de celulosa encontrado en la base del tallo, pero no explica el mayor contenido de lignina en la sección apical. El contenido de lignina puede incrementarse a modo de defensa cuando las plantas están expuestas a estrés (biótico o abiótico) sin embargo se desconoce el mecanismo que desencadena el cambio del contenido de lignina en situaciones de estrés (Moura *et al.*, 2010).

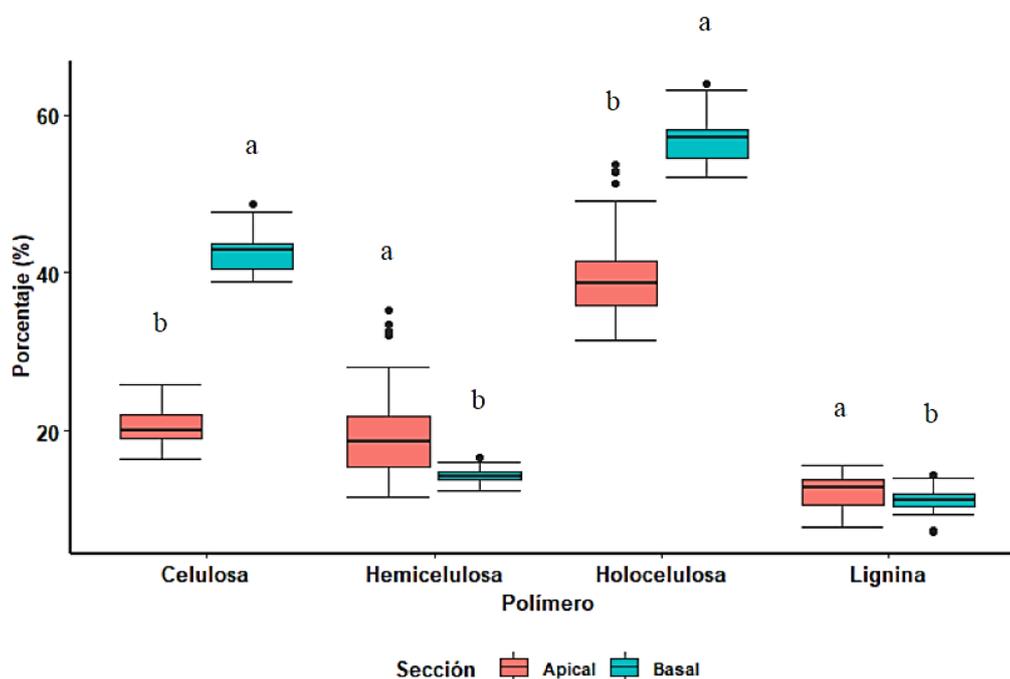


Figura 31: Composición química en la sección apical y basal de Bolaina

4.3. ASIGNACIÓN DE BANDAS EN FTIR

Los espectros FTIR muestran picos que evidencian la presencia de grupos funcionales y enlaces químicos, así como su abundancia, lo que permite analizar la composición de muestras complejas (Coates, 2006; Shurvell, 2006). En el caso de Capirona y Bolaina, el espectro FTIR se usó para determinar la variación composicional entre la sección apical y basal del tallo. En la Figura 32 se muestran los picos de la sección apical y basal de los tallos de Capirona y Bolaina. Asimismo, en la Tabla 8 muestra las frecuencias de los picos y los enlaces asignados de acuerdo con la literatura revisada.

En los espectros FTIR de madera se distinguió dos regiones ricas en información: la región

de frecuencia de grupos ($3800 - 2800 \text{ cm}^{-1}$) y la región de huella dactilar ($1800 - 600 \text{ cm}^{-1}$) (Popescu *et al.*, 2009). Especialmente, esta última región contiene mayor información espectral (Acquah *et al.*, 2016) a modo de picos que indican la presencia de los principales componentes de la madera (Funda *et al.*, 2020).

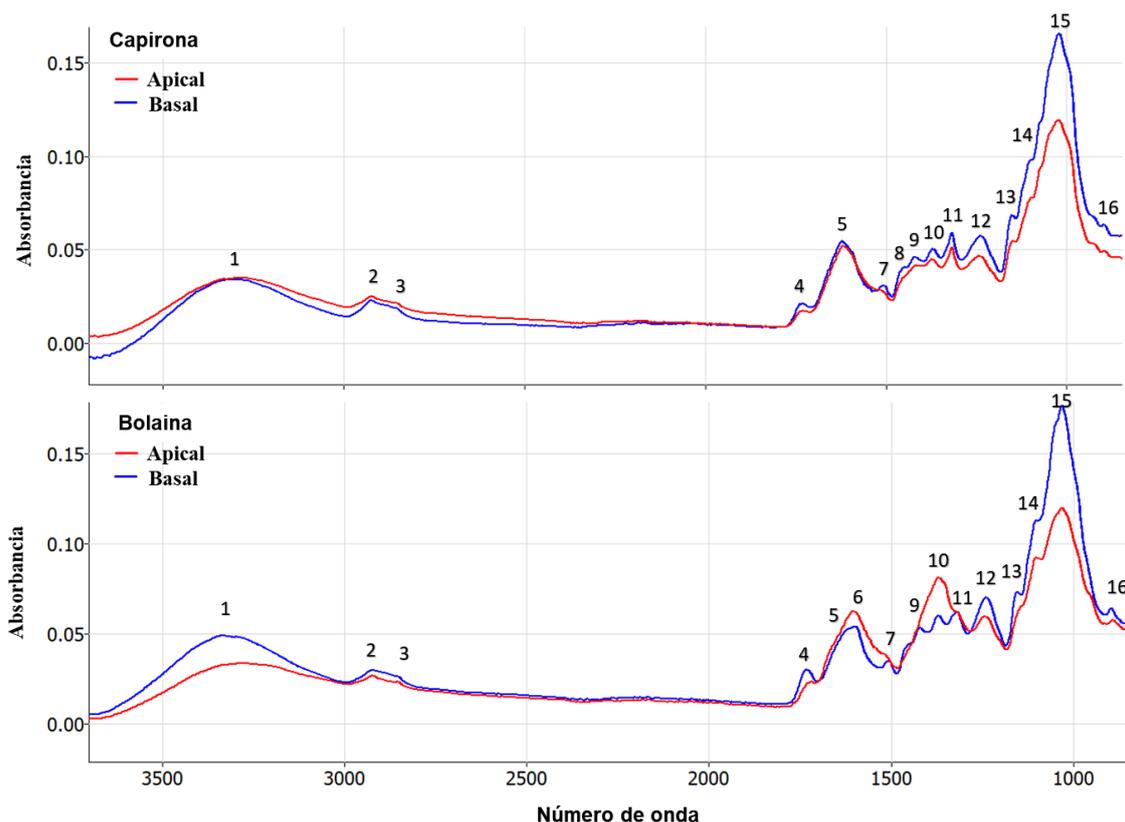


Figura 32: Espectro FTIR de Capirona y Bolaina

En la Figura 32, en la región de frecuencias de grupos ($3800 - 2800 \text{ cm}^{-1}$) se distinguieron los picos 1, 2 y 3; asignados a estiramientos de C-H y O-H, estos enlaces están presentes en todos los componentes de la madera. En la huella dactilar ($1800 - 850 \text{ cm}^{-1}$) se distinguieron: los picos 6 y 7 asignados a la vibración de los anillos aromáticos de lignina y compuestos extractivos (1600 y 1510 cm^{-1}), el pico 4 asignado a estiramiento de enlace C=O en hemicelulosa y compuestos extractivos (1730 cm^{-1}) y los picos 11, 13 y 16 asignados enlaces de celulosa (1320 , 1150 y 897 cm^{-1}) (Popescu *et al.*, 2007; Rana *et al.*, 2010). Los picos 8, 9 y 10 evidencian la vibración de enlaces C-H y O-H en carbohidratos y lignina (1447 , 1420 y 1371 cm^{-1}) (Funda *et al.*, 2020; Traoré *et al.*, 2018; Xu *et al.*, 2013). Finalmente, los picos 12, 14, y 15 (1240 , 1100 y 1032 cm^{-1}) se encuentran en la región de carbohidratos, pero

tienen contribución de enlaces de lignina (Acquah *et al.*, 2016; Chen *et al.*, 2010; Traoré *et al.*, 2018; Zhang *et al.*, 2016).

Tabla 8: Cuadro de frecuencias de Capirona y Bolaina

Pico	Bolaina		Capirona		Enlace en el polímero	Referencia
	Apical	Basal	Apical	Basal		
1	3266	3340	3281	3315	Lignina+Celulosa+Hemicelulosa: estiramiento intra e intermolecular de -OH	Karunakaran et al., 2015. Popescu, 2007
2	2923	2922	2923	2923	Lignina+Celulosa+Hemicelulosa: estiramiento simétrico de C-H	Karunakaran et al., 2015. Popescu, 2007
3	2853	2857	2854	2855	Lignina+Hemicelulosa: estiramiento asimétrico de C-H	Karunakaran et al., 2015. Popescu, 2007
4	1718	1733	1730	1731	Hemicelulosa: estiramiento C=O en xilano. Extractivos: estiramiento C=O de ester carbonilos	Chen et al., 2010. Zhang et al., 2016. Zhou et al., 2015.
5	-	1622	1621	1623	Flavonas y oxalato de calcio: estiramiento de C=O	Zhou et al., 2015.
6	1603	1598	-	-	Lignina: vibración de anillos aromáticos y estiramiento de C=O. Extractivos: vibración de anillos aromáticos	Zhang et al., 2016. Zhou et al., 2015.
7	1518	1508	1516	1511	Lignina: vibración de anillos aromáticos. Extractivos: vibración de amillos benceno	Zhang et al., 2016. Zhou et al., 2015.
8	-	-	-	1447	Celulosa+Lignina+Hemicelulosa: flexión de O-H	Xu et al., 2013
9	-	1422	1417	1419	Lignina: flexión de C-H y estiramiento de anillo aromático. Carbohidratos: Flexión C-H	Popescu et al., 2007; Funda et al., 2020; Chen et al., 2010, Zhou et al., 2015
10	1371	1372	1373	1371	Celulosa y hemicelulosa: flexión de C-H	Popescu, 2007. Zhang et al., 2016.
11	1320	1319	1318	1318	Celulosa: flexión fuera del plano de CH ₂	Popescu, 2007. Traoré et al., 2018
12	1243	1240	1244	1239	Polisacáridos: Estiramiento C-O y deformación O-H	Acquah et al., 2016
13	-	1155	1150	1152	Celulosa: estiramiento asimétrico de C-O-C	Popescu, 2007
14	1104	-	-	1099	Celulosa y hemicelulosa: estiramiento C-O-C. Lignina: Flexión C-H en Guaiacil.	Traoré, 2018, Popescu et al., 2007
15	1032	1032	1024	1021	Holocelulosa: estiramiento de enlace C-O. Lignina: flexión de C-O en alcoholes primarios	Zhang et al., 2016
16	893	897	896	897	Celulosa: flexión de enlace C-H de enlaces β en los anillos de glucosa	Zhang, 2016, Traoré et al., 2018

FUENTE: Elaboración propia.

Los espectros de la sección apical y basal de Capirona y Bolaina presentaron un patrón de picos similar, pero la absorbancia de algunos picos fue mayor en la sección basal (Figura 32). Es así que los picos 12, 13, 14, 15 y 16, correspondientes a la vibración de enlace C-O y C-H en celulosa y hemicelulosa, exhibieron mayor absorbancia en la sección basal que en la apical de ambas especies. Estos picos se encuentran en la región de carbohidratos ubicada entre 1200 a 900 cm⁻¹ (una de las áreas más prominentes del espectro). Estas observaciones concuerdan con los análisis químicos (tabla 4 y 5); donde el contenido de holocelulosa (celulosa + hemicelulosa) fue mayor en la base del tallo (45.6%) que en el ápice (39.9%) en

Capirona al igual que Bolaina con 56.7% en la base y 39.5% en el ápice.

En Capirona, la absorbancia los picos 8, 9, 10 y 11 (correspondientes a enlaces comunes entre carbohidratos y lignina: flexión C-H, O-H y CH₂) es mayor en la sección basal que en la sección apical. Mientras que, en Bolaina los picos 9, 10 y 11 presentaron mayor absorbancia en la sección apical que en la basal con el pico 8 ausente (ver figura 32). Estos resultados concuerdan con los análisis químicos de Bolaina, los cuales mostraron que el contenido de hemicelulosa y lignina es mayor en la sección apical con 19.3% y 12.2% que en la sección basal con 14.2% y 11.2%, respectivamente.

Según la Tabla 8, los picos 6 (1600 cm⁻¹) y 7 (1510 cm⁻¹) son los picos más característicos de lignina y corresponden a vibración de los esqueletos aromáticos (G. Zhou *et al.*, 2011). Estos picos también se atribuyen a compuestos extractivos como flavonas (Zhang *et al.*, 2016). En el caso de Bolaina, el contenido de lignina fue ligeramente mayor en el ápice que en la base del tallo; esta pequeña diferencia (1%) se ve amplificada en los espectros, probablemente debido a la presencia de compuestos extractivos. En contraste, en Capirona el pico 7 tuvo mayor absorbancia en la sección basal, lo que concuerda con el análisis químico (ver tabla 4). Además, se observó que los espectros de Capirona no presentaron el pico 6 definido (anillo aromático de lignina). Por otro lado, el pico 5 ubicado en 1621/1623 cm⁻¹ se atribuyó principalmente a oxalato de calcio y flavonas (Zhang *et al.*, 2016). Este pico es prominente en Capirona mientras que en Bolaina está poco definido y muy cerca al pico 6. Esto podría indicar una fuerte presencia de compuestos extractivos en Capirona.

El pico 4 corresponde a vibración de enlaces C=O en hemicelulosa y compuestos extractivos como ácidos risínicos esterificados, ceras y grasas (Zhang *et al.*, 2016; C. Zhou *et al.*, 2015) y se observó que la absorbancia de la sección basal es mayor en ambas especies. Sin embargo, estas observaciones contradicen los análisis químicos, donde el contenido de hemicelulosa fue mayor en la sección apical del tallo para ambas especies (tabla 4 y 5). La presencia de compuestos extractivos podría explicar estas diferencias debido a que la señal de hemicelulosa podría estar enmascarada por la señal de los compuestos extractivos (Traoré *et al.*, 2018).

4.4. CALIBRACIÓN DEL MODELO DE PREDICCIÓN PLS EN CAPIRONA

Para desarrollar modelos predictivos se empleó la regresión PLS que correlacionó los espectros FTIR con la composición química de la madera. Las muestras (sólidas) fueron analizadas en un espectrofotómetro FTIR – ATR por lo que fueron propensas a fenómenos como: la dispersión de luz (*scattering*), ruido ambiental y variación del paso de la luz (*path leng*) (Butler *et al.*, 2018; Rinnan, Norgaard, *et al.*, 2009). Debido a que estos fenómenos incorporan información poco relevante, causan desplazamiento de línea base y perjudican la linealidad entre la absorbancia y concentración de un analito (ley de Beer) (Butler *et al.*, 2018), se aplicaron pretratamientos que redujeron los efectos de estos fenómenos, mejoraron la calidad y precisión de los espectros y permitió construir modelos predictivos robustos (Butler *et al.*, 2018; Rinnan, Norgaard, *et al.*, 2009).

Es así que se evaluaron 8 pretratamientos en los espectros de Capirona y se consideraron dos regiones espectrales: $3700\text{ cm}^{-1} - 850\text{ cm}^{-1}$ y $1800\text{ cm}^{-1} - 850\text{ cm}^{-1}$. Estos modelos fueron construidos por validación cruzada y los modelos resultantes se muestran en el Anexo 13. Con los 2 mejores pretratamientos obtenidos de la validación cruzada, que fueron 1° y 2° derivada, se construyeron modelos por validación externa (grupo de calibración y validación) en la región $3700\text{ cm}^{-1} - 850\text{ cm}^{-1}$. Los valores de R_c^2 y RMSEC del grupo de calibración y R_p^2 , RMSEP y RPD del grupo de predicción se usaron para evaluar el poder predictivo de estos últimos modelos. La Tabla 7 muestra los modelos construidos por validación cruzada y externa con los 2 mejores pretratamientos en la región $3700\text{ cm}^{-1} - 850\text{ cm}^{-1}$ para celulosa, hemicelulosa, holocelulosa y lignina en Capirona. Los modelos con mayor valor de R^2 , menor RMSEP y mayor RPD son considerados como los mejores modelos según Zhou *et al.* (2015) y Li *et al.* (2015).

El modelamiento de los espectros sin pretratamiento para celulosa, hemicelulosa, holocelulosa y lignina arrojó un R_c^2/R_p^2 entre 0.5 a 0.65. Según He y Hu (2013) estos modelos solo permitirán discriminar muestras con altas y bajas concentraciones (ver tabla 9). (He & Hu, 2013). Asimismo, los valores RPD de celulosa, hemicelulosa y holocelulosa no superaron el valor mínimo de 1.5 para realizar tamizados y predicciones preliminares (Karlinsari *et al.*, 2014). En la Tabla 7 también muestra que los modelos construidos a partir 1° y 2° derivada tuvieron una mejor capacidad predictiva (RPD). Los valores de RPD en los espectros tratados estuvieron en el rango de 1.5 a 2.5, que de acuerdo con Hein (2010) es

suficiente para estimar las propiedades de la madera. Por lo tanto, el pretratamiento de 1° y 2° derivada demostró ser un paso necesario en la construcción de modelos quimiométricos en Capirona. Luego se determinó el número de factores óptimos por el estadístico J propuesto por Gowen et al (2011) para evitar el sobreajuste obtenido por el estadístico F propuesto por Halaand et al (1988) (ver Anexo 13).

Tabla 9: Modelos PLS de celulosa, hemicelulosa, holocelulosa y lignina en Capirona

Compuesto	CELULOSA			HEMICELULOSA			HOLOCELULOSA			LIGNINA		
	NP	1°D	2°D	NP	1°D	2°D	NP	1°D	2°D	NP	1°D	2°D
Factores	4	4	4	7	5	5	4	4	3	4	4	3
R ²	44.9	64.1	65.8	38.9	46.0	36.1	46.4	59.5	56.4	40.0	67.6	70.4
RMSECV	3	2.43	2.37	0.86	0.80	0.87	2.86	2.49	2.59	0.74	0.56	0.53
Factores	4	4	4	6	5	4	4	4	4	4	4	4
R _c ²	51.3	76.4	85.1	43.6	77.3	76.9	52.6	71.4	83.0	57.9	75.8	90.2
RMSEC	2.82	1.97	1.57	0.81	0.52	0.52	2.69	2.10	1.62	0.63	0.48	0.31
R _p ²	51.5	70.2	74.0	39.6	61.1	48.5	59.2	67.6	72.1	57.6	72.0	75.4
RMSEP	2.82	2.21	2.06	0.84	0.68	0.78	2.50	2.23	2.07	0.63	0.52	0.48
RPD _a	1.44	1.83	1.96	1.29	1.60	1.39	1.57	1.76	1.89	1.55	1.89	2.02
RPD _b	1.44	1.85	1.96	1.29	1.60	1.41	1.46	1.77	1.90	1.54	1.90	2.06

NP: no pretratamiento. 1° D: primera derivada. 2°D: segunda derivada. a. Cálculo de RPD según Santos et al., 2016. b. Cálculo de RPD según Li et al., 2015.

A partir de los valores estadísticos mostrados en la Tabla 9, se escogieron los mejores pretratamientos para cada compuesto. En primer lugar, los modelos para celulosa presentaron el mismo número de factores (4) con ambos pretratamientos, sin embargo el modelo con 2° derivada presentó valores inferiores de RMSECV, RMSEC y RMSEP, y valores superiores de R^2 , R_c^2 , R_p^2 , y RPD. En el caso de hemicelulosa, el modelo con 1° derivada presentó los valores más bajos de RMSECV, RMSEC y RMSEP y el mayor RPD, cuyo valor es suficiente para estimar propiedades de madera ($RPD > 1.5$) según Hein et al (2010). Además presenta valores estadísticos cercanos entre el grupo de calibración y validación, lo cual indicia un buen ajuste según Li et al.(2015). Seguidamente, se escogió la 2° derivada como mejor pretratamiento para holocelulosa debido a los valores inferiores de error (RMSEC y RMSEP), altos coeficiente de correlación (R_c^2 y R_p^2) y alto poder predictivo (RPD). Finalmente, ambos pretratamientos en lignina presentaron el mismo número de factores (4), pero el modelo con 2° derivada reportó mejores valores estadísticos (bajo error, alto coeficiente de correlación y RPD).

La Tabla 10 resume los mejores modelos PLS para los componentes de madera de Capirona.. Los modelos PLS para celulosa, hemicelulosa y lignina tienen 4 factores y fueron contruidos con espectros tratados con 2° derivada. Por otro lado, el modelo PLS para hemicelulosa tiene 5 factores y se construyó con espectros tratados con 1° derivada.

Tabla 10: Mejores modelos PLS de celulosa, hemicelulosa, holocelulosa y lignina de Capirona

Compuesto	Mejor pretrat.	Factores	R_c^2	RMSEC	R_p^2	RMSEP	RPD _a	RPD _b
Celulosa	2°D	4	85.1	1.57	74.0	2.06	1.96	1.96
Hemicelulosa	1°D	5	77.3	0.52	61.1	0.68	1.60	1.60
Holocelulosa	2°D	4	83.0	1.62	72.1	2.07	1.89	1.90
Lignina	2°D	4	90.2	0.31	75.4	0.48	2.02	2.06

1°D: primera derivada. 2°D: segunda derivada. RPD_a: cálculo de RPD según Santos et al., 2016. RPD_b: cálculo de RPD según Li et al., 2015.

Los modelos PLS de lignina y hemicelulosa presentaron buena capacidad predictiva (RMSEP bajo), estos modelos predicen el porcentaje de los compuestos con apenas 0.48% y 0.68% de desviación, respectivamente. En el caso de lignina, el modelo también presentó un buen ajuste de datos ($R_c^2 = 90.2$) y alto valor de RPD (2.06), este último valor es indicador de un modelo con muy buena predicción según Karlinasari et al. (2014). Sin embargo, el modelo de hemicelulosa presentó bajo ajuste ($R_c^2 = 77.3$) y un valor de RPD apenas superior a 1.5, el cual indica un modelo útil solo para tamizajes iniciales y predicciones preliminares (Karlinasari *et al.*, 2014). De igual forma, Funda et al. (2020) reportaron un modelo de lignina como de alto valore predictivo (RMSEP = 0.47/0.49) en pino silvestre (*Pinus sylvestris L.*), mientras que el de hemicelulosa fue un modelo con poder predictivo medio a bajo. Igualmente, Zhou et al. (2015) publicaron que el modelo de lignina fue ideal para análisis cuantitativo mientras que el de hemicelulosa fue más apropiado para tamizaje en 4 géneros de árboles de madera dura.

Los modelos PLS de celulosa y holocelulosa fueron buenos, pero no tan eficientes como el modelo de lignina (Figura 33). En particular, los valores de RMSEP fueron altos con 2.06% en celulosa y 2.07% en holocelulosa. Mientras que Funda et al (2020) y Zhou et al (2015) reportaron RMSEP de 0.71% y 0.8% en celulosa, respectivamente. Sin embargo, los valores de R_c^2 de celulosa y holocelulosa se encontraron entre 82 y 90, lo que revela buenas predicciones según He y Hu (2013). En función de RPD, estos modelos se clasificaron como

modelos ideales para tamizados y predicciones iniciales ($RPD = 1.5 - 2$), aunque los valores de RPD de ambos compuestos (1.96 y 1.90) son cercanos a 2, que es el valor mínimo en modelos con buen poder predictivo según Karlinasari et al. (2014). Los valores de RPD reportados por Acquah et al. (2016) fueron de 1.61 y 1.53 para celulosa y holocelulosa usando solo la huella dactilar, estos modelos solo pueden ser usados en tamizados preliminares. Nuestros modelos están en concordancia con Acquah et al. (2016), Funda et al. (2020) y Zhou et al. (2015), quienes reportaron que los modelos PLS construidos con espectros FTIR para carbohidratos no fueron tan buenos como los modelos predictivos para lignina.

A pesar de la baja correlación del modelo de hemicelulosa y los altos valores de RMSEP de celulosa y holocelulosa, todos los valores RPD de estos modelos se encuentran en el rango de 1.5 a 2.5. Por lo tanto, los modelos son aptos para realizar predicción de las propiedades de la madera, según Hein et al (2010) y Karlinasari et al. (2014).

La tendencia a desarrollar mejores modelos predictivos para lignina que para carbohidratos puede deberse a la vibración de los enlaces detectados por FTIR (Acquah *et al.*, 2016). La estructura química de los enlaces de lignina son únicos (Figura 32) mientras que los carbohidratos presentan estructura química semejante (Acquah *et al.*, 2016). Así, los enlaces de celulosa y hemicelulosa comparten picos (1244-1239, 1150, 1099, 1024-1021) en la región de carbohidratos mientras que los picos más representativos de lignina (1600 cm^{-1} y 1510 cm^{-1}) no comparten enlaces con otros polímeros pero sí con compuestos extractivos como flavonas (Zhang *et al.*, 2016). Especialmente en Capirona, el pico en 1600 cm^{-1} parece estar fusionado con el pico en 1621/1623 (atribuido a oxalato de calcio y flavonas según Zhang et al., 2016). A pesar de la interferencia de compuestos extractivos se pudo construir un buen modelo predictivo para lignina.

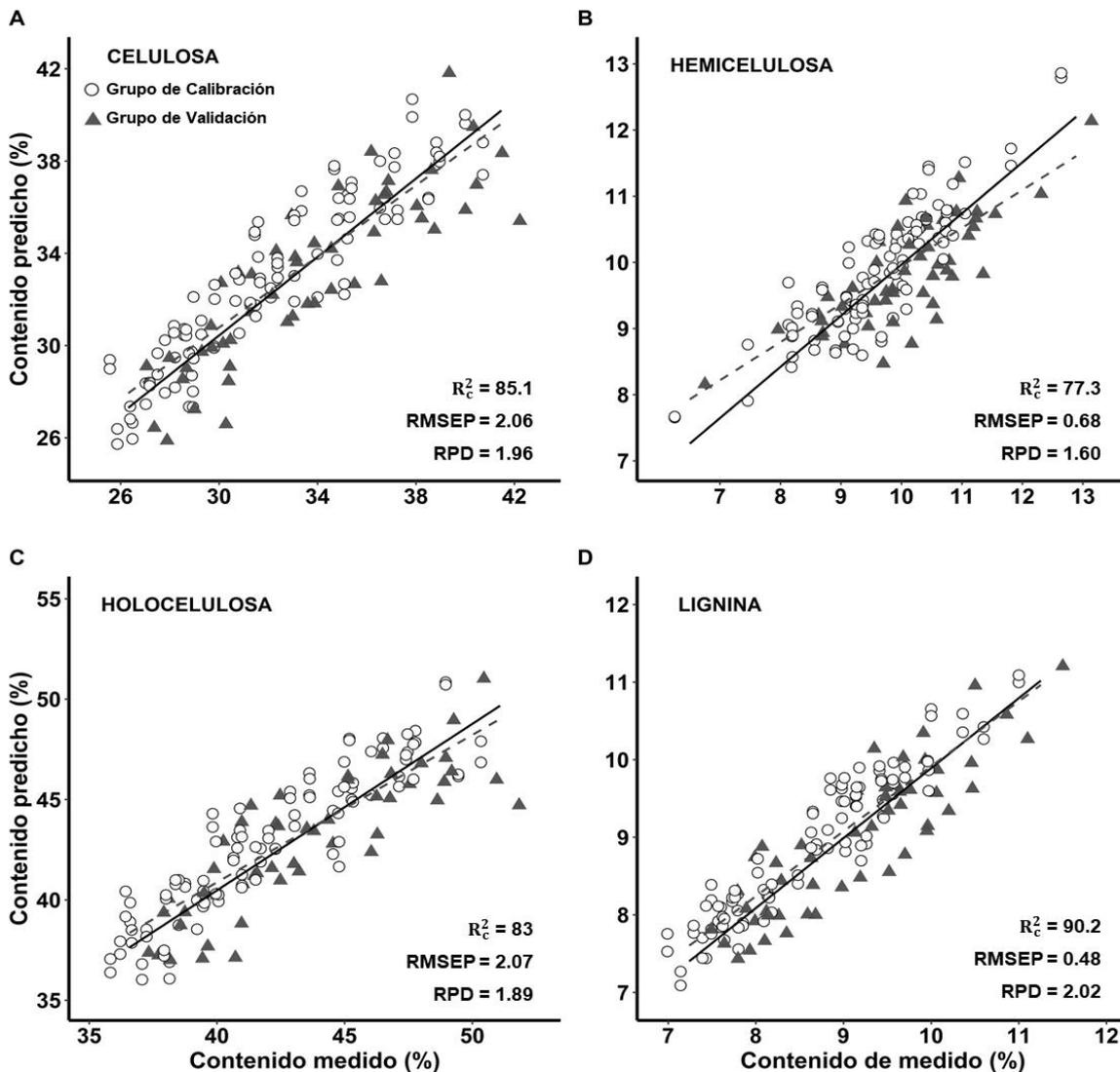


Figura 33: Modelos PLS de celulosa, hemicelulosa, holocelulosa y lignina en Capirona

Por el contrario, los modelos PLS construidos con espectros NIR celulosa y hemicelulosa fueron buenos predictores. Via et al. (2014) obtuvieron modelos PLS de celulosa y hemicelulosa en 4 géneros de árboles de madera dura con valores RPD de 2.04 y 2.72, respectivamente. He y Hu (2013) reportaron un modelo en 116 especies de árboles con un RPD especialmente alto de 5.13 para celulosa. Al contrario, Jiang et al. (2014) reportaron modelos PLS excelentes para lignina y holocelulosa en coníferas. Mientras el modelo de celulosa fue bueno y el de hemicelulosa solo útil para aproximaciones (Jiang *et al.*, 2014). A pesar de reportarse buenos modelos PLS para celulosa y hemicelulosa con espectros NIR, no se puede afirmar que la región NIR es mejor para carbohidratos que para lignina, pero puede ser usada como metodología complementaria.

En relación a la distribución de datos, para obtener modelos robustos se requiere una variación amplia entre los datos obtenidos por el método de referencia, de forma que se evite muestras con concentraciones similares de algún compuesto (Xu *et al.*, 2013). No obstante, una alta variación no siempre significa alta correlación (R^2) de los datos aunque estas medidas están correlacionadas (Kjeldahl y Bro, 2010). En la Figura 33 se observa la distribución de los principales componentes de la madera. En celulosa, holocelulosa y lignina hubo una buena distribución a lo largo de la línea de regresión. En el caso de los datos de hemicelulosa se concentraron en la parte central de la línea de regresión. Como resultado, la correlación de datos en hemicelulosa fue la más baja ($R_c^2 = 77.3$). Lo que se interpreta que el modelo PLS solo explicó el 77.3 % de la variación de datos.

Sin embargo, una mayor correlación de datos (R^2) no significa un mejor modelo. Además, no es un parámetro que sea válido para comparaciones entre modelos, ya que este valor se relaciona con diferentes grupos de datos (Kjeldahl y Bro, 2010). El RMSEP es un parámetro útil para hacer mejores comparaciones. El modelo de hemicelulosa presentó uno de los RMSEP más bajos, por lo que el modelo fue satisfactorio para predicciones al igual que el modelo de lignina. De forma adicional los análisis se complementaron con el parámetro RPD para evaluar la capacidad predictiva de los modelos.

4.5. CALIBRACIÓN DEL MODELO DE PREDICCIÓN PLS EN BOLAINA

Para la calibración de celulosa, hemicelulosa, holocelulosa y lignina en Bolaina se aplicaron 8 pretratamientos a los espectros de las muestras. Los modelos PLS se construyeron por validación cruzada con los espectros pretratados y usando dos regiones: la huella dactilar y toda la región FTIR. Los resultados se muestran en el Anexo 14. Cada modelo fue evaluado por los valores estadísticos RMSECV y R^2 . En la Tabla 11 se muestran los modelos PLS contruidos con los dos mejores pretratamientos mediante validación cruzada y externa.

Los modelos sin pretratamiento (NP) de hemicelulosa y lignina presentaron un alto número de factores (9), lo que con lleva a una reducción de la robustez de futuras predicciones e incorpora ruido al modelo (Faber y Rajkó, 2007; Rinnan, Berg, *et al.*, 2009). Un modelo se considera óptimo cuando tiene el menor número de factores (Acquah *et al.*, 2016). Con la aplicación de pretratamientos, el número de factores de los modelos de hemicelulosa y

lignina se redujo a 4 y 6 respectivamente. Asimismo, los valores de RMSEP se reducen y los valores de R^2 aumentaron (ver Tabla 11).

Tabla 11: Modelos PLS de celulosa, hemicelulosa, holocelulosa y lignina

Compuesto	CELULOSA			HEMICELULOSA			HOLOCELULOSA			LIGNINA		
	NP	MM-N	V-N	NP	1° D	2° D	NP	V-N	2° D	NP	1° D	V-N
Factores	5	4	5	9	4	4	4	6	3	9	6	6
R ²	89.8	97	97.4	83.3	74.2	76.5	78.7	85.9	81.9	67.5	68.9	66.9
RMSECV	3.6	1.94	1.8	2	2.49	2.37	4.35	3.54	4.02	0.97	0.96	0.99
Factores	5	4	5	5	5	6	3	6	4	7	6	6
R _c ²	90.9	97.3	97.7	73.9	85.1	90.5	79.2	89.3	92	67.7	82.6	74.1
RMSEC	3.37	1.83	1.68	2.5	1.98	1.51	4.31	3.09	2.67	0.97	0.72	0.87
R _p ²	91	97.3	97.7	74.6	81.3	83.1	79.3	88.8	86.7	66.5	77.7	72.9
RMSEP	3.34	1.82	1.68	2.47	2.12	2.02	4.3	3.16	3.45	0.99	0.81	0.89
RPD _a	3.35	6.14	6.67	1.98	2.31	2.43	2.2	2.99	2.74	1.73	2.12	1.92
RPD _b	3.34	6.14	6.67	1.98	2.32	2.44	2.20	2.99	2.74	1.73	2.12	1.93

NP: no pretratamiento. MMN: mínimo máximo de normalización. V-N: vector de normalización. 1°D: primera derivada. 2°D: segunda derivada. RPD_a: cálculo de RPD según Santos et al., 2016. RPD_b: cálculo de RPD según Li et al., 2015.

Por otro lado, los modelos de celulosa y holocelulosa sin pretratamiento presentaron valores de RPD mayores a 2, lo que indica buen poder predictivo. Los espectros FTIR de celulosa y holocelulosa sin pretratamiento presentaron la sensibilidad suficiente para su detección en Bolaina. La aplicación de pretratamientos mejoró aún más los modelos PLS, se redujeron los valores de RMSEP y aumentaron los de R^2 en los 4 compuestos.

En función de los valores estadísticos presentados en la Tabla 11 se escogió el mejor pretratamiento para cada compuesto. En primer lugar, los modelos de celulosa presentaron valores estadísticos cercanos (R_c^2 , R_p^2 , RMSEC, RMSEP y RPD) con ambos pretratamientos, pero se eligió mínimo máximo de normalización como el mejor por presentar el menor número de factores (4). Seguidamente, el modelo con 1° derivada fue escogido para hemicelulosa por presentar el menor número de factores (5) y tener valores estadísticos cercanos entre el grupo de calibración y validación, lo cual indica un buen ajuste según Li et al.(2015). En el caso de holocelulosa, el vector de normalización fue escogido como el mejor pretratamiento a pesar de presentar el mayor número de factores (6) pero con mayor valor de RPD, menor RMSEP y valores estadísticos cercanos entre el grupo de calibración y validación. Finalmente, el modelo de lignina con 1° derivada presentó los valores más altos de R_c^2/R_p^2 , bajo RMSEC/RMSEP y un alto RPD que el modelo con vector de normalización. En la Tabla 12 se resumen los mejores pretratamientos para cada compuesto. SE decidió usar solo la validación externa por contener mejores valores.

Tabla 12: Mejores modelos PLS de celulosa, hemicelulosa, holocelulosa y lignina en Bolaina

Compuesto	Mejor pretrat.	Factores	R_c^2	RMSEC	R_p^2	RMSEP	RPD _a	RPD _b
Celulosa	MM-N	4	97.3	1.83	97.3	1.82	6.14	6.14
Hemicelulosa	1° D	5	85.1	1.98	81.3	2.12	2.31	2.32
Holocelulosa	V-N	6	89.3	3.09	88.8	3.16	2.99	2.99
Lignina	1° D	6	82.6	0.72	77.7	0.81	2.12	2.12

MM-N: mínimo máximo de normalización. V-N: vector de normalización. 1°D: primera derivada. RPD_a: cálculo de RPD según Santos et al., 2016. RPD_b: cálculo de RPD según Li et al., 2015.

El modelo PLS de lignina presentó una buena capacidad predictiva. En este modelo se reportó el RMSEP = 0.81% más bajo entre los demás compuestos y valor RPD mayor a 2, este último valor indicó un buen poder predictivo según Karlinasari et al. (2014). Además, el modelo presentó buen ajuste de datos ($R_c^2 = 82.6$), que según Hein et al. (2010) describe un modelo eficiente para predicciones. De forma similar, otros autores concluyen que los modelos PLS de lignina construidos con espectros FTIR tuvieron buena capacidad predictiva y que puede aplicar a un gran número de muestras (Acquah *et al.*, 2016; C. Zhou *et al.*, 2015; G. Zhou *et al.*, 2011). Estos autores reportaron valores de $R_c^2 \approx 90 - 85$ y RMSEP $\approx 0.80\%$ similares a los presentados en este trabajo, sin embargo, valores de RPD (entre 1.4 y 1.72) más bajos.

En el caso de hemicelulosa y holocelulosa, los modelos presentaron mejor capacidad predictiva (RPD) y ajuste de datos (R_c^2) que lignina (figura 35). Sin embargo, la desviación de predicción (RMSEP) de ambos modelos fue mayor que el error en lignina y el error reportador por otros autores (Funda et al., 2020; Jiang et al., 2014 y Zhou et al., 2015). Los modelos presentados de hemicelulosa y holocelulosa revelan buenas predicciones en función del ajuste de datos ($R_c^2 = 82 - 91$) según He y Hu (2013). Por otro lado, la capacidad predictiva en holocelulosa (RPD= 2.99) es muy eficiente mientras que la de hemicelulosa (RPD=2.32) es buena según Karlinasari et al. (2014), ver tabla 10. Otros autores como Zhou et al. (2015) y Acquah et al. (2016) reportaron menores valores de RPD (2.06, 0.85 y 1.45) y consideraron a los modelos aptos solo para tamizaje. Sin embargo los modelos en NIR tuvieron buena capacidad predictiva en hemicelulosa (Via *et al.*, 2014) y excelente en holocelulosa (Jiang *et al.*, 2014).

El modelo de celulosa tiene excelente capacidad predictiva y ajuste de datos (valores más altos de RPD y R_c^2). Sin embargo, el modelo presenta un error alto (RMSEP =1.82%), comparado con el error reportado por Funda et al. (2020), He y Hu (2013), Via et al. (2014) y Jiang et al. (2014). Por otro lado, el valor de RPD =6.14 fue mayor que los valores reportados por los autores previamente mencionados.

En la evaluación de modelos se priorizó el valor de RMSEP y RPD en lugar de R^2 para comparar los modelos. Kjeldahl y Bro (2010) señalaron que un alto valor de R^2 no significa un mejor modelo. La varianza explicada por el modelo (R^2) no solo depende de la eficiencia del mismo sino también de la distribución de los datos. Debido a que este valor está en función del grupo de datos no debe ser usado para comparar modelos. En particular, el modelo de celulosa presentó la mayor variación de datos, lo que conllevó a un alto coeficiente de correlación (R^2) pero con alto error. El parámetro RMSEP es más útil para evaluar un modelo y compararlo con otros (Kjeldahl y Bro, 2010). En ese sentido, los modelos de celulosa, hemicelulosa y holocelulosa presentaron el RMSEP alto. Mientras que el RMSEP más bajo fue reportado en el modelo de lignina. De forma adicional se usó el parámetro RPD para evaluar los modelos PLS de los 4 componentes de Bolaina. Todos los modelos presentaron RPD mayor a 1.5, valor mínimo para estimar las propiedades de maderas según según Hein et al (2010) y Karlinasari et al. (2014).

V. CONCLUSIONES

1. La composición de los principales componentes de la madera como celulosa, hemicelulosa, holocelulosa y lignina es significativamente diferente entre la sección apical y basal de los tallos de Capirona y Bolaina.
2. Los picos del espectro FTIR de Capirona y Bolaina evidenciaron la presencia de anillos aromáticos de lignina (picos 5, 6 y 7), enlaces C=O característico de hemicelulosa (pico 4), enlaces C-H y O-H en holocelulosa y lignina (picos 8, 9, 10), enlaces C-O y C-H de celulosa (picos 11, 13 y 16) y enlaces C-O-C y C-H en holocelulosa (picos 12, 13, 14, 15 y 16).
3. En Capirona, debido a los valores del modelo de lignina (bajo RMSEP y $RPD > 2.0$) se le consideró como un buen modelo predictivo. Mientras que los modelos de celulosa, hemicelulosa y holocelulosa presentaron poder predictivo suficiente ($RPD > 1.5$) para la cuantificación del contenido químico en madera.
4. En Bolaina, el modelo de celulosa tiene excelente capacidad predictiva ($R^2 > 0.91$ y $RPD > 3.0$). Asimismo, los modelos PLS de hemicelulosa y lignina presentaron buena capacidad predictiva ($RPD > 2.0$) y en el modelo de holocelulosa la predicción fue eficiente ($RPD > 3.0$).
5. En Bolaina, los modelos PLS de hemicelulosa, holocelulosa y lignina se consideran modelos con buena capacidad predictiva ($0.81 < R^2 < 0.90$ y $RPD > 2.0$); y el modelo de celulosa, excelente.
6. El FTIR-ATR junto a PLS demostró ser un modelo predictivo económico y preciso para la determinación de la composición química de la madera de Capirona y Bolaina en estadios juveniles.

VI. RECOMENDACIONES

1. Se recomienda el uso de más muestras que contengan la mayor variabilidad posible para construir modelos más robustos.
2. En estudios posteriores, se puede cuantificar el contenido de compuestos extractivos, contenido energético, reactividad termal de Capirona y Bolaina por modelos PLS en base a espectros FTIR.
3. Desarrollo de método para determinar las propiedades físicas de la madera en plantas jóvenes de Capirona y Bolaina, como criterio adicional para la selección del material.
4. Desarrollo de métodos para la determinación química de madera en estadios de 6 meses en árboles.

VII. BIBLIOGRAFÍA

- Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *WIREs Computational Statistics*, 2(1), 97–106. <https://doi.org/10.1002/wics.51>
- Acquah, G. E., Via, B. K., Fasina, O. O., & Eckhardt, L. G. (2016). Rapid Quantitative Analysis of Forest Biomass Using Fourier Transform Infrared Spectroscopy and Partial Least Squares Regression. *Journal of Analytical Methods in Chemistry*, 2016, 1–10. <https://doi.org/10.1155/2016/1839598>
- Akon-Technology. (2016). *Determining Acid Detergent Lignin in beakers*. Akon Technology. https://www.ankom.com/sites/default/files/document-files/Method_8_Lignin_in_beakers.pdf
- Akon-Technology. (2017). *Acid Detergent Fiber in Feeds Filter Bag Technique (for A200 and A200I)*. AKON Technology. https://www.ankom.com/sites/default/files/document-files/Method_5_ADF_A200.pdf
- Akon-Techonology. (2017). *Neutral Detergent Fiber in Feeds-Filter Bag Technique (for A200 and A200I)*. AKON Technology. https://www.ankom.com/sites/default/files/document-files/Method_6_NDF_A200.pdf
- Alonso-Simón, A., García-Angulo, P., Mérida, H., Encina, A., Álvarez, J. M., & Acebes, J. L. (2011). The use of FTIR spectroscopy to monitor modifications in plant cell wall architecture caused by cellulose biosynthesis inhibitors. *Plant Signaling & Behavior*, 6(8), 1104–1110. <https://doi.org/10.4161/psb.6.8.15793>
- Aróstegui, A. (1987). Informe Técnico Semestral. Proyecto “Estudios Básicos Aplicados de la Madera, estudio propiedades físicas y mecánicas de madera de selva baja.” In *Centro Forestal XII: Vol. I*.
- Barros, J., Escamilla-trevino, L., Song, L., Rao, X., Serrani-yarce, J. C., Palacios, M. D., Engle, N., Choudhury, F. K., Tschaplinski, T. J., Venables, B. J., Mittler, R., & Dixon, R. A. (2019). 4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase. *Nature Communications*, 10(1994), 1–11. <https://doi.org/https://doi.org/10.1038/s41467-019-10082-7>

- Barros, J., Serk, H., Granlund, I., & Pesquet, E. (2015). The cell biology of lignification in higher plants. *Annals of Botany*, *115*(7), 1053–1074. <https://doi.org/10.1093/aob/mcv046>
- Baucher, M., Monties, B., Montagu, M. Van, & Boerjan, W. (1998). Biosynthesis and Genetic Engineering of Lignin. *Critical Reviews in Plant Sciences*, *17*(2), 125–197. <https://doi.org/10.1080/07352689891304203>
- Blanco, M., & Alcalà, M. (2009). Multivariate Calibration for Quantitative Analysis. In *Infrared Spectroscopy for Food Quality Analysis and Control* (pp. 51–82). Elsevier. <https://doi.org/10.1016/B978-0-12-374136-3.00003-1>
- Boerjan, W., Ralph, J., & Baucher, M. (2003). Lignin Biosynthesis. *Annual Review of Plant Biology*, *54*(1), 519–546. <https://doi.org/10.1146/annurev.arplant.54.031902.134938>
- Brett, C. T. (2000). Cellulose Microfibrils in Plants : Biosynthesis , Deposition , and Integration into the Cell Wall. *International Review of Cytology*, *199*, 161–199.
- Brown, R. M. (2004). Cellulose Structure and Biosynthesis: What is in Store for the 21st Century? *Journal of Polymer Science, Part A: Polymer Chemistry*, *42*(3), 487–495. <https://doi.org/10.1002/pola.10877>
- Butler, H. J., Smith, B. R., Fritsch, R., Radhakrishnan, P., Palmer, D. S., & Baker, M. J. (2018). Optimised spectral pre-processing for discrimination of biofluids via ATR-FTIR spectroscopy. *The Analyst*, *143*(24), 6121–6134. <https://doi.org/10.1039/C8AN01384E>
- Camel, V. F., Quispe-Melgar, H. R., Ames-Martínez, F. N., Navarro Romo, W. C., Segovia-Salcedo, M. C., & Kessler, M. (2019). Estructura forestal de tres especies endémicas del género *Polylepis* (Rosaceae) en la Región Central del Perú. *Ecología Austral*, *29*(3), 285–295. <https://doi.org/10.25260/EA.19.29.3.0.812>
- Chen, H., Ferrari, C., Angiuli, M., Yao, J., Raspi, C., & Bramanti, E. (2010). Qualitative and quantitative analysis of wood samples by Fourier transform infrared spectroscopy and multivariate analysis. *Carbohydrate Polymers*, *82*(3), 772–778. <https://doi.org/10.1016/j.carbpol.2010.05.052>
- Christian, G. (2009). *Química Analítica* (M.-H. EDITORES (ed.); Sexta Edic). Mc Graw Hill.
- Coates, J. (2006). Interpretation of Infrared Spectra, A Practical Approach. *Encyclopedia of Analytical Chemistry*, 1–23. <https://doi.org/10.1002/9780470027318.a5606>
- Contreras, A., Cárdenas, A., & Gonzáles, H. (2020). Caracterización física y mecánica de compuestos de Guazuma crinita Mart. a base de polipropileno virgen. *Revista Mexicana*

<https://doi.org/https://https://doi.org/10.29298/rmcf.v11i57.621>

- Cornelius, J. P., Pinedo-Ramírez, R., Sotelo Montes, C., Ugarte-Guerra, L. J., & Weber, J. C. (2018). Efficiency of early selection in *Calycophyllum spruceanum* and *Guazuma crinita*, two fast-growing timber species of the Peruvian Amazon. *Canadian Journal of Forest Research*, 48(5), 517–523. <https://doi.org/10.1139/cjfr-2017-0407>
- Ellwanger, J. H., Kulmann-LEAL, B., Kaminski, V. L., Valverde-Villegas, J. M., Veiga, A. B. G. DA, Spilki, F. R., Fearnside, P. M., Caesar, L., Giatti, L. L., Wallau, G. L., Almeida, S. E. M., Borba, M. R., Hora, V. P. DA, & Chies, J. A. B. (2020). Beyond diversity loss and climate change: Impacts of Amazon deforestation on infectious diseases and public health. *Anais Da Academia Brasileira de Ciências*, 92(1), 1–33. <https://doi.org/10.1590/0001-3765202020191375>
- Faber, N. M., & Rajkó, R. (2007). How to avoid over-fitting in multivariate calibration-The conventional validation approach and an alternative. *Analytica Chimica Acta*, 595(1-2 SPEC. ISS.), 98–106. <https://doi.org/10.1016/j.aca.2007.05.030>
- FAO, & UNEP. (2020). *The State of the World's Forests 2020*. FAO and UNEP. <https://doi.org/10.4060/ca8642en>
- Festucci-Buselli, R. A., Otoni, W. C., & Joshi, C. P. (2007). Structure, organization, and functions of cellulose synthase complexes in higher plants. *Brazilian Journal of Plant Physiology*, 19(1), 1–13. <https://doi.org/10.1590/S1677-04202007000100001>
- FitzPatrick, M., Champagne, P., & Cunningham, M. F. (2012). Quantitative determination of cellulose dissolved in 1-ethyl-3-methylimidazolium acetate using partial least squares regression on FTIR spectra. *Carbohydrate Polymers*, 87(2), 1124–1130. <https://doi.org/10.1016/j.carbpol.2011.08.086>
- Flores, Y. (2007). *Bolaina blanca (Guazuma crinita Mart)* (pp. 1–8). Instituto Nacional de Investigación Agraria.
- Fromm, J. (2013). Cellular Aspects of Wood Formation. In J. Fromm (Ed.), *Plant Cell Monographs* (Vol. 20). Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-36491-4>
- Funda, T., Fundova, I., Gorzsás, A., Fries, A., & Wu, H. X. (2020). Predicting the chemical composition of juvenile and mature woods in Scots pine (*Pinus sylvestris* L.) using FTIR spectroscopy. *Wood Science and Technology*, 54(2), 289–311. <https://doi.org/10.1007/s00226-020-01159-4>
- Galarza, E., & Serna, K. La. (2005). ¿Son sostenibles las concesiones forestales en el Perú?

Economía y Sociedad, 56, 34–41.

- Gautam, R., Vanga, S., Ariese, F., & Umaphathy, S. (2015). Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Techniques and Instrumentation*, 2(1). <https://doi.org/10.1140/epjti/s40485-015-0018-6>
- Geladi, P. (2003). Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 58(5), 767–782. [https://doi.org/10.1016/S0584-8547\(03\)00037-5](https://doi.org/10.1016/S0584-8547(03)00037-5)
- Gibb, R., Redding, D. W., Chin, K. Q., Donnelly, C. A., Blackburn, T. M., Newbold, T., & Jones, K. E. (2020). Zoonotic host diversity increases in human-dominated ecosystems. *Nature*, 584(7821), 398–402. <https://doi.org/10.1038/s41586-020-2562-8>
- Gowen, A. A., Downey, G., Esquerre, C., & O'Donnell, C. P. (2011). Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients. *Journal of Chemometrics*, 25(7), 375–381. <https://doi.org/10.1002/cem.1349>
- Haaland, D. M., & Thomas, E. V. (1988). Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Analytical Chemistry*, 60(11), 1193–1202. <https://doi.org/10.1021/ac00162a020>
- Harakava, R. (2005). Genes encoding enzymes of the lignin biosynthesis pathway in Eucalyptus. *Genetics and Molecular Biology*, 28(3 suppl), 601–607. <https://doi.org/10.1590/S1415-47572005000400015>
- He, W., & Hu, H. (2013). Prediction of hot-water-soluble extractive, pentosan and cellulose content of various wood species using FT-NIR spectroscopy. *Bioresource Technology*, 140, 299–305. <https://doi.org/10.1016/j.biortech.2013.04.115>
- Héberger, K. (2008). Chemoinformatics—multivariate mathematical–statistical methods for data evaluation. In *Medical Applications of Mass Spectrometry* (pp. 141–169). Elsevier. <https://doi.org/10.1016/B978-044451980-1.50009-4>
- Hecht, S. B., Morrison, K. D., & Padoch, C. (2014). The Social Lives of Forests. In *Journal of Physics A: Mathematical and Theoretical* (Vol. 44, Issue 8). University of Chicago Press. <https://doi.org/10.7208/chicago/9780226024134.001.0001>
- Hein, P. R. G. (2010). Multivariate regression methods for estimating basic density in Eucalyptus wood from near infrared spectroscopic data. *Cerne, Lavras*, 16, 90–96.
- Holtzapple, M. T. (2003). Hemicelluloses. In B. Caballero, P. Finglas, & F. Toldra (Eds.), *Encyclopedia of Food Sciences and Nutrition* (pp. 3060–3071). Elsevier.

<https://doi.org/10.1016/B0-12-227055-X/00589-7>

- Hubert, M., & Branden, K. Vanden. (2003). Robust methods for partial least squares regression. *Journal of Chemometrics*, 17(10), 537–549. <https://doi.org/10.1002/cem.822>
- IIAP. (2009). Evaluación económica de parcelas de regeneración natural y plantaciones de Bolaina Blanca, Guazuma crinita, en el VRAEM - AYNA San Francisco. In *Avances Economicos N°11* (primera ed).
- Jiang, W., Han, G., Via, B. K., Tu, M., Liu, W., & Fasina, O. (2014). Rapid assessment of coniferous biomass lignin-carbohydrates with near-infrared spectroscopy. *Wood Science and Technology*, 48(1), 109–122. <https://doi.org/10.1007/s00226-013-0590-3>
- Jin, X., Chen, X., Shi, C., Li, M., Guan, Y., Yu, C. Y., Yamada, T., Sacks, E. J., & Peng, J. (2017). Determination of hemicellulose, cellulose and lignin content using visible and near infrared spectroscopy in *Miscanthus sinensis*. *Bioresource Technology*, 241(May 2017), 603–609. <https://doi.org/10.1016/j.biortech.2017.05.047>
- Karlinasari, L., Sabed, M., Wistara, I. N. J., & Purwanto, Y. A. (2014). Near infrared (NIR) spectroscopy for estimating the chemical composition of (*Acacia mangium* Willd.) wood. *Journal of the Indian Academy of Wood Science*, 11(2), 162–167. <https://doi.org/10.1007/s13196-014-0133-z>
- Karunakaran, C., Christensen, C. R., Gaillard, C., Lahlali, R., Blair, L. M., Perumal, V., Miller, S. S., & Hitchcock, A. P. (2015). Introduction of Soft X-Ray Spectromicroscopy as an Advanced Technique for Plant Biopolymers Research. *PLOS ONE*, 10(3), e0122959. <https://doi.org/10.1371/journal.pone.0122959>
- Kelly, J. G., Trevisan, J., Scott, A. D., Carmichael, P. L., Pollock, H. M., Martin-Hirsch, P. L., & Martin, F. L. (2011). Biospectroscopy to metabolically profile biomolecular structure: a multistage approach linking computational analysis with biomarkers. *Journal of Proteome Research*, 10(4), 1437–1448. <https://doi.org/10.1021/pr101067u>
- Kjeldahl, K., & Bro, R. (2010). Some common misunderstandings in chemometrics. *Journal of Chemometrics*, 24(7–8), 558–564. <https://doi.org/10.1002/cem.1346>
- Kubo, S., & Kadla, J. F. (2005). Hydrogen bonding in lignin: A fourier transform infrared model compound study. *Biomacromolecules*, 6(5), 2815–2821. <https://doi.org/10.1021/bm050288q>
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer New York. <https://doi.org/10.1007/978-1-4614-6849-3>
- L. G. Wade, J. (2004). Carbohidratos y ácidos nucleicos. In *Química Analítica* (Quinta, p.

- 1296). PEARSON EDUCATION S.A.
- Lee, L. C., Liong, C.-Y., & Jemain, A. A. (2017). A contemporary review on Data Preprocessing (DP) practice strategy in ATR-FTIR spectrum. *Chemometrics and Intelligent Laboratory Systems*, *163*, 64–75. <https://doi.org/10.1016/j.chemolab.2017.02.008>
- Li, S., Bashline, L., Lei, L., & Gu, Y. (2014). Cellulose Synthesis and Its Regulation. *The Arabidopsis Book*, *12*, e0169. <https://doi.org/10.1199/tab.0169>
- Li, X., Sun, C., Zhou, B., & He, Y. (2015). Determination of Hemicellulose, Cellulose and Lignin in Moso Bamboo by Near Infrared Spectroscopy. *Scientific Reports*, *5*(1), 17210. <https://doi.org/10.1038/srep17210>
- Liu, Q., Luo, L., & Zheng, L. (2018). Lignins: Biosynthesis and Biological Functions in Plants. *International Journal of Molecular Sciences*, *19*(2), 335. <https://doi.org/10.3390/ijms19020335>
- Lohmöller, J.-B. (1989). *Latent Variable Path Modeling with Partial Least Squares*. Germany: Physica-Verlag. <https://www.springer.com/gp/book/9783642525148>
- Maiti, R., Gonzalez Rodriguez, H., & Kumari, A. (2016). Wood Density of Ten Native Trees and Shrubs and Its Possible Relation with a Few Wood Chemical Compositions. *American Journal of Plant Sciences*, *07*(08), 1192–1197. <https://doi.org/10.4236/ajps.2016.78114>
- Mcnamara, J. T., Morgan, J. L. W., & Zimmer, J. (2016). A Molecular Description of Cellulose Biosynthesis. *Annual Review of Biochemistry*, *84*, 895–921. <https://doi.org/10.1146/annurev-biochem-060614-033930.A>
- Moura, J. C. M. S., Bonine, C. A. V., de Oliveira Fernandes Viana, J., Dornelas, M. C., & Mazzafera, P. (2010). Abiotic and biotic stresses and changes in the lignin content and composition in plants. *Journal of Integrative Plant Biology*, *52*(4), 360–376. <https://doi.org/10.1111/j.1744-7909.2010.00892.x>
- Orrego Medina, D. R., & Bustamante Guillén, N. R. (2017). Trabajabilidad de la madera de capirona (*Calycophyllum spruceanum*) procedente de plantaciones de la cuenca del río Aguaytia en la región de Ucayali – Perú. *Revista Forestal Del Perú*, *32*(2), 97. <https://doi.org/10.21704/rfp.v32i2.1041>
- Otarola, E., & Martinez, P. (2007). Análisis de rentabilidad económica y desarrollo de ecuaciones alométricas de los bosques aluviales de *Calycophyllum spruceanum* (BENTHAM) HOOKER F. ex SCHUMANN “Capirona” para determinación del valor maderable y del carbono almacenado para servicios de RED. *Folia Amazónica*, *16*,

121–135.

- Pandey, K. K. (1999). A study of chemical structure of soft and hardwood and wood polymers by FTIR spectroscopy. *Journal of Applied Polymer Science*, 71(12), 1969–1975. [https://doi.org/10.1002/\(SICI\)1097-4628\(19990321\)71:12<1969::AID-APP6>3.3.CO;2-4](https://doi.org/10.1002/(SICI)1097-4628(19990321)71:12<1969::AID-APP6>3.3.CO;2-4)
- Paredez, A. R., Somerville, C. R., & Ehrhardt, D. w. (2006). Visualization of Cellulose Synthase with Microtubules. *Science*, 312(5779), 1491–1495. <https://doi.org/10.1126/science.1126551>
- Pasquini, C. (2003). Review Near Infrared Spectroscopy : Fundamentals , Practical Aspects and Analytical Applications. *Sociedade Brasileira de Química*, 14(2), 198–219.
- Pauly, M., Gille, S., Liu, L., Mansoori, N., de Souza, A., Schultink, A., & Xiong, G. (2013). Hemicellulose biosynthesis. *Planta*, 238(4), 627–642. <https://doi.org/10.1007/s00425-013-1921-1>
- Peak, D. (2005). Fourier transform infrared spectroscopy. *Elsevier*, 80–85.
- Pereira, H., Graca, J., & Rodrigues, J. C. (2004). Wood Chemistry in Relation to Quality. *ChemInform*, 35(46). <https://doi.org/10.1002/chin.200446298>
- Pereira, M. A., Silva, J. R. M. da, Rabelo, G. F., & Andrade, A. C. de A. (2014). Propriedades anatômicas, químicas e de densidade da madeira de Coffea arabica L. *CERNE*, 20(3), 363–367. <https://doi.org/10.1590/0104776020142003833>
- Plomion, C., Stokes, A., France, C. P., & Rhe, L. De. (2018). Wood Formation in Trees 1. *Plant Physiology*, 127(December), 1513–1523. <https://doi.org/10.1104/pp.010816.1>
- Popescu, C.-M., Popescu, M.-C., Singurel, G., Vasile, C., Argyropoulos, D. S., & Willfor, S. (2007). Spectral Characterization of Eucalyptus Wood. *Applied Spectroscopy*, 61(11), 1168–1177. <https://doi.org/10.1366/000370207782597076>
- Popescu, C.-M., Singurel, G., Popescu, M.-C., Vasile, C., Argyropoulos, D. S., & Willför, S. (2009). Vibrational spectroscopy and X-ray diffraction methods to establish the differences between hardwood and softwood. *Carbohydrate Polymers*, 77(4), 851–857. <https://doi.org/10.1016/j.carbpol.2009.03.011>
- Putzel, L., Cronkleton, P., Larson, A., Pinedo-Vasquez, Salazar, O., & Sears, R. (2013). Peruvian smallholder production and marketing of bolaina (*Guazuma crinita*), a fast-growing Amazonian timber species : call for a pro-livelihoods policy environment. In *Center for International Forestry Research Peruvian*. <https://doi.org/10.17528/cifor/004257>
- Ralph, J., Lundquist, K., Brunow, G., Lu, F., Kim, H., Paul, F., Marita, J. M., Hatfield, R.

- D., Ralph, S. A., Christensen, J. H., & Boerjan, W. (2004). Lignins : Natural polymers from oxidative coupling of 4-hydroxyphenyl- propanoids. *Phytochemistry Reviews*, 3(January), 29–60. <https://doi.org/10.1023/B>
- Rana, R., Langenfeld-heyser, R., Finkeldey, R., & Polle, A. (2010). Characterisation of wood and lignin of five tropical timber wood species of the family Dipterocarpaceae. *Wood Science and Technology*, 44, 225–242. <https://doi.org/10.1007/s00226-009-0281-2>
- Ren, J.-L., & Sun, R.-C. (2017). Polysaccharides in Medicinal Applications. In S. Dumitriu (Ed.), *Polysaccharides in Medicinal Applications* (1st ed., pp. 73–130). Routledge. <https://doi.org/10.1201/9780203742815>
- Reynel, C., Pennington, R. T., Pennington, T. D., Flores, C., & Daza, A. (2003). Árboles útiles de la Amazonía peruana y sus usos: Un manual con apuntes de identificación, ecología y propagación de las especies. In *Árboles útiles de la Amazonía Peruana*. Herbario de la Facultad de Ciencias Forestales de la Universidad Nacional Agraria-La Molina, Royal Botanic Gardens Kew, Royal Botanic Gardens Edinburgh e ICRAF.
- Richmond, T. (2000). Protein family review Higher plant cellulose synthases Gene organization and evolutionary history. *Genome Biology*, 1(REVIEWS3001), 1–5. <https://doi.org/10.1186/gb-2000-1-4-reviews3001>
- Rinnan, A. (2014). Pre-processing in vibrational spectroscopy – when, why and how. *Anal. Methods*, 6(18), 7124–7129. <https://doi.org/10.1039/C3AY42270D>
- Rinnan, A., Berg, F. Van Den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, 28(10), 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- Rinnan, A., Norgaard, L., Berg, F. van den, Thygesen, J., Bro, R., & Engelsen, S. B. (2009). Data Pre-processing. In D. Sun (Ed.), *Infrared Spectroscopy for Food Quality Analysis and Control* (Vol. 3, pp. 29–50). Elsevier. <https://doi.org/10.1016/B978-0-12-374136-3.00002-X>
- Santana, M. A. E., & Okino, E. Y. A. (2007). Chemical composition of 36 Brazilian Amazon forest wood species. *Holzforschung*, 61(5), 469–477. <https://doi.org/10.1515/HF.2007.084>
- Saxena, I. M., & Brown, R. M. (2005). Cellulose Biosynthesis: Current Views and Evolving Concepts. *Annals of Botany*, 96(1), 9–21. <https://doi.org/10.1093/aob/mci155>
- Scheller, H. V., & Ulvskov, P. (2010). Hemicelluloses. *Annual Review of Plant Biology*, 61, 263–289. <https://doi.org/10.1146/annurev-arplant-042809-112315>

- Sears, R., Cronkleton, P., Arco, M. P., Robiglio, V., Putzel, L., & Cornelius, J. P. (2014). Producción de madera en sistemas agroforestales de pequeños productores. In *Programa de Investigación de CGIAR sobre Bosques, Árboles y Agroforestería (CRP-FTA)*.
- SERFOR. (2017). Programa forestal maderable. *Programa Nacional Sectorial de Ciencia, Tecnología e Innovación Forestal Maderable*, 1–49.
- Shi, J., Xing, D., & Lia, J. (2012). FTIR Studies of the Changes in Wood Chemistry from Wood Forming Tissue under Inclined Treatment. *Energy Procedia*, 16, 758–762. <https://doi.org/10.1016/j.egypro.2012.01.122>
- Shurvell, H. F. (2006). Spectra- Structure Correlations in the Mid- and Far-Infrared. In *Handbook of Vibrational Spectroscopy*. <https://doi.org/10.1002/0470027320.s4101>
- Sibille, A. (2006). Capirona Calycophyllum spruceanum. In *Guía de Procesamiento Industrial* (pp. 1–15). http://plantarperu.com/guia_capirona.pdf
- Skoog, D., Holler, J., & Crouch, S. (2008). *Principios de Análisis Instrumental* (S. Cervantes (ed.); Sexta Edic). Cengage Learning.
- Somerville, C. (2006). Cellulose Synthesis in Higher Plants. *Annual Review of Cell and Developmental Biology*, 22(1), 53–78. <https://doi.org/10.1146/annurev.cellbio.22.022206.160206>
- Stuart, B. H. (2004). Infrared Spectroscopy: Fundamentals and Applications. In K. David J. Ando, Consultant, Dartford (Ed.), *John Wiley & Sons, Ltd* (Vol. 8). John Wiley & Sons, Ltd. <https://doi.org/10.1002/0470011149>
- Subramanian, A., & Rodriguez-Saona, L. (2009). Fourier Transform Infrared (FTIR) Spectroscopy. In *Infrared Spectroscopy for Food Quality Analysis and Control* (pp. 145–178).
- Sun, D. (2009). Fundamentals and Principles of Infrared Spectroscopy. In A. Press (Ed.), *Infrared Spectroscopy for Food Quality Analysis and Control* (1st Editio, pp. 3–25). Elsevier Inc.
- Traoré, M., Kaal, J., & Martínez Cortizas, A. (2018). Differentiation between pine woods according to species and growing location using FTIR-ATR. *Wood Science and Technology*, 52(2), 487–504. <https://doi.org/10.1007/s00226-017-0967-9>
- Trevisan, J., Angelov, P. P., Carmichael, P. L., Scott, A. D., & Martin, F. L. (2012). Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *The Analyst*, 137(14), 3202. <https://doi.org/10.1039/c2an16300d>

- Tuisima-Coral, L. L., Odicio-Guevara, J. E., Weber, J. C., Lluncor-Mendoza, D., & Lojka, B. (2017). Variation in wood physical properties within stems of *Guazuma crinita*, a timber tree species in the Peruvian Amazon. *Madera y Bosques*, *23*(1), 53–61. <https://doi.org/10.21829/myb.2017.2311534>
- Tutus, A., Ezici, A. C., & Ates, S. (2010). Chemical, morphological and anatomical properties and evaluation of cotton stalks (*Gossypium Hirsutum* L.) in pulp industry. *Scientific Research and Essays*, *5*(12), 1553–1560.
- Ushiñahua, D. (2016). Comportamiento fenológico preliminar de Capirona en la provincia de San Martín, región de San Martín. *HOJA DIVULGATIVA N° 002 - 2016, 01*.
- van de Voort, F. R. (1992). Fourier transform infrared spectroscopy applied to food analysis. *Food Research International*, *25*(5), 397–403. [https://doi.org/10.1016/0963-9969\(92\)90115-L](https://doi.org/10.1016/0963-9969(92)90115-L)
- Van Soest, P. J., Robertson, J. B., & Lewis, B. A. (1991). Methods for Dietary Fiber, Neutral Detergent Fiber, and Nonstarch Polysaccharides in Relation to Animal Nutrition. *Journal of Dairy Science*, *74*(10), 3583–3597. [https://doi.org/10.3168/jds.S0022-0302\(91\)78551-2](https://doi.org/10.3168/jds.S0022-0302(91)78551-2)
- Van Soest, P., & Robertson, J. (1979). Systems of Analysis for Evaluating Fibrous Feeds. *Standardization of Analytical Methodology for Feeds*, *4*, 15.
- Vanholme, R., Demedts, B., Morreel, K., Ralph, J., & Boerjan, W. (2010). Lignin Biosynthesis and Structure. *Plant Physiology*, *153*(3), 895–905. <https://doi.org/10.1104/pp.110.155119>
- Via, B. K., Zhou, C., Acquah, G., Jiang, W., & Eckhardt, L. (2014). Near infrared spectroscopy calibration for wood chemistry: Which chemometric technique is best for prediction and interpretation? *Sensors (Switzerland)*, *14*(8), 13532–13547. <https://doi.org/10.3390/s140813532>
- Villa, P. M., Martins, S. V., Delgado Monsanto, L., de Oliveira Neto, S. N., & Mota Cancio, N. (2015). La agroforestería como estrategia para la recuperación y conservación de reservas de carbono en bosques de la Amazonía. *Bosque (Valdivia)*, *36*(3), 347–356. <https://doi.org/10.4067/S0717-92002015000300002>
- Weber, J. C., & Montes, C. S. (2005). Variation and Correlations Among Stem Growth and Wood Traits of *Calycophyllum spruceanum* Benth. from the Peruvian Amazon. *Silvae Genetica*, *54*(1–6), 31–41. <https://doi.org/10.1515/sg-2005-0005>
- Weber, J. C., & Sotelo Montes, C. (2008). Geographic variation in tree growth and wood density of *Guazuma crinita* Mart. in the Peruvian Amazon. *New Forests*, *36*(1), 29–52.

<https://doi.org/10.1007/s11056-007-9080-5>

- Wightman, R., & Turner, S. (2010). Trafficking of the cellulose synthase complex in developing xylem vessels. *Biochemical Society Transactions*, 38(3), 755–760. <https://doi.org/10.1042/bst0380755>
- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)
- Xu, F., Yu, J., Tesso, T., Dowell, F., & Wang, D. (2013). Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review. *Applied Energy*, 104, 801–809. <https://doi.org/10.1016/j.apenergy.2012.12.019>
- Ye, Z.-H., & Zhong, R. (2015). Molecular control of wood formation in trees. *Journal of Experimental Botany*, 66(14), 4119–4131. <https://doi.org/10.1093/jxb/erv081>
- Zhang, F.-D., Xu, C.-H., Li, M.-Y., Chen, X.-D., Zhou, Q., & Huang, A.-M. (2016). Identification of *Dalbergia cochinchinensis* (CITES Appendix II) from other three *Dalbergia* species using FT-IR and 2D correlation IR spectroscopy. *Wood Science and Technology*, 50(4), 693–704. <https://doi.org/10.1007/s00226-016-0815-3>
- Zhao, X., Tong, T., Li, H., Lu, H., Ren, J., Zhang, A., Deng, X., Chen, X., & Wu, A.-M. (2017). Characterization of hemicelluloses from *Neolamarckia cadamba* (Rubiaceae) during xylogenesis. *Carbohydrate Polymers*, 156(September), 333–339. <https://doi.org/10.1016/j.carbpol.2016.09.041>
- Zhou, C., Jiang, W., Cheng, Q., & Via, B. K. (2015). Multivariate Calibration and Model Integrity for Wood Chemistry Using Fourier Transform Infrared Spectroscopy. *Journal of Analytical Methods in Chemistry*, 2015, 1–9. <https://doi.org/10.1155/2015/429846>
- Zhou, G., Taylor, G., & Polle, A. (2011). FTIR-ATR-based prediction and modelling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars. *Plant Methods*, 7(9), 1–10.

VIII. ANEXOS

Anexo 1: Contenido de celulosa, hemicelulosa, holocelulosa y lignina en Capirona

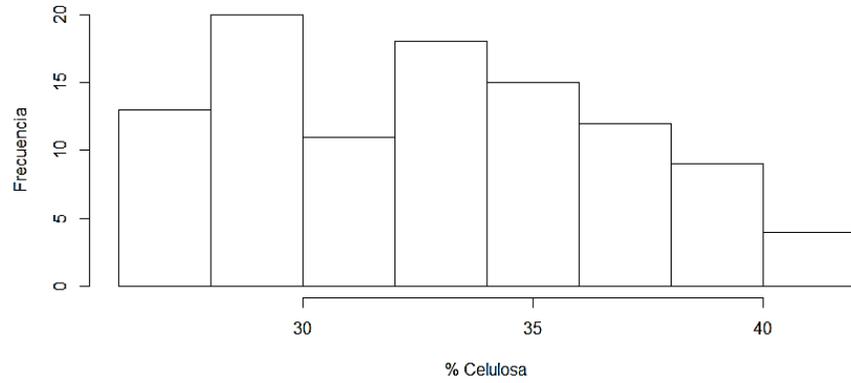
Apical					Basal				
Muestra	Celulosa (%)	Hemicelulosa (%)	Hemicelulosa (%)	Lignina (%)	Muestra	Celulosa (%)	Hemicelulosa (%)	Holocelulosa (%)	Lignina (%)
1A	29.87	10.61	40.48	8.78	1B	32.19	9.52	41.71	9.03
1A	29.40	10.53	39.93	8.68	1B	32.49	9.09	41.58	9.11
2A	29.73	10.34	40.07	7.88	2B	32.62	10.41	43.03	8.14
2A	28.93	9.99	38.92	7.60	2B	32.19	10.31	42.50	8.39
3A	27.93	10.64	38.57	8.45	3B	30.91	10.41	41.32	9.38
3A	29.16	10.76	39.92	8.30	3B	32.22	9.89	42.11	9.63
4A	29.44	10.52	39.96	7.87	4B	37.91	10.28	48.18	9.40
4A	29.61	10.36	39.98	7.48	4B	38.08	9.48	47.56	9.50
5A	34.69	10.86	45.55	7.95	5B	37.23	9.42	46.66	9.93
5A	34.80	10.21	45.01	8.13	5B	37.34	9.58	46.91	9.96
6A	29.09	8.78	37.87	7.90	6B	36.99	8.57	45.56	9.88
6A	28.74	8.08	36.82	7.80	6B	37.99	8.48	46.47	9.76
7A	30.97	10.14	41.11	8.53	7B	35.87	9.97	45.84	9.62
7A	30.10	10.40	40.50	8.27	7B	36.04	10.25	46.29	9.23
8A	32.91	9.15	42.06	8.01	8B	41.69	8.97	50.66	11.05
8A	32.78	10.07	42.85	7.95	8B	41.24	10.25	51.49	10.64
9A	29.30	11.72	41.02	7.73	9B	33.65	10.43	44.08	10.45
9A	28.62	10.88	39.50	7.55	9B	34.55	10.10	44.65	10.05
10A	27.22	8.63	35.84	9.52	10B	31.83	8.77	40.60	9.48
10A	28.32	8.99	37.30	9.88	10B	32.57	7.99	40.57	9.84
11A	29.01	9.08	38.08	9.24	11B	31.38	7.89	39.27	9.72
11A	30.37	9.31	39.67	9.54	11B	31.46	7.53	38.99	9.69
12A	33.01	9.84	42.85	8.92	12B	39.87	10.31	50.19	9.92
12A	31.46	9.14	40.60	9.62	12B	38.64	11.56	50.21	9.70
13A	27.82	10.61	38.43	7.63	13B	33.56	10.40	43.96	8.87
13A	26.63	10.56	37.19	7.45	13B	32.68	9.43	42.11	8.89
14A	32.01	9.70	41.71	8.18	14B	36.16	10.04	46.20	9.14
14A	32.03	9.06	41.09	8.02	14B	35.93	8.85	44.79	9.31
15A	26.31	11.36	37.67	7.81	15B	41.29	7.50	48.79	10.71
15A	26.31	10.35	36.66	7.67	15B	40.21	5.49	45.71	10.50
16A	28.56	8.64	37.20	8.09	16B	35.83	8.71	44.54	9.28
16A	27.94	9.68	37.62	7.92	16B	35.39	8.83	44.22	9.17
17A	28.37	10.29	38.66	8.94	17B	34.32	9.65	43.97	9.45
17A	27.49	9.83	37.32	8.94	17B	33.29	9.95	43.24	9.07
18A	30.12	10.20	40.32	7.86	18B	35.99	9.97	45.97	9.09
18A	29.94	10.18	40.12	7.77	18B	34.87	11.02	45.88	9.11
19A	29.67	8.94	38.60	8.14	19B	39.57	9.08	48.65	9.61
19A	29.15	9.74	38.89	8.53	19B	39.58	8.80	48.38	9.74
20A	33.14	9.90	43.04	7.05	20B	39.12	11.21	50.33	11.28
20A	33.11	11.41	44.52	7.43	20B	38.07	10.98	49.05	11.22
21A	29.88	11.04	40.92	7.98	21B	35.89	10.55	46.44	8.95
21A	29.53	10.95	40.48	7.75	21B	36.36	8.65	45.01	8.84
22A	30.11	10.82	40.93	8.61	22B	38.00	9.99	47.99	10.12
22A	31.03	11.13	42.16	8.24	22B	37.75	10.66	48.41	10.31
23A	35.54	9.32	44.86	7.88	23B	39.70	8.97	48.67	8.81
23A	36.13	10.05	46.19	7.89	23B	39.71	7.90	47.60	9.96
24A	26.45	10.10	36.55	7.87	24B	34.51	8.61	43.11	9.54
24A	27.82	9.53	37.35	7.72	24B	33.10	8.29	41.39	9.25
25A	26.24	11.94	38.18	7.42	25B	35.14	12.66	47.80	10.21
25A	26.92	12.11	39.03	7.31	25B	36.57	12.36	48.93	10.34
25A	26.69	12.13	38.82	7.44	25B	34.95	13.64	48.59	10.08

Anexo 2: Contenido de celulosa, hemicelulosa, holocelulosa y lignina en Bolaina

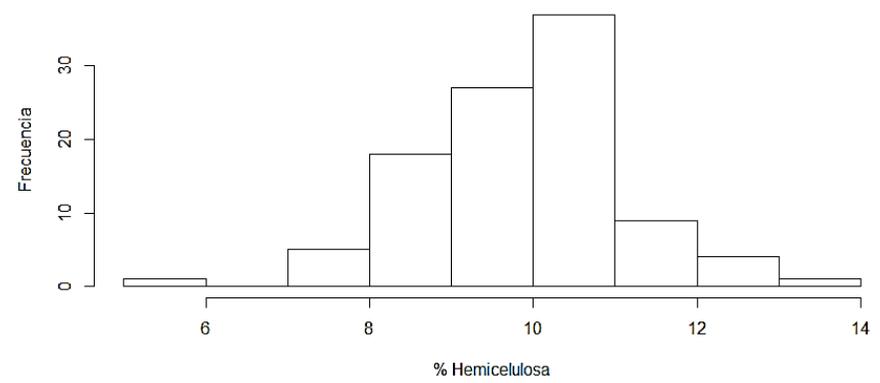
Muestra	Apical				Muestra	Basal			
	Celulosa (%)	Hemicelulosa (%)	Holocelulosa (%)	Lignina (%)		Celulosa (%)	Hemicelulosa (%)	Holocelulosa (%)	Lignina (%)
1A	16.64	24.82	41.46	7.89	1B	44.65	13.93	58.58	7.26
1A	16.48	26.84	43.32	7.82	1B	44.45	13.93	58.38	7.46
2A	20.14	13.84	33.98	11	2B	42.9	14.23	57.13	11.52
2A	20.32	12.47	32.79	10.59	2B	42.96	16.04	59	11.11
3A	18.74	19.02	37.76	10.52	3B	38.83	13.21	52.04	13.91
3A	18.98	18.86	37.84	9.99	3B	39.82	13.55	53.37	14.33
4A	17.14	17.23	34.37	10.26	4B	40.23	14.58	54.81	10.36
4A	17.88	17.1	34.98	10.48	4B	41.14	15.11	56.25	10.31
5A	18.84	19.33	38.17	10.74	5B	40.33	15.66	55.99	10.11
5A	19.27	20.24	39.51	10.82	5B	40.25	13.73	53.98	10.27
6A	19.96	18.78	38.74	11.01	6B	42.61	15.12	57.73	11.75
6A	20.17	18.26	38.43	10.95	6B	43.16	13.93	57.09	12.01
7A	18.91	21.81	40.72	10.27	7B	43.5	13.96	57.46	10.52
7A	18.16	22.8	40.96	10.79	7B	43.01	14.57	57.58	10.87
8A	19.26	18.79	38.05	10.08	8B	40.17	14.32	54.49	10.86
8A	18.97	20.8	39.77	9.72	8B	39.95	14.02	53.97	10.51
9A	19.93	20.01	39.94	10.07	9B	40.19	13.85	54.04	10.13
9A	19.4	21.89	41.29	10.12	9B	40.06	13.09	53.15	10.3
10A	19.13	16.83	35.96	10.38	10B	42.81	14.03	56.84	11.3
10A	18.43	18.04	36.47	10.61	10B	41.56	14.3	55.86	11.17
11A	19.39	16.37	35.76	12.05	11B	43.1	13.42	56.52	11.3
11A	18.67	17.79	36.46	11.68	11B	43.18	15.02	58.2	11.78
12A	19.43	33.44	52.87	13.14	12B	43.79	14.29	58.08	12.14
12A	18.5	35.27	53.77	13.09	12B	43.04	15.09	58.13	12.46
13A	20.17	32.65	52.82	12.56	13B	40.31	12.88	53.19	9.48
13A	19.36	31.98	51.34	12.29	13B	41.69	13.04	54.73	9.46
14A	21.79	12.29	34.08	12.69	14B	39.12	14.42	53.54	10.88
14A	21.75	12.93	34.68	12.92	14B	39.34	14.87	54.21	10.82
15A	19.7	21.78	41.48	12.77	15B	41.56	14.1	55.66	11.68
15A	19.36	22.84	42.2	12.95	15B	40.16	13.73	53.89	11.22
16A	22.62	12.34	34.96	14.32	16B	40.74	15.3	56.04	10.35
16A	22	11.6	33.6	13.82	16B	41.56	14.63	56.19	10.74
17A	23.47	15.8	39.27	14.57	17B	42.85	14.39	57.24	13.32
17A	22.17	15.45	37.62	14.74	17B	42.46	13.98	56.44	13.31
18A	21.05	11.56	32.61	13.48	18B	43.05	14.38	57.43	12.78
18A	21.92	11.74	33.66	14.06	18B	43.81	14.3	58.11	12.59
19A	20.88	11.92	32.8	13.6	19B	42.09	12.42	54.51	10.81
19A	19.62	11.92	31.54	13.6	19B	41.37	13.01	54.38	10.77
20A	22.08	25.1	47.18	14.52	20B	43.93	13.49	57.42	11.09
20A	23.32	22.49	45.81	14.67	20B	43.22	13.95	57.17	11.58
21A	22.37	16.75	39.12	13.59	21B	42.9	15.49	58.39	10.36
21A	22.24	16.44	38.68	13.5	21B	43.9	16.53	60.43	10.56
22A	21.13	28	49.13	12.96	22B	43.78	13.78	57.56	11.58
22A	20.72	27.23	47.95	12.98	22B	42.84	15.34	58.18	11.51
23A	21.94	18.97	40.91	14.63	23B	48.67	15.31	63.98	10.84
23A	21.27	21.58	42.85	14.05	23B	47.64	15.48	63.12	10.4
24A	25.83	16.12	41.95	14.22	24B	44.97	13.91	58.88	12.66
24A	25.05	15.62	40.67	13.87	24B	45.14	13.33	58.47	12.95
25A	23.62	15.04	38.66	15.27	25B	44.63	13.01	57.64	12.47
25A	23.65	14.66	38.31	15.56	25B	44.56	14.25	58.81	12.95

Anexo 3: Histograma de frecuencia de Capirona

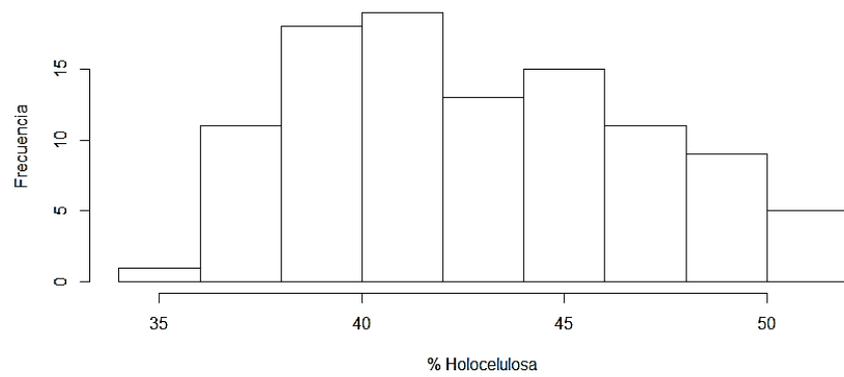
Histograma de Celulosa



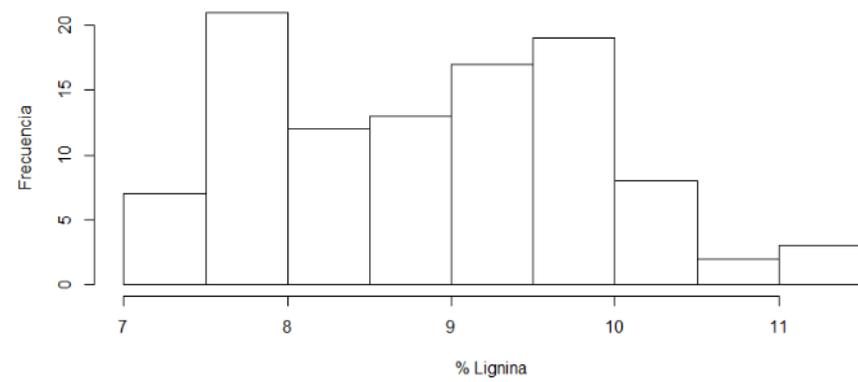
Histograma de Hemicelulosa



Histograma de Holocelulosa



Histograma de Lignina



Anexo 4: Prueba de normalidad de los componentes de Capirona

CELULOSA:

Shapiro-wilk normality test

data: Celulosa

w = 0.96419, p-value = 0.007257

HEMICELULOSA:

Shapiro-wilk normality test

data: Hemicelulosa

w = 0.97847, p-value = 0.09434

HOLOCELULOSA:

Shapiro-wilk normality test

data: Holocelulosa

w = 0.96107, p-value = 0.004274

LIGNINA:

Shapiro-wilk normality test

data: Lignina

w = 0.96507, p-value = 0.008444

Anexo 5: Modelo lineal mixto en función de la sección apical y basal en tallos de Capirona

CELULOSA:

Linear mixed model fit by maximum likelihood ['lmerModLmerTest']

Formula: Celulosa ~ Seccion + (1 | Seccion) + (1 | Arbol)

Data: data_completa

AIC	BIC	logLik	deviance	df.resid
453.9466	467.0715	-221.9733	443.9466	97

Random effects:

Groups	Name	Std.Dev.
Arbol	(Intercept)	2.112e+00
Seccion	(Intercept)	2.357e-05
Residual		1.665e+00

Number of obs: 102, groups: Arbol, 25; Seccion, 2

Fixed Effects:

(Intercept)	SeccionBasal
29.844	6.159

convergence code 0; 0 optimizer warnings; 1 lme4 warnings

Shapiro-wilk normality test

data: resid(Cel_modelo)

w = 0.98851, p-value = 0.5321

HOLOCELULOSA:

Linear mixed model fit by maximum likelihood ['lmerModLmerTest']

Formula: HoloCelulosa ~ Seccion + (1 | Seccion) + (1 | Arbol)

Data: data_completa

AIC	BIC	logLik	deviance	df.resid
460.1161	473.2409	-225.0580	450.1161	97

Random effects:

Groups	Name	Std.Dev.
Arbol	(Intercept)	2.251
Seccion	(Intercept)	0.000
Residual		1.700

Number of obs: 102, groups: Arbol, 25; Seccion, 2

Fixed Effects:

(Intercept)	SeccionBasal
39.97	5.63

convergence code 0; 0 optimizer warnings; 1 lme4 warnings

Shapiro-wilk normality test

data: resid(Holo_modelo)

w = 0.98908, p-value = 0.5766

LIGNINA:

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [glmerMod]

Family: Gamma (log)

Formula: Lignina ~ Seccion + (1 | Arbol) + (1 | Seccion)

Data: data_completa

AIC	BIC	logLik	deviance	df.resid
190.4903	203.6151	-90.2451	180.4903	97

Random effects:

Groups	Name	Std.Dev.
Arbol	(Intercept)	3.562e-02
Seccion	(Intercept)	8.271e-07
Residual		6.423e-02

Number of obs: 102, groups: Arbol, 25; Seccion, 2

Fixed Effects:

(Intercept)	SeccionBasal
2.0955	0.1711

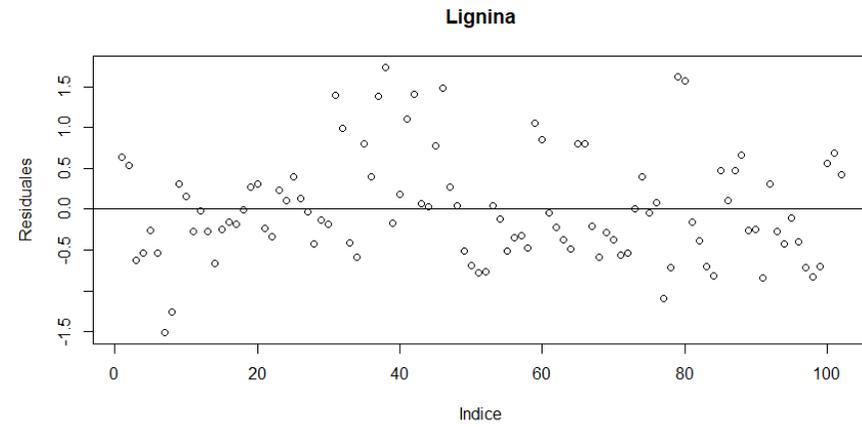
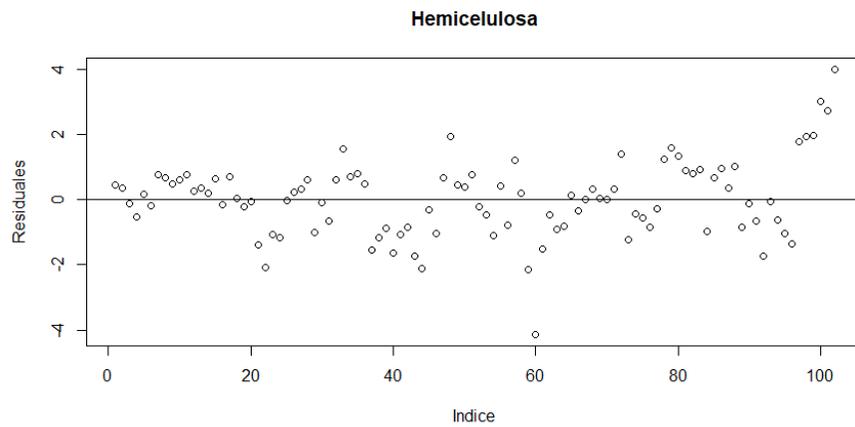
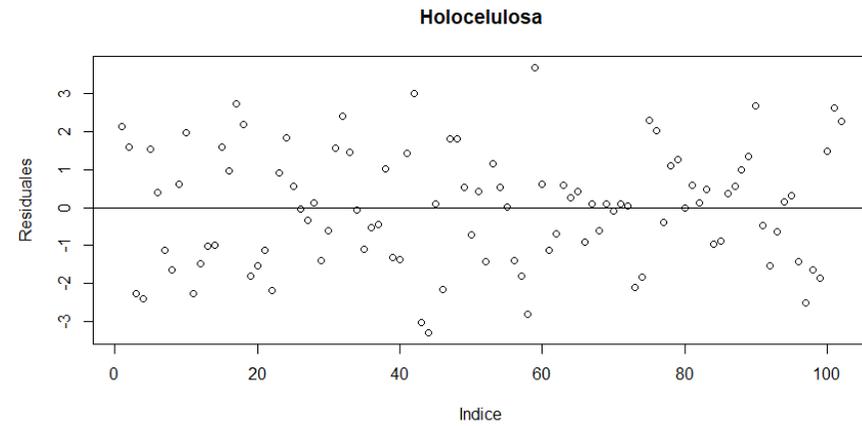
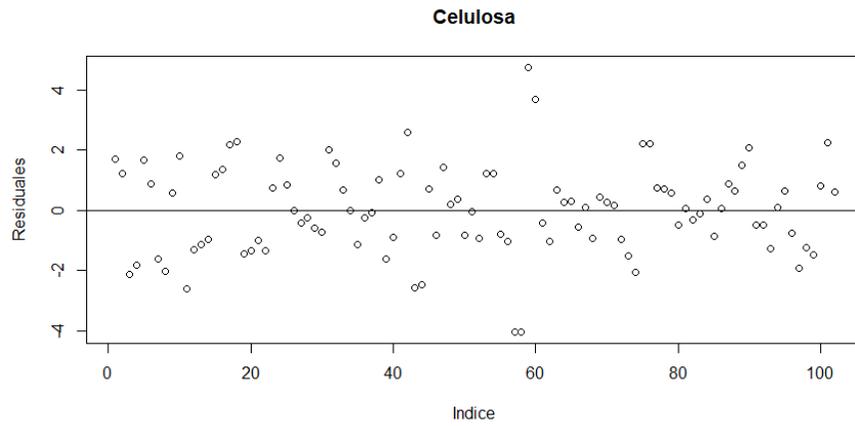
convergence code 0; 0 optimizer warnings; 1 lme4 warnings

Shapiro-wilk normality test

data: resid(Lignina2.Model)

w = 0.97458, p-value = 0.0462

Anexo 6: Homogeneidad de varianza de los modelos generalizados mixtos en compuestos de Capirona



Anexo 7: Prueba de Tukey en los componentes de Capirona.

CELULOSA:

```
Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: Tukey Contrasts
Fit: lmer(formula = Celulosa ~ Seccion + (1 | Seccion) + (1 | Arbol
),
data = data_completa, REML = FALSE, na.action = na.fail)
```

Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z)
Basal - Apical == 0	6.1586	0.3296	18.68	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```
> cld(comp,level=0.05,Letters=letters, decreasing=TRUE)
Apical Basal
 "b"    "a"
```

HEMICELULOSA:

```
Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: Tukey Contrasts
Fit: lm(formula = Hemicelulosa ~ Seccion)
```

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)
Basal - Apical == 0	-0.5284	0.2334	-2.265	0.0257 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```
> cld(comp,level=0.05, decreasing=TRUE)
Apical Basal
 "a"    "b"
```

HOLOCELULOSA:

```
Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: Tukey Contrasts
Fit: lmer(formula =HoloCelulosa ~ Seccion +(1 | Seccion) +(1 |Arbol
),
data = data_completa, REML = FALSE, na.action = na.fail)
```

Linear Hypotheses:

	Estimate	Std. Error	z	value	Pr(> z)
Basal - Apical == 0	5.6302	0.3367	16.72	<2e-16	**

*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```
> cld(comp,level=0.05,Letters=letters, decreasing=TRUE)
```

Apical	Basal
"b"	"a"

LIGNINA:

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

```
Fit: glmr(formula = Lignina ~ Seccion + (1 | Arbol) + (1 | Seccion),  
data = data_completa, family = Gamma(link = log), na.action = na.fail)
```

Linear Hypotheses:

	Estimate	Std. Error	z	value	Pr(> z)
Basal - Apical == 0	0.17107	0.01177	14.53	<2e-16	***

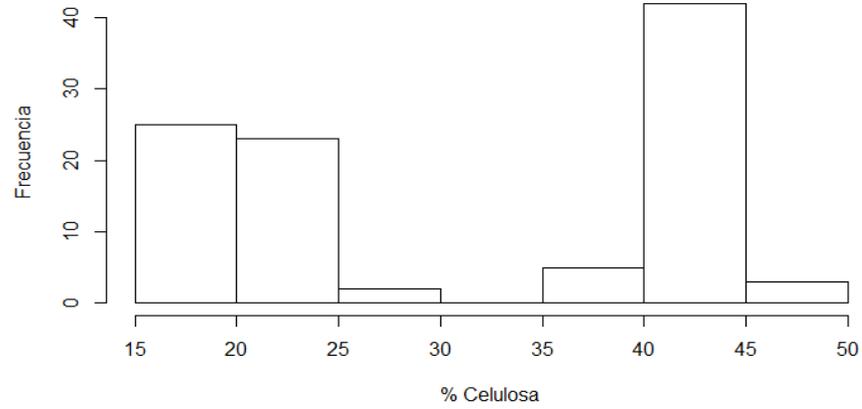
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```
> cld(comp,level=0.05, decreasing=TRUE)
```

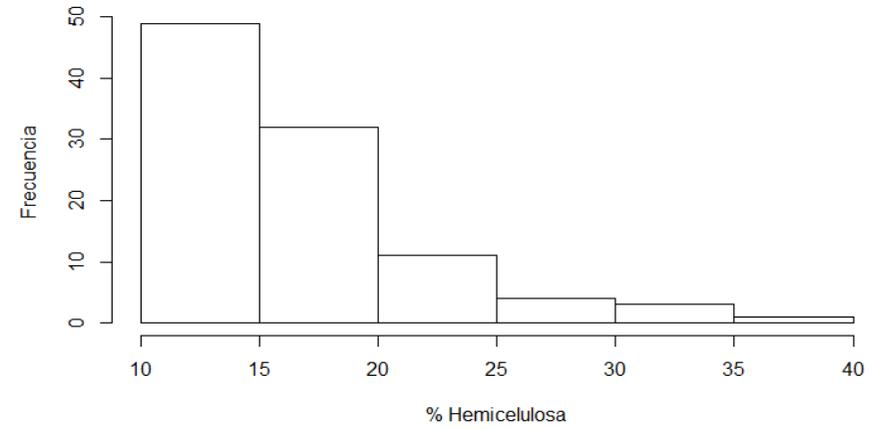
Apical	Basal
"b"	"a"

Anexo 8: Histograma de frecuencias en Bolaina

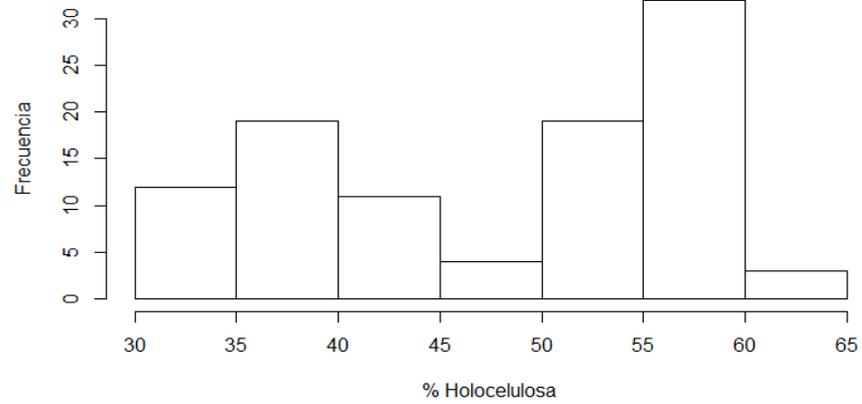
Histograma de Celulosa



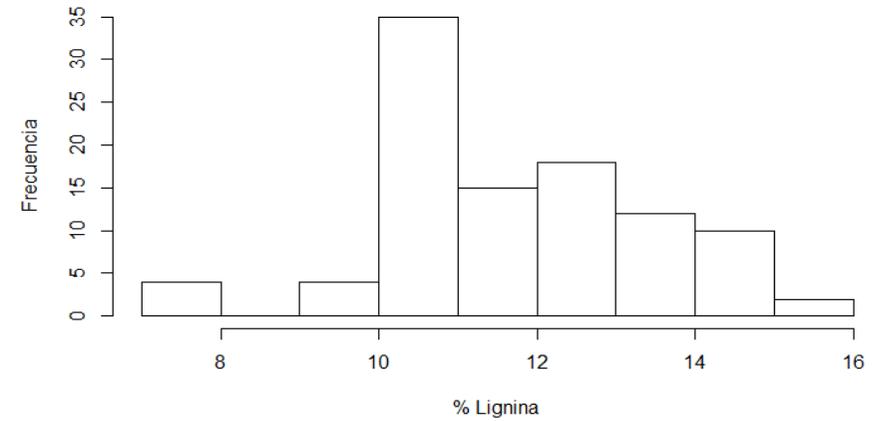
Histograma de Hemicelulosa



Histograma de Holocelulosa



Histograma de Lignina



Anexo 9: Prueba de normalidad de los principales componentes de Bolaina.

CELULOSA:

Shapiro-wilk normality test

data: CELULOSA
w = 0.78846, p-value = 1.112e-10

HEMICELULOSA:

Shapiro-wilk normality test

data: HEMICELULOSA
w = 0.78228, p-value = 7.325e-11

HOLOCELULOSA:

Shapiro-wilk normality test

data: Holocelulosa
w = 0.87763, p-value = 1.443e-07

LIGNINA:

Shapiro-wilk normality test

data: LIGNINA
w = 0.96633, p-value = 0.01167

Anexo 10: Modelo lineal mixto en función de la sección apical y basal en Bolaina

CELULOSA:

Linear mixed model fit by maximum likelihood [Eigen and Moritz]

Formula: CELULOSA ~ SECCIÓN + (1 | ARBOL) + (1 | SECCIÓN)

Data: datos1

AIC	BIC	logLik	deviance	df.resid
402.0869	415.1127	-196.0434	392.0869	95

Random effects:

Groups	Name	Std.Dev.
ARBOL	(Intercept)	1.491
SECCIÓN	(Intercept)	0.000
Residual		1.384

Number of obs: 100, groups: ARBOL, 25; SECCIÓN, 2

Fixed Effects:

(Intercept)	SECCIÓNBasal
20.44	22.00

Shapiro-wilk normality test
data: resid(Celulosa.Model)
w = 0.99115, p-value = 0.7567

HEMICELULOSA:

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [Eigen and Moritz]

Family: Gamma (log)

Formula: HEMICELULOSA ~ SECCIÓN + (1 | ARBOL) + (1 | SECCIÓN)

Data: datos1

AIC	BIC	logLik	deviance	df.resid
515.9336	528.9594	-252.9668	505.9336	95

Random effects:

Groups	Name	Std.Dev.
ARBOL	(Intercept)	0.1130
SECCIÓN	(Intercept)	0.0000
Residual		0.1813

Number of obs: 100, groups: ARBOL, 25; SECCIÓN, 2

Fixed Effects:

(Intercept)	SECCIÓNBasal
2.9284	-0.2681

Shapiro-wilk normality test
data: resid(Hemicelulosa.Model)
w = 0.98013, p-value = 0.1361

HOLOCELULOSA:

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [glmerMod]

Family: Gamma (log)

Formula: HoLocelulosa ~ SECCIÓN + (1 | ARBOL) + (1 | SECCIÓN)

Data: datos1

AIC	BIC	logLik	deviance	df.resid
564.5921	577.6179	-277.2960	554.5921	95

Random effects:

Groups	Name	Std.Dev.
ARBOL	(Intercept)	0.05509
SECCIÓN	(Intercept)	0.00000
Residual		0.07986

Number of obs: 100, groups: ARBOL, 25; SECCIÓN, 2

Fixed Effects:

(Intercept)	SECCIÓNBasal
3.6753	0.3622

Shapiro-wilk normality test

data: resid(HoLocelulosa.Model)

w = 0.96246, p-value = 0.006047

LIGNINA:

Linear mixed model fit by maximum likelihood [lmerModLmerTest']

Formula: LIGNINA ~ SECCIÓN + (1 | ARBOL) + (1 | SECCIÓN)

Data: datos1

AIC	BIC	logLik	deviance	df.resid
347.0805	360.1064	-168.5403	337.0805	95

Random effects:

Groups	Name	Std.Dev.
ARBOL	(Intercept)	1.291e+00
SECCIÓN	(Intercept)	1.981e-09
Residual		1.015e+00

Number of obs: 100, groups: ARBOL, 25; SECCIÓN, 2

Fixed Effects:

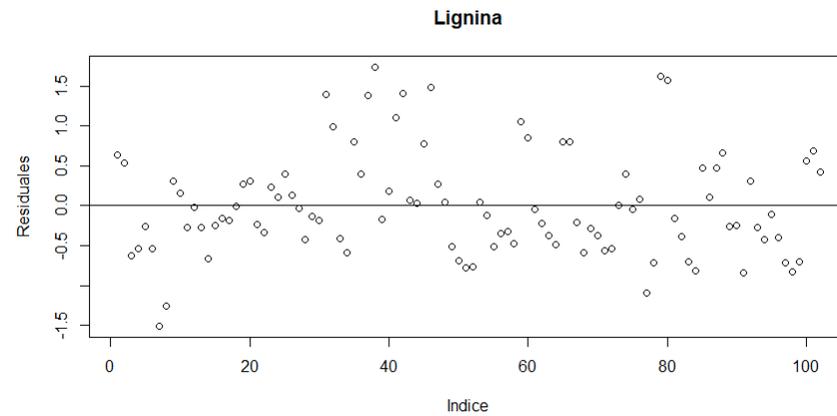
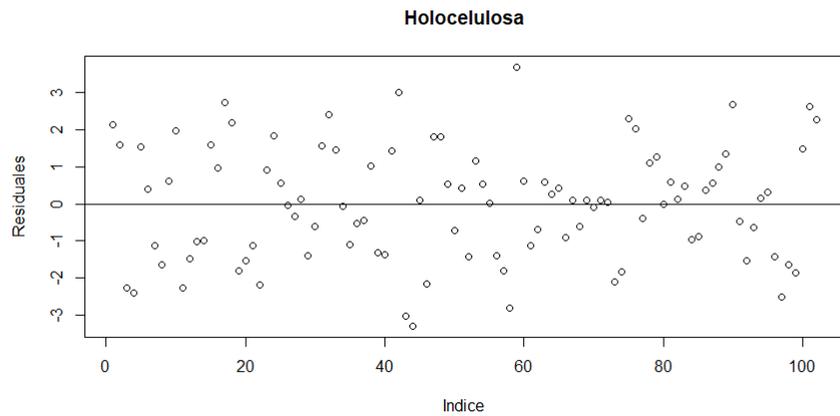
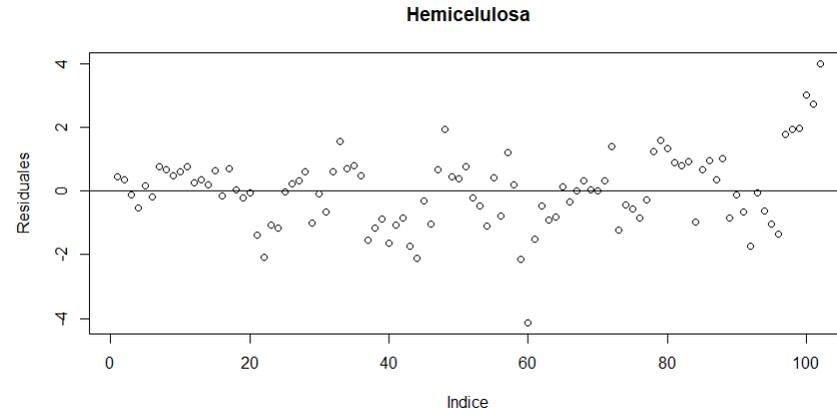
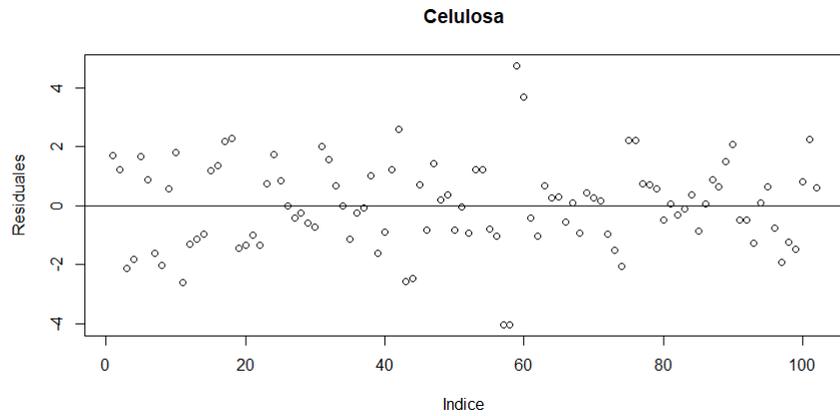
(Intercept)	SECCIÓNBasal
12.245	-1.027

Shapiro-wilk normality test

data: resid(Lignina.Model)

w = 0.98558, p-value = 0.3492

Anexo 11: Homogeneidad de varianza de los modelos mixtos lineales generalizados en Bolaina



Anexo 12: Contraste de medias en la sección basal y apical por teste de Tukey en Bolaina

CELULOSA:

```
Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: Tukey Contrasts
Fit: lmer(formula = CELULOSA ~ SECCIÓN + (1 | ARBOL) + (1 | SECCIÓN),
data = datos1, REML = FALSE, na.action = na.fail)
```

Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z)
Basal - Apical == 0	22.0032	0.2769	79.46	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```
> cld(comp,level=0.05, decreasing=TRUE)
```

Apical	Basal
"b"	"a"

HEMICELULOSA:

```
Fit: glmer(formula = HEMICELULOSA ~ SECCIÓN + (1 | ARBOL) + (1 | SECC
IÓN),
data = datos1, family = (Gamma(link = log)), na.action = na.fail)
```

Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z)
Basal - Apical == 0	-0.26807	0.03318	-8.08	6.66e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```
> cld(comp,level=0.05, decreasing=TRUE)
```

Apical	Basal
"a"	"b"

HOLOCELULOSA:

```
Fit: glmer(formula = Holocelulosa ~ SECCIÓN + (1 | ARBOL) + (1 | SECC
IÓN),
data = datos1, family = Gamma(link = log), na.action = na.fail)
```

Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z)
Basal - Apical == 0	0.36224	0.01422	25.47	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```
> cld(comp,level=0.05, decreasing=TRUE)
```

Apical	Basal
"b"	"a"

LIGNINA:

```
Fit: lmer(formula = LIGNINA ~ SECCIÓN + (1 | ARBOL) + (1 | SECCIÓN),
data = datos1, REML = FALSE, na.action = na.fail)
```

Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z)
Basal - Apical == 0	-1.027	0.203	-5.059	4.21e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```
> cld(comp,level=0.05, decreasing=TRUE)
```

Apical	Basal
"a"	"b"

Anexo 13: Dos regiones y dos pretratamientos de Capirona

Species			Capirona (<i>Callycophyllum spruceanum</i>) ^a						
Compound	Pre-processing	Method of determination os latent variables	3700-850 cm ⁻¹			1800-850 cm ⁻¹			
			Latent Variables	RMSECV	R ²	Latent Variables	RMSECV	R ²	
Cellulose	1 st derivative	Gowen	4	2.43	64.10	4	2.52	61.16	
		Haaland	8	2.00	75.44	9	1.97	76.33	
	2 nd derivative	Gowen	4	2.37	65.85	4	2.50	61.85	
		Haaland	5	2.25	69.21	4	2.50	61.85	
	Smoothing	Gowen	7	2.63	57.74	3	2.83	51.10	
		Haaland	10	2.13	72.28	10	2.23	69.47	
	Vector norm.	Gowen	6	2.35	66.42	3	2.53	60.81	
		Haaland	9	2.01	75.17	9	2.05	74.27	
	Cte offset norm.	Gowen	6	2.77	53.31	4	2.77	53.25	
		Haaland	10	2.26	68.81	10	2.15	71.61	
	Straight line sub.	Gowen	6	2.34	66.41	4	2.69	55.98	
		Haaland	9	2.01	75.17	10	2.15	71.80	
	Min-Max norm.	Gowen	6	2.42	64.34	3	2.57	59.60	
		Haaland	10	1.93	77.20	9	2.10	73.07	
	MSC	Gowen	6	2.32	67.10	3	2.54	60.79	
		Haaland	10	1.77	80.82	9	2.04	74.57	
	Hemicellulose	1 st derivative	Gowen	5	0.80	46.04	5	0.89	32.23
			Haaland	7	0.71	56.71	8	0.65	63.26
2 nd derivative		Gowen	5	0.87	36.09	4	0.93	25.76	
		Haaland	4	0.91	29.95	6	0.87	35.90	
Smoothing		Gowen	8	0.83	41.68	7	0.85	38.71	
		Haaland	10	0.75	51.65	7	0.85	38.71	
Vector norm.		Gowen	8	0.81	44.10	3	1.02	11.39	
		Haaland	9	0.77	49.02	7	0.95	23.06	
Cte offset norm.		Gowen	8	0.83	41.21	3	1.05	6.06	
		Haaland	10	0.75	52.29	8	0.83	40.70	
Straight line sub.		Gowen	6	0.88	33.90	6	0.88	33.93	
		Haaland	9	0.74	52.89	8	0.83	40.84	
Min-Max norm.		Gowen	7	0.88	33.55	4	0.99	15.62	
		Haaland	9	0.80	46.05	7	0.94	24.48	
MSC		Gowen	8	0.79	46.92	3	1.02	11.27	
		Haaland	9	0.77	48.7	7	0.95	22.94	
Holo cellulose		1 st derivative	Gowen	4	2.49	59.53	4	2.56	54.16
			Haaland	9	1.81	78.53	10	1.6	83.3
	2 nd derivative	Gowen	3	2.59	56.41	4	2.57	56.83	
		Haaland	5	2.32	64.90	5	2.47	59.93	
	Smoothing	Gowen	7	2.57	56.82	5	2.62	55.02	
		Haaland	10	2.07	72.05	10	2.17	69.3	
	Vector norm.	Gowen	6	2.38	62.99	3	2.53	58.22	
		Haaland	9	1.95	75.15	9	2.02	73.17	
	Cte offset norm.	Gowen	5	2.76	50.35	4	2.71	52.14	
		Haaland	10	2.18	68.98	9	2.19	68.51	
	Straight line sub.	Gowen	6	2.57	56.97	4	2.63	54.77	
		Haaland	10	2.05	72.69	9	2.22	67.81	
	Min-Max norm.	Gowen	6	2.39	62.77	5	2.41	61.95	
		Haaland	10	1.92	75.78	9	2.09	71.41	
	MSC	Gowen	6	2.38	63.20	7	2.25	66.95	
		Haaland	10	1.62	82.86	10	1.88	76.93	
	Lignin	1 st derivative	Gowen	4	0.56	67.60	4	0.6	62.26
			Haaland	8	0.42	81.34	9	0.4	82.4
2 nd derivative		Gowen	3	0.53	70.42	4	0.53	70.07	
		Haaland	4	0.51	73.10	4	0.53	70.07	
Smoothing		Gowen	4	0.65	55.68	5	0.63	57.23	
		Haaland	10	0.51	73.1	9	0.56	66.38	
Vector norm.		Gowen	4	0.63	58.67	7	0.57	64.96	
		Haaland	9	0.54	69.85	10	0.48	75.86	
Cte offset norm.		Gowen	6	0.63	58.74	5	0.63	57.89	
		Haaland	10	0.51	73.18	10	0.48	75.82	
Straight line sub.		Gowen	3	0.65	55.84	5	0.61	60.2	
		Haaland	10	0.46	77.38	9	0.49	74.76	
Min-Max norm.		Gowen	5	0.62	59.79	6	0.61	60.56	
		Haaland	10	0.51	72.37	9	0.52	41.46	
MSC		Gowen	4	0.63	58.6	7	0.57	65.64	
		Haaland	9	0.53	70.39	9	0.51	72.64	

Anexo 14: Dos regiones y los pretratamientos de Bolaina

Species			Bolaina (<i>Guazuma crinita</i>) ^a					
Compound	Pre-processing	Method of determination of latent variables	3700-850 cm ⁻¹			1800-850 cm ⁻¹		
			Latent Variables	RMSECV	R ²	Latent Variables	RMSECV	R ²
Cellulose	1 st derivative	Gowen	6	3.19	91.84	6	3.23	91.64
		Haaland	9	2.5	94.99	9	2.69	94.2
	2 nd derivative	Gowen	3	3.59	89.64	6	3.16	91.97
		Haaland	6	3.35	90.99	8	2.81	93.64
	Smoothing	Gowen	5	3.57	89.76	5	3.63	89.42
		Haaland	9	3.09	92.34	9	3.14	92.09
	Vector norm.	Gowen	5	1.79	97.41	5	1.84	97.26
		Haaland	10	1.51	98.15	10	1.51	98.17
	Cte offset norm.	Gowen	5	3.54	89.95	6	3.52	90.09
		Haaland	9	3.07	92.42	9	2.97	92.94
	Straight line sub.	Gowen	4	3.59	89.64	6	3.49	90.25
		Haaland	9	2.97	92.91	9	3.05	92.54
	Min-Max norm.	Gowen	4	1.94	96.97	3	1.95	96.93
		Haaland	8	1.69	97.71	8	1.66	97.77
	MSC	Gowen	5	1.82	97.32	7	1.72	97.62
		Haaland	8	1.61	97.9	10	1.5	98.18
Hemicellulose	1 st derivative	Gowen	4	2.49	74.15	6	2.45	74.94
		Haaland	10	1.91	85.47	10	1.75	87.13
	2 nd derivative	Gowen	6	2.24	79.06	6	2.24	79.05
		Haaland	10	1.77	86.89	9	1.89	84.99
	Smoothing	Gowen	5	2.71	69.32	6	2.59	71.93
		Haaland	9	2	83.18	10	2.08	81.79
	Vector norm.	Gowen	6	2.84	66.27	5	3.2	57.12
		Haaland	9	2.43	75.37	10	2.47	74.43
	Cte offset norm.	Gowen	5	2.68	70.04	6	2.47	74.47
		Haaland	9	1.99	83.39	9	2.16	80.49
	Straight line sub.	Gowen	6	2.14	80.74	6	2.65	70.59
		Haaland	8	1.99	83.33	10	2.1	81.5
	Min-Max norm.	Gowen	6	2.87	65.46	6	3.04	61.36
		Haaland	9	2.41	75.68	10	2.37	76.44
	MSC	Gowen	6	2.83	66.48	5	3.24	56.04
		Haaland	9	2.42	75.56	10	2.51	73.67
Holocellulose	1 st derivative	Gowen	7	3.63	85.23	6	3.95	82.47
		Haaland	9	2.89	90.6	9	3.14	88.95
	2 nd derivative	Gowen	3	4.01	81.9	6	3.73	84.35
		Haaland	4	3.92	82.71	8	3.45	86.61
	Smoothing	Gowen	5	4.28	79.4	3	4.35	78.77
		Haaland	9	3.69	84.66	8	3.7	84.58
	Vector norm.	Gowen	6	3.54	85.93	7	3.49	86.31
		Haaland	9	2.17	88.68	9	3	89.86
	Cte offset norm.	Gowen	5	4.25	79.68	5	4.21	80.04
		Haaland	9	3.68	84.8	9	3.75	84.19
	Straight line sub.	Gowen	4	4.23	79.87	5	4.21	80.04
		Haaland	8	3.56	85.79	9	3.75	84.19
	Min-Max norm.	Gowen	7	3.43	86.74	6	3.49	86.31
		Haaland	9	3.14	88.93	10	2.87	90.74
	MSC	Gowen	6	3.53	85.97	7	3.49	86.34
		Haaland	9	3.17	88.7	10	2.94	90.27
Lignin	1 st derivative	Gowen	6	0.95	68.93	6	1.04	63.08
		Haaland	8	0.77	79.47	10	0.73	81.44
	2 nd derivative	Gowen	5	1.08	60.21	7	0.91	71.64
		Haaland	7	0.93	70.42	9	0.8075	77.8
	Smoothing	Gowen	7	1.13	56.31	6	1.17	52.97
		Haaland	9	0.98	67.12	10	0.91	71.54
	Vector norm.	Gowen	6	0.98	66.9	4	1.18	51.91
		Haaland	10	0.77	79.32	10	0.79	78.77
	Cte offset norm.	Gowen	5	1.18	51.95	6	1.16	53.45
		Haaland	9	0.94	69.34	10	0.88	73.27
	Straight line sub.	Gowen	7	1.05	62.36	6	1.15	54.71
		Haaland	9	0.93	69.95	9	0.93	70.16
	Min-Max norm.	Gowen	5	1.12	57.16	6	1.15	54.72
		Haaland	9	0.85	75.2	10	0.89	72.94
	MSC	Gowen	6	1.04	62.92	4	1.19	51.29
		Haaland	10	0.79	78.4	10	0.79	78.59

Anexo 15: Capirona: determinación del número de factores en modelo PLS de celulosa con 2° derivada

Validación cruzada

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM
1	3.55	1	23.01	1893.58	2.7342	0	0	1
2	3.02	0.61	44.54	1364.13	1.97	136878.1	0.05	0.66
3	2.56	0.28	60.09	981.67	1.42	316248.0	0.11	0.40
4	2.37	0.15	65.85	839.93	1.21	715364.8	0.26	0.41
5	2.25	0.06	69.21	757.19	1.09	1219738.0	0.44	0.50
6	2.20	0.03	70.49	725.77	1.05	1713765.9	0.62	0.65
7	2.18	0.01	70.99	713.54	1.03	2015616.6	0.73	0.74
8	2.17	0.00	71.30	705.77	1.02	2305465.9	0.83	0.84
9	2.15	-0.01	71.84	692.56	1	2556449.7	0.92	0.91
10	2.16	0	71.48	701.41	1.01	2767084.1	1	1

Gowen
Haaland

Grupo de calibración

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM
1	3.44	1	27.846	1183.08	144.92	0	0	1
2	2.84	0.81	50.977	803.80	98.46	137152.21	0.05	0.86
3	2.14	0.59	72.104	457.39	56.03	328708.3	0.13	0.71
4	1.57	0.41	85.062	244.94	30.00	777297.5	0.30	0.70
5	1.08	0.25	92.832	117.53	14.40	1346932.8	0.51	0.77
6	0.82	0.17	95.856	67.95	8.32	1862700	0.71	0.88
7	0.60	0.10	97.83	35.57	4.36	2151466.7	0.82	0.92
8	0.45	0.05	98.759	20.35	2.49	2405238.7	0.92	0.97
9	0.37	0.03	99.164	13.71	1.68	2545531	0.97	1.00
10	0.29	0	99.502	8.16	1	2628631	1	1

Gowen

Haaland

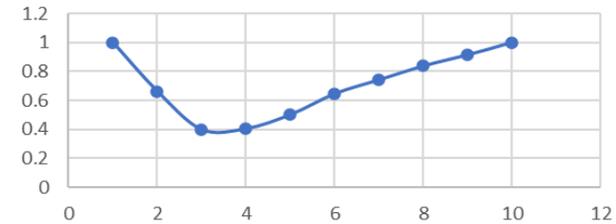
Grupo de validación

Rank	RMSECV	R2	PRESS	Fvalue
1	3.47	26.69	601.05	2.82
2	2.86	50.24	407.92	1.91
3	2.30	67.70	264.82	1.24
4	2.06	74.00	213.13	1.00
5	2.10	73.19	219.79	1.03
6	2.12	72.57	224.86	1.06
7	2.08	73.72	215.47	1.01
8	2.11	72.94	221.86	1.04
9	2.11	72.78	223.19	1.05
10	2.13	72.30	227.13	1.07

Gowen y Halaand

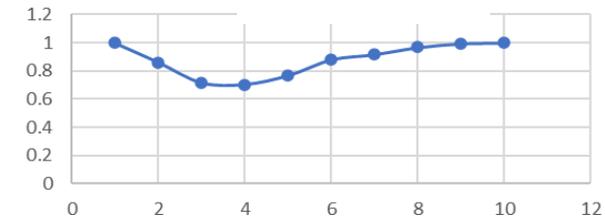
Gowen et al. (2010)

BM + RVM



Gowen et al. (2010)

BM + RVM



Anexo 16: Capirona: determinación del número de factores en modelo PLS de hemicelulosa con 1° derivada

Validación cruzada

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	1.08	1	0.51	174.87	2.56	0	0	1	
2	1.07	0.97	2.79	170.88	2.50	184.0	0.00	0.97	
3	1.01	0.83	12.81	153.26	2.25	1117.4	0.01	0.84	
4	0.96	0.69	22.04	137.03	2.01	18196.3	0.16	0.86	
5	0.80	0.30	46.04	94.85	1.39	28475.6	0.25	0.55	Gowen
6	0.81	0.33	44.17	98.14	1.44	60738.7	0.54	0.87	
7	0.71	0.09	56.79	75.96	1.11	74500.8	0.66	0.75	Haaland
8	0.70	0.06	58.54	72.88	1.07	87234.1	0.77	0.83	
9	0.69	0.04	59.06	71.96	1.05	98496.4	0.87	0.92	
10	0.67	0	61.18	68.24	1	113141.1	1	1	

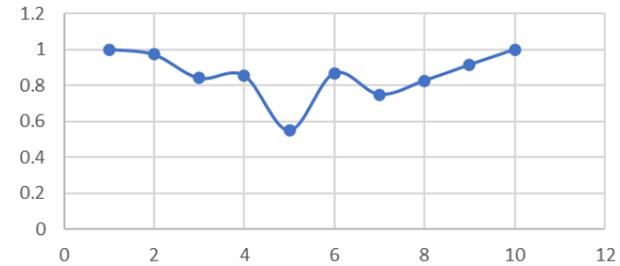
Grupo de calibración

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	1.06	1	3.24	113.39	55.47	0	0	1	
2	1.03	0.96	9.11	106.51	52.104	194.4	0.00	0.97	
3	0.83	0.75	40.91	69.25	33.874	1208.7	0.01	0.76	
4	0.73	0.63	55.14	52.57	25.717	21350.5	0.19	0.82	
5	0.52	0.40	77.33	26.56	12.995	32395.7	0.29	0.69	Gowen
6	0.41	0.28	85.98	16.43	8.0378	62535.4	0.55	0.84	
7	0.29	0.16	92.94	8.28	4.0488	78214.8	0.69	0.85	
8	0.21	0.08	96.15	4.51	2.2043	96531.4	0.85	0.93	
9	0.18	0.04	97.28	3.19	1.5588	108323.6	0.96	1.00	
10	0.14	0	98.26	2.04	1	113082.0	1	1	Haaland

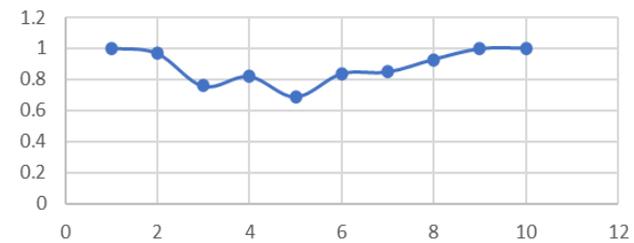
Grupo de validación

Rank	RMSECV	R2	PRESS	Fvalue	
1	1.06	3.30	56.66	2.77	
2	1.04	8.55	53.59	2.62	
3	0.91	28.89	41.67	2.04	
4	0.86	36.41	37.26	1.82	
5	0.68	61.09	22.80	1.11	Gowen y Haaland
6	0.66	62.63	21.90	1.07	
7	0.65	63.40	21.45	1.05	
8	0.66	62.42	22.02	1.08	
9	0.66	62.34	22.07	1.08	
10	0.64	65.08	20.46	1.00	

BM+RVM



BM+RVM



Anexo 17: Capirona: determinación del número de factores en modelo PLS de holocelulosa con 2° derivada

Validación cruzada

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	3.51	1	19.66	1849.49	2.58	0	0	1	
2	3.01	0.62	40.96	1359.11	1.89	108675.8	0.04	0.66	
3	2.59	0.30	56.41	1003.44	1.40	297060.5	0.11	0.41	Gowen
4	2.42	0.17	61.97	875.46	1.22	677090.9	0.25	0.42	
5	2.32	0.10	64.90	808.02	1.13	1175828.0	0.43	0.53	Haaland
6	2.27	0.06	66.56	769.81	1.07	1655877.8	0.60	0.66	
7	2.24	0.04	67.42	750.00	1.04	1989887.2	0.72	0.76	
8	2.21	0.02	68.09	734.60	1.02	2284610.3	0.83	0.85	
9	2.19	0.00	68.76	719.21	1.00	2532428.4	0.92	0.92	
10	2.19	0	68.81	718.08	1	2762720.0	1	1	

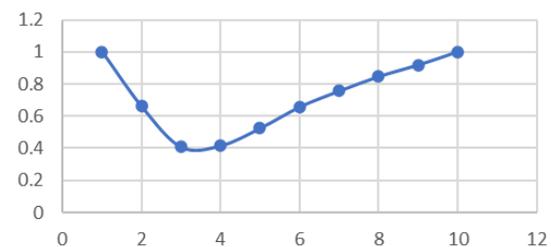
Grupo de calibración

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	3.41	1	24.15	1164.07	152.57	0	0	1	
2	2.83	0.81	47.85	800.35	104.90	109606.2	0.04	0.86	
3	2.18	0.61	68.96	476.30	62.43	309366.3	0.12	0.73	
4	1.62	0.43	82.97	261.42	34.26	730986.2	0.28	0.70	Gowen
5	1.14	0.27	91.60	128.95	16.90	1327487.0	0.50	0.78	
6	0.85	0.18	95.32	71.81	9.41	1855517.8	0.70	0.88	
7	0.63	0.11	97.41	39.70	5.20	2171897.5	0.82	0.93	
8	0.52	0.08	98.23	27.23	3.57	2413083.7	0.91	0.99	
9	0.40	0.04	98.94	16.23	2.13	2520634.1	0.95	0.99	
10	0.28	0	99.50	7.63	1	2642071.8	1	1	Haaland

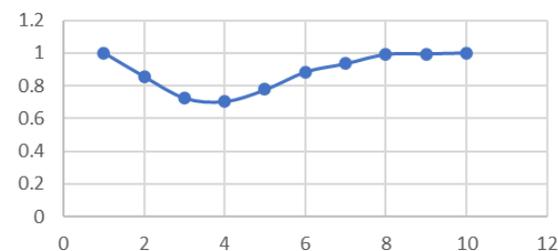
Grupo de validación

Rank	RMSECV	R2	PRESS	Fvalue	
1	3.44	22.88	591.78	2.76	
2	2.85	47.06	406.22	1.90	
3	2.28	66.08	260.29	1.22	
4	2.07	72.10	214.09	1.00	Gowen y Haaland
5	2.13	70.39	227.21	1.06	
6	2.19	68.65	240.54	1.12	
7	2.15	69.81	231.64	1.08	
8	2.17	69.28	235.72	1.10	
9	2.19	68.69	240.24	1.12	
10	2.21	68.04	245.25	1.15	

BM+RVM



BM+RVM



Anexo 18: Capirona: determinación del número de factores en modelo PLS de lignina con 2° derivada

Grupo de Calibración

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	0.79	1	34.94	62.12	210.93	0	0	1	
2	0.59	0.73	63.77	34.60	117.46	30152.1	0.05	0.78	
3	0.44	0.53	79.55	19.53	66.31	78142.3	0.14	0.67	
4	0.31	0.34	90.16	9.39	31.90	159940.2	0.29	0.63	Gowen
5	0.23	0.23	94.66	5.10	17.30	299604.3	0.54	0.78	
6	0.18	0.17	96.58	3.27	11.09	375524.5	0.68	0.85	
7	0.13	0.11	98.16	1.76	5.97	430315.8	0.78	0.89	
8	0.10	0.06	98.94	1.01	3.43	490848.5	0.89	0.96	
9	0.08	0.03	99.41	0.56	1.91	526874.2	0.96	0.99	
10	0.05	0	99.69	0.29	1	550241.0	1	1	Haaland

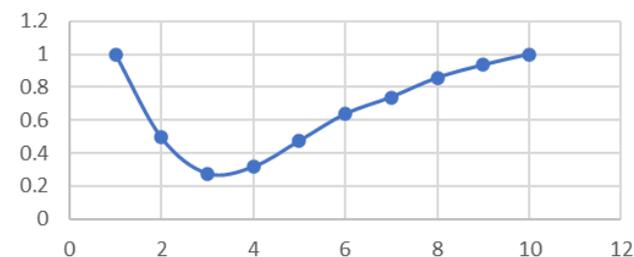
Validación cruzada

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	0.81	1	31.30	98.40	2.83	0	0	1	
2	0.63	0.45	58.75	59.09	1.70	29730.9	0.05	0.49	
3	0.53	0.15	70.42	42.37	1.22	75316.9	0.12	0.28	Gowen
4	0.51	0.08	73.10	38.53	1.11	146067.6	0.24	0.32	Haaland
5	0.49	0.03	74.68	36.26	1.04	271298.0	0.44	0.47	
6	0.49	0.03	74.58	36.42	1.05	370430.6	0.60	0.64	
7	0.49	0.04	74.54	36.47	1.05	432743.0	0.70	0.74	
8	0.49	0.04	74.54	36.46	1.05	505070.4	0.82	0.86	
9	0.49	0.02	75.19	35.53	1.02	565592.7	0.92	0.94	
10	0.48	0	75.75	34.74	1	615880.8	1	1	

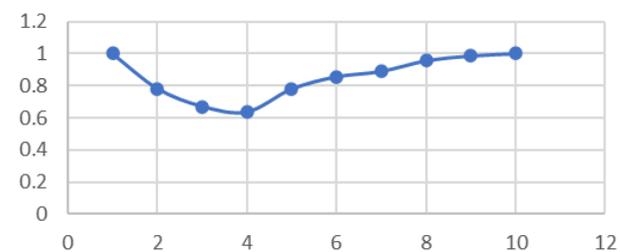
Grupo de Validación

Rank	RMSECV	R2	PRESS	Fvalue	
1	0.80	33.76	31.63	2.71	
2	0.60	62.91	17.71	1.52	
3	0.50	73.61	12.60	1.08	Haaland
4	0.48	75.43	11.73	1.00	Gowen
5	0.48	75.53	11.69	1.00	
6	0.52	72.23	13.26	1.13	
7	0.52	71.70	13.51	1.16	
8	0.51	72.81	12.98	1.11	
9	0.50	74.31	12.27	1.05	
10	0.49	74.89	11.99	1.03	

BM+RVM



BM+RVM



Anexo 19: Bolaina: determinación del número de factores en modelo PLS de celulosa con mínimo máximo de normalización

Validación cruzada

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	3.10	1	92.29	1445.62	3.70	0.00	0	1	
2	2.76	0.77	93.91	1143.15	2.93	1.62	0.01	0.78	
3	2.44	0.55	95.25	891.24	2.28	9.63	0.06	0.61	
4	1.94	0.22	96.98	566.93	1.45	18.16	0.11	0.34	Gowen
5	1.89	0.18	97.15	533.98	1.37	31.31	0.20	0.38	
6	1.81	0.13	97.39	488.84	1.25	36.12	0.23	0.36	
7	1.80	0.13	97.40	487.22	1.25	52.93	0.33	0.46	
8	1.69	0.05	97.71	429.38	1.10	69.33	0.43	0.49	Haaland
9	1.66	0.03	97.80	412.87	1.06	122.91	0.77	0.80	
10	1.61	0	97.92	390.21	1	159.81	1	1	

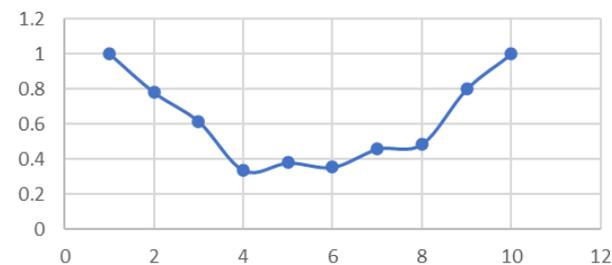
Grupo de calibración

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	3.08	1	92.43	946.23	4.96	0.00	0	1	
2	2.58	0.71	94.66	667.31	3.50	1.62	0.01	0.72	
3	2.23	0.50	96.02	497.77	2.61	10.13	0.06	0.56	
4	1.83	0.26	97.32	334.71	1.76	18.59	0.11	0.37	Gowen
5	1.77	0.23	97.50	313.19	1.64	32.52	0.19	0.42	
6	1.69	0.18	97.72	284.56	1.49	37.58	0.22	0.40	
7	1.64	0.16	97.84	270.17	1.42	55.30	0.32	0.48	
8	1.51	0.07	98.19	226.90	1.19	71.80	0.42	0.49	
9	1.45	0.04	98.33	209.32	1.10	131.78	0.77	0.81	Haaland
10	1.38	0	98.48	190.67	1	171.46	1	1	

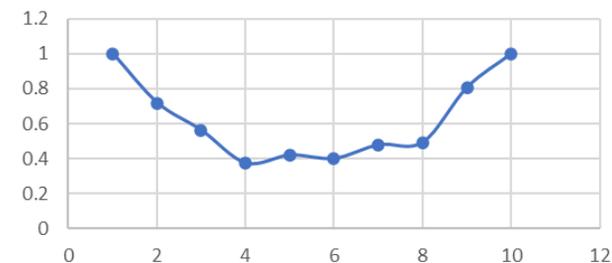
Grupo de validación

Rank	RMSECV	R2	PRESS	Fvalue	
1	3.05	92.55	465.62	4.81	
2	2.57	94.71	330.76	3.41	
3	2.23	96.03	248.52	2.57	
4	1.82	97.35	165.73	1.71	Gowen
5	1.77	97.51	155.84	1.61	
6	1.67	97.78	138.88	1.43	
7	1.62	97.89	132.01	1.36	
8	1.50	98.20	112.85	1.16	
9	1.44	98.34	104.10	1.07	Haaland
10	1.39	98.45	96.87	1	

BM+RVM



BM+RVM



Anexo 20: Bolaina: determinación del número de factores en modelo PLS de hemicelulosa con 1° derivada

Validación cruzada

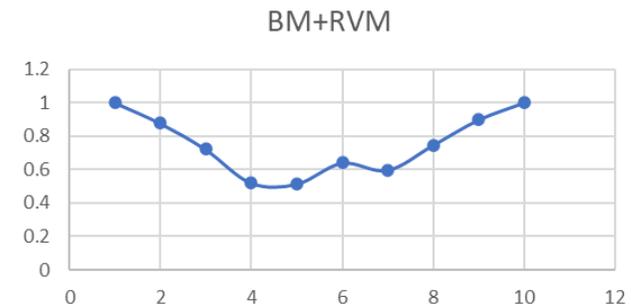
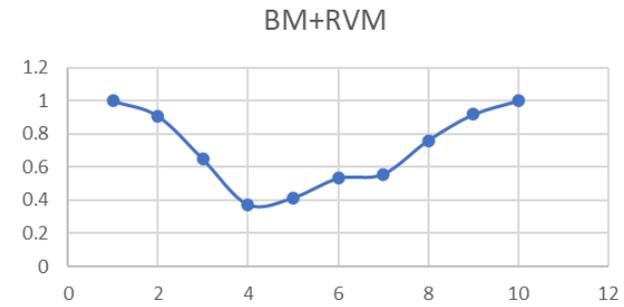
Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	4.18	1	27.13	2622.09	5.63	0	0	1	
2	3.92	0.89	36.04	2301.58	4.94	2490.9	0.02	0.91	
3	3.24	0.61	56.35	1570.49	3.37	5372.5	0.04	0.65	
4	2.49	0.30	74.15	930.07	2.00	9987.7	0.07	0.37	Gowen
5	2.35	0.24	77.02	826.85	1.77	23639.5	0.17	0.41	
6	2.25	0.20	78.96	757.15	1.63	45782.0	0.33	0.53	
7	2.11	0.14	81.53	664.74	1.43	57334.9	0.41	0.56	
8	2.03	0.11	82.82	618.11	1.33	89797.0	0.65	0.76	
9	1.93	0.07	84.54	556.23	1.19	117541.9	0.85	0.92	
10	1.76	0	87.05	465.86	1	138408.3	1	1	Haaland

Grupo de calibración

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	4.13	1	32.06	1629.77	28.18	0	0	1	
2	3.67	0.86	41.73	1397.71	24.16	2310.12	0.02	0.88	
3	3.12	0.69	62.70	894.82	15.47	3870.88	0.03	0.72	
4	2.30	0.45	79.01	503.58	8.71	10868.22	0.08	0.52	
5	1.98	0.35	85.13	356.71	6.17	24278.97	0.17	0.51	Gowen
6	1.84	0.30	86.88	314.75	5.44	48526.07	0.34	0.64	
7	1.48	0.20	91.34	207.69	3.59	57854.91	0.40	0.60	
8	1.24	0.12	94.09	141.73	2.45	89631.23	0.62	0.75	
9	1.08	0.07	95.88	98.79	1.71	118676.93	0.82	0.90	
10	0.84	0	97.59	57.84	1	143853.76	1	1	Haaland

Grupo de validación

Rank	RMSECV	R2	PRESS	Fvalue	
1	4.02	32.64	807.94	5.79	
2	3.72	42.23	692.97	4.96	
3	3.00	62.57	448.90	3.21	
4	2.33	77.35	271.65	1.95	
5	2.12	81.30	224.26	1.61	Gowen
6	2.02	83.04	203.38	1.46	
7	1.82	86.15	166.09	1.19	
8	1.81	86.41	163.06	1.17	
9	1.78	86.73	159.21	1.14	
10	1.67	88.36	139.66	1	Haaland



Anexo 21: Bolaina: determinación del número de factores en modelo PLS de holocelulosa con vector de normalización

Validación cruzada

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	4.29	1	79.36	2762.49	1.96	0	0	1	
2	4.11	0.85	81.05	2536.26	1.80	35.78	0.01	0.86	
3	3.99	0.76	82.12	2393.89	1.70	118.84	0.02	0.78	
4	3.95	0.72	82.52	2339.59	1.66	175.32	0.03	0.75	
5	3.82	0.62	83.65	2189.14	1.56	317.60	0.05	0.67	
6	3.54	0.39	85.93	1882.96	1.34	1107.61	0.17	0.56	Gowen
7	3.39	0.27	87.09	1728.51	1.23	2156.27	0.34	0.61	
8	3.29	0.19	87.84	1627.16	1.16	3383.09	0.53	0.72	
9	3.18	0.09	88.68	1514.72	1.08	4728.41	0.74	0.84	Haaland
10	3.06	0	89.48	1407.65	1	6354.89	1	1	

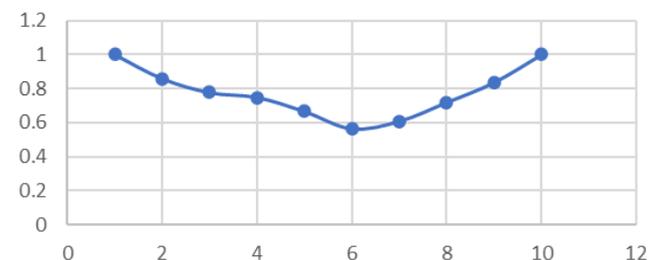
Grupo de calibración

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	4.23	1	79.91	1792.88	3.24	0	0	1	
2	3.90	0.82	82.94	1522.72	2.75	35.92	0.01	0.83	
3	3.77	0.75	84.08	1420.60	2.57	114.22	0.02	0.77	
4	3.67	0.70	84.87	1350.22	2.44	181.28	0.03	0.73	
5	3.35	0.53	87.41	1123.67	2.03	337.95	0.05	0.58	
6	3.09	0.39	89.31	954.07	1.73	1212.17	0.18	0.57	Gowen
7	2.89	0.29	90.64	835.38	1.51	2340.18	0.34	0.63	
8	2.71	0.19	91.74	736.97	1.33	3653.10	0.53	0.73	
9	2.54	0.10	92.76	646.41	1.17	5093.62	0.74	0.85	
10	2.35	0	93.80	552.99	1	6840.88	1	1	Haaland

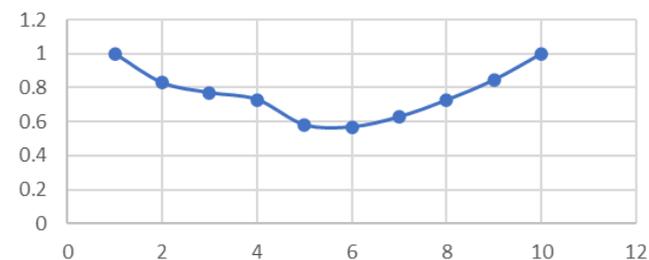
Grupo de validación

Rank	RMSECV	R2	PRESS	Fvalue	
1	4.23	79.99	892.62	2.64	
2	3.91	82.86	764.65	2.26	
3	3.75	84.25	702.73	2.08	
4	3.65	85.03	667.92	1.97	
5	3.38	87.20	570.96	1.69	
6	3.16	88.81	499.41	1.48	Gowen
7	2.96	90.16	438.89	1.30	
8	2.86	90.81	410.06	1.21	
9	2.74	91.60	374.77	1.11	
10	2.60	92.41	338.52	1	Haaland

BM+RVM



BM+RVM



Anexo 22: Bolaina: determinación del número de factores en modelo PLS de lignina con 1° derivada

Validación cruzada

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	1.64	1	8.38	403.79	4.78	0	0	1	
2	1.47	0.81	26.02	326.07	3.86	769.0	0.01	0.83	
3	1.41	0.75	31.92	300.03	3.55	3709.6	0.06	0.80	
4	1.22	0.52	49.70	221.70	2.62	4759.8	0.07	0.60	
5	1.11	0.41	57.95	185.35	2.19	14716.8	0.23	0.63	
6	0.96	0.23	68.94	136.89	1.62	24401.5	0.37	0.60	Gowen
7	0.86	0.13	74.61	111.91	1.32	37454.8	0.58	0.70	
8	0.78	0.03	79.48	90.45	1.07	47058.9	0.72	0.75	Haaland
9	0.79	0.04	79.02	92.48	1.09	56964.4	0.87	0.91	
10	0.75	0	80.83	84.47	1	65114.6	1	1	

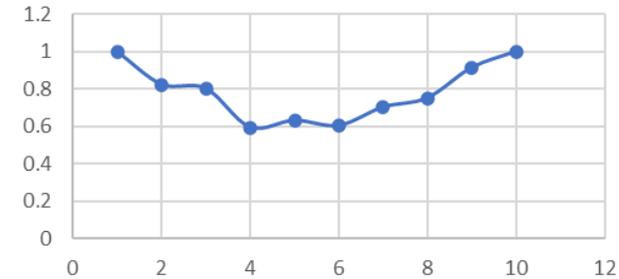
Grupo de calibración

Rank	RMSECV	BM	R2	PRESS	Fvalue	J	RVM	BM+RVM	
1	1.61	1	11.94	258.75	27.83	0	0	1	
2	1.40	0.84	33.58	195.16	20.99	788.3	0.01	0.85	
3	1.34	0.79	39.00	179.24	19.28	3953.7	0.05	0.84	
4	1.06	0.58	61.66	112.65	12.12	5136.0	0.07	0.65	
5	0.89	0.45	73.12	78.99	8.50	15671.9	0.20	0.65	
6	0.72	0.32	82.57	51.22	5.51	27501.9	0.36	0.67	Gowen
7	0.60	0.23	87.62	36.39	3.91	42118.1	0.54	0.77	
8	0.48	0.13	92.19	22.94	2.47	52379.7	0.68	0.81	
9	0.38	0.06	95.04	14.57	1.57	64959.9	0.84	0.90	
10	0.30	0	96.84	9.30	1	77439.8	1	1	Haaland

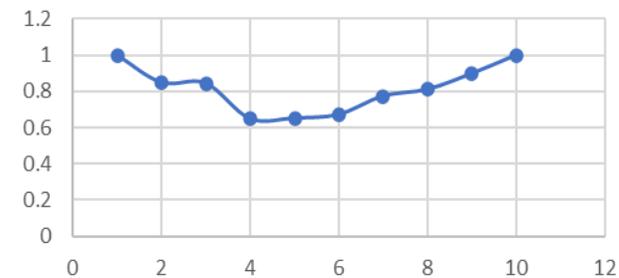
Grupo de validación

Rank	RMSECV	R2	PRESS	Fvalue	
1	1.61	11.95	129.35	6.13	
2	1.42	31.00	101.36	4.80	
3	1.37	36.12	93.84	4.45	
4	1.13	56.79	63.48	3.01	
5	0.98	67.04	48.42	2.29	
6	0.81	77.74	32.70	1.55	Gowen
7	0.76	80.60	28.51	1.35	
8	0.68	84.40	22.92	1.09	
9	0.66	85.04	21.98	1.04	
10	0.65	85.64	21.10	1	Haaland

BM+RVM



BM+RVM



Anexo 23: Registro fotográfico



Casa malla de Bolainas y Capironas



Secado de muestras vegetales



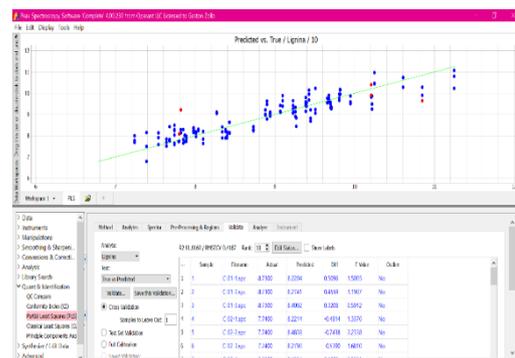
Sección apical (A) y basal (B) de Capirona



Analizador de fibra AKOM



Incubadora Daisy



Procesamiento de datos en el software
Peak Spectroscopy