



Universidad de Concepción  
Dirección de Postgrado  
Facultad de Farmacia -Programa de Doctorado en Ciencias y Tecnología Analítica

**Caracterización química localizada de pulpas lignocelulósicas  
mediante técnicas microscópicas y análisis multivariado  
aplicado a imágenes**

Tesis para optar al grado de Doctor en Ciencias y Tecnología Analítica

JUAN ANDRÉS ARAYA QUINTANA  
CONCEPCIÓN-CHILE  
2017

Profesor Guía: Dra. Rosario Castillo Felices  
Profesor Co-Guía: Dra. Juanita Freer Calderón  
Dpto. de Análisis instrumental, Facultad de Farmacia  
Universidad de Concepción

©2017 JUAN ANDRÉS ARAYA QUINTANA

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.



## Tabla de Contenidos

Índice de Figuras.....	v
Índice de Tablas.....	vii
Resumen.....	viii
<i>Summary</i> .....	x
<b>Capítulo 1:</b> Introducción, Teoría, Hipótesis y Objetivos.....	1
1. Introducción.....	2
1.1 Biomasa lignocelulósica y producción de bioetanol.....	2
1.2 Factores que afectan la recalcitrancia de la fibra lignocelulósica.....	4
1.3 Pretratamiento de biomasa lignocelulósica.....	5
1.4 Métodos analíticos para el estudio de materiales lignocelulósicos.....	7
1.5 Micro-espectroscopía infrarroja y estudio químico in situ.....	8
1.6 Otras técnicas microscópicas: microscopía electrónica de barrido y microscopía de fluorescencia.....	10
2. Teoría.....	12
2.1 Análisis multivariado de imágenes espectrales.....	12
2.2 Pretratamientos de los datos espectrales.....	13
2.3 Métodos Quimiométricos.....	14
3. Hipótesis.....	20
4. Objetivo General.....	20
5. Objetivos Específicos.....	20
Referencias.....	21
<b>Capítulo 2:</b> Estrategia Analítica.....	25
1. Obtención de pulpas lignocelulósicas.....	28
2. Caracterización química de pulpas.....	30
3. Hidrólisis enzimática.....	30
4. Obtención de imágenes microscópicas.....	31
5. Análisis quimiométrico de las imágenes microscópicas obtenidas.....	32

<b>Capítulo 3:</b> Resultados: Enfoque Cualitativo.....	34
<b>Sección 1:</b> Análisis químico superficial a nivel microscópico de fibras pre-tratadas de <i>Pinus radiata</i> mediante SEM, CLSM y microscopía FT-IR.....	37
<b>Sección 2:</b> Efecto de diferentes técnicas de pre-tratamiento de material lignocelulósico en la relocalización de lignina al interior de la pared celular.....	51
Efecto del pre-tratamiento por autohidrólisis.....	52
Efectos del pre-tratamiento por hidrólisis con ácido diluido.....	60
Efectos del pre-tratamiento por Organosolv.....	64
Referencias.....	68
<b>Capítulo 4:</b> Resultados: Enfoque Cuantitativo.....	69
<b>Sección 1:</b> Manuscrito publicado	
<i>Single pixel quantification strategies using middle infrared hyperspectral imaging of lignocellulosic fibers and MCR-ALS analysis</i> .....	72
<b>Sección 2:</b> Manuscrito enviado	
<i>Fourier transformed infrared imaging and quantitative analysis of pre-treated wood fibres: a comparison between partial least squares (PLS) and multivariate curve resolution with alternating least squares (MCR-ALS) methods</i> .....	86
<b>Capítulo 5:</b> Conclusiones finales.....	119

# Índice de Figuras

## Capítulo 1

Figura 1.1: Modelo conceptual simplificado de la estructura de biomasa lignocelulósica...3

Figura 1.2: Representación esquemática del desdoblamiento del hiper-cubo (imagen espectral) de dimensiones X, Y, Z en una matriz bidimensional.....12

## Capítulo 2

Figura 2.1: Esquema de la estrategia analítica para la caracterización química in situ de fibras lignocelulósicas.....27

Figura 2.2: Esquema de análisis de imágenes espectrales mediante métodos quimiométricos.....33

## Capítulo 3

### Sección 1

Figura 3.1: Imágenes representativas obtenidas mediante microscopía confocal y SEM de fibras de la muestra PR2, de alto rendimiento en hidrólisis enzimática.....39

Figura 3.2: Imágenes representativas obtenidas mediante microscopía confocal y SEM de fibras de la muestra PR3, de alto rendimiento en hidrólisis enzimática.....40

Figura 3.3: Espectros FT-IR de patrones de lignina alcalina Kraft y celulosa.....42

Figura 3.4: Microscopía FT-IR de fibra PR2, de alto rendimiento en hidrólisis enzimática..43

Figura 3.5: Microscopía FT-IR de fibra PR3, de bajo rendimiento en hidrólisis enzimática..44

Figura 3.6: Imagen reconstruida por PCA de la fibra PR2.....46

Figura 3.7: Imagen reconstruida por PCA de la fibra PR3.....47

Figura 3.8: Análisis por MCR-ALS de la fibra PR2.....49

Figura 3.9: Análisis por MCR-ALS de la fibra PR3.....50

### Sección 2

Figura 3.10: Cortes transversales de fibras pre-tratadas.....51

Figura 3.11: Imágenes obtenidas por CLSM de fibras provenientes de pulpas pre-tratadas por autohidrólisis.....53

Figura 3.12: Imágenes obtenidas por CLSM espectral, obteniendo el espectro de emisión de fluorescencia, de fibras provenientes de pulpas pre-tratadas por autohidrólisis.....53

Figura 3.13: Espectros de emisión de fluorescencia reconstruidos por PCA (8 componentes) obtenidos de dos regiones de coloración diferente en imágenes obtenidas por CLSM espectral de las muestras AU-1 y AU-5.....	54
Figura 3.14: Imágenes obtenidas por CLSM espectral (emisión de fluorescencia) y reconstruidas mediante desmezclado lineal, de las muestras AU-1 y AU-5 pre-tratadas por autohidrólisis a distintas severidades.....	55
Figura 3.15: Espectros de emisión de fluorescencia de celulosa y lignina.....	56
Figura 3.16: Análisis por MCR-ALS de imágenes obtenidas por CLSM espectral (emisión de fluorescencia) de fibras pre-tratadas por autohidrólisis.....	58
Figura 3.17: Análisis por MCR-ALS (mapa de concentración relativa y espectro puro) de imágenes obtenidas por microscopía FT-IR de cortes transversales de las muestras AU-1 y AU-5.....	59
Figura 3.18: Imágenes obtenidas de la muestra HAD-5 de bajo rendimiento en hidrólisis enzimática, tratada por hidrólisis ácida diluida a baja severidad.....	61
Figura 3.19: Imágenes obtenidas de la muestra HAD-11 de alto rendimiento en hidrólisis enzimática, tratada por hidrólisis ácida diluida a alta severidad.....	61
Figura 3.20: Análisis por MCR-ALS de imágenes obtenidas por CLSM espectral (emisión de fluorescencia) de la muestra HAD-11.....	62
Figura 3.21: Análisis por MCR-ALS (mapa de concentración relativa y espectro puro) de imágenes obtenidas por microscopía FT-IR de cortes transversales de las muestras HAD-5 y HAD-11.....	63
Figura 3.22: Imágenes obtenidas de la muestra PR2 de alto rendimiento en hidrólisis enzimática, tratada por Organosolv.....	65
Figura 3.23: Imágenes obtenidas de la muestra PR3 de alto rendimiento en hidrólisis enzimática, tratada por Organosolv.....	65
Figura 3.24: Distribución y espectros de emisión puros de celulosa y lignina obtenidos mediante MCR-ALS de imágenes obtenidas por CLSM espectral (emisión de fluorescencia) de fibras pre-tratadas por Organosolv.....	66
Figura 3.25: Análisis por MCR-ALS (mapa de concentración relativa y espectro puro) de imágenes obtenidas por microscopía FT-IR de cortes transversales de las muestras PR2 y PR3.....	67

# Índice de Tablas

## Capítulo 1

Tabla 1.1: Resumen de características primarias del sustrato y su hipotético impacto en la accesibilidad de celulasas en base a un ranking de impacto.....	4
--	---

## Capítulo 2

Tabla 2.1: Condiciones de pretratamiento por hidrólisis con ácido diluido para <i>Eucalyptus globulus</i> .....	29
---	----

Tabla 2.2: Condiciones de pretratamiento por autohidrólisis para <i>Eucalyptus globulus</i> .....	29
---	----

Tabla 2.3: Condiciones de pre-tratamiento por Organosolv para <i>Pinus radiata</i> .....	30
--	----

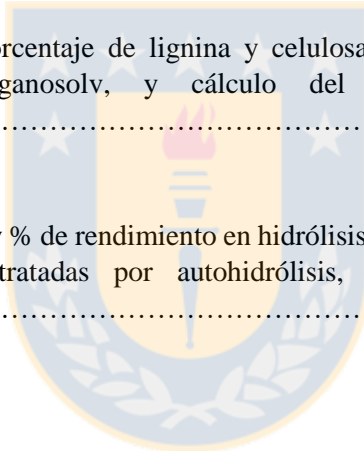
## Capítulo 3

### Sección 1

Tabla 3.1: Cuantificación del porcentaje de lignina y celulosa total, obtenida mediante método Klason, de dos pulpas Organosolv, y cálculo del rendimiento de la hidrólisis enzimática.....	37
---	----

### Sección 3

Tabla 3.2: Composición química y % de rendimiento en hidrólisis enzimática (%R.H.E.) de muestras a alta y baja severidad pre-tratadas por autohidrólisis, hidrólisis con ácido diluido y Organosolv.....	52
--	----



## Resumen

El rendimiento en el proceso de obtención de etanol mediante sacarificación y fermentación de material lignocelulósico, depende en gran medida de la interacción de las enzimas con la fibra de celulosa, donde la presencia de lignina y hemicelulosas suponen una barrera para la accesibilidad enzimática, fenómeno conocido como recalcitrancia. La mayoría de los métodos analíticos existentes estudian la composición química de la madera a través de técnicas destructivas. Para estudiar más de cerca el fenómeno de la recalcitrancia es necesario desarrollar una estrategia que permita el análisis cualitativo y cuantitativo en la superficie misma de la fibra lignocelulósica. La visualización de la distribución espacial de los principales componentes químicos presentes en la fibra es posible mediante obtención de imágenes espectrales a través de micro-espectroscopía FT-IR, donde se obtiene un espectro infrarrojo desde cada píxel de la imagen microscópica. Debido a la gran cantidad de datos obtenidos desde una imagen espectral, es útil la utilización de diferentes técnicas de análisis multivariado, a fin de extraer matemáticamente la información relevante que permita la separación e identificación de los diferentes componentes de la matriz.

En este trabajo se analizaron por FT-IR de imágenes fibras pre-tratadas de material lignocelulósico. Las imágenes espectrales obtenidas fueron sometidas a análisis de componentes principales (PCA), resolución multivariada de curvas con mínimos cuadrados alternantes (MCR-ALS) y regresión por mínimos cuadrados parciales (PLS). La reconstrucción quimiométrica de las imágenes espectrales obtenidas permitió identificar *in situ* los diferentes compuestos químicos presentes en la fibra, así como describir su distribución. Además, se desarrollaron diferentes estrategias para análisis cuantitativo en imágenes espectrales, permitiendo cuantificar glucanos, lignina y hemicelulosa directamente desde la superficie de la fibra.

Se aplicaron otras técnicas microscópicas, como microscopía electrónica de barrido (SEM) y microscopía láser confocal (CLSM) como métodos complementarios para el estudio de la ultraestructura de la fibra lignocelulósica. CLSM permitió la obtención de imágenes espectrales usando el espectro de emisión de compuestos auto-fluorescentes presentes en la madera. Mediante esta técnica fue posible identificar lignina y celulosa en la pared celular de fibras pre-tratadas y describir los patrones de distribución sin necesidad de métodos de marcaje molecular.



Con las estrategias analíticas desarrolladas se observó que diferentes pre-tratamientos producen diferentes patrones de relocalización de lignina y que no es necesariamente el contenido total de lignina lo que afecta la accesibilidad enzimática, sino que la distribución en que ésta se encuentra en la pared celular.

La plataforma analítica desarrollada permite el estudio químico micro-localizado, selectivo y no destructivo para matrices reales complejas y de estructura irregular, lo que proporcionaría una herramienta avanzada para el estudio del fenómeno de recalcitrancia. Este tipo de análisis tendría una aplicación transversal en el estudio de ambientes químicos en matrices microscópicas, por lo que su proyección en otras áreas del conocimiento constituye otra ventaja relativa a su desarrollo.



## Summary

The yield in the process of ethanol obtainment, through saccharification and fermentation of lignocellulosic material, is highly dependent on the interaction on enzymes with the cellulosic fiber, where the presence of lignin and hemicelluloses constitute a barrier for enzymatic accessibility, a phenomena known as recalcitrance. Most of the existing analytical methods study the chemical composition of wood through destructive techniques. To study recalcitrance phenomena it is necessary to develop a strategy that allows qualitative and quantitative analysis in the very surface of the lignocellulosic fiber. Visualization of the spatial distribution of the main chemical components present in the fiber is possible through the obtention of hyperspectral images by FT-IR imaging, where an infrared spectrum is obtained from each pixel of the microscopic image. Due to the high amount of data obtained from a single hyperspectral image, it is necessary to use different multivariate analysis techniques, to mathematically extract relevant information that allows identification of the different components in the matrix

In this work, pre-treated lignocellulosic fibers were analyzed by FT-IR imaging. Hyperspectral images obtained underwent principal component analysis (PCA), multivariate curve resolution-alternating least squares (MCR-ALS) and partial least squares regression (PLS). Chemometric reconstruction of the hyperspectral images obtained allowed to identify *in situ* the different chemical components present in the fiber, along with their distribution. Also, different strategies for quantitative analysis in hyperspectral images were developed, which allowed to quantify glucans, lignin and hemicelluloses directly from the fiber's surface.

Other microscopic techniques were applied as complementary methods, such as scanning electron microscopy (SEM) and confocal laser scanning microscopy (CLSM), for the study of the ultrastructure of the lignocellulosic fiber. CLSM allowed the obtention of hyperspectral images using the emission spectra of auto-fluorescent components present in wood. It was possible to identify lignin and cellulose in the cell wall of pre-treated fibers and to describe distribution patterns without using molecular labeling methods.

With the analytical strategies developed it was observed that different pre-treatments produce different patterns of lignin re-localization, and that enzymatic accessibility is not necessarily affected by the total content of lignin, but the distribution of it in the cell wall.

The developed analytical platform allows a selective, non-destructive, chemically localized study for naturally complex matrices of irregular structure, which would represent an advanced tool to the study of the recalcitrance phenomena. This type of analysis would have a broad application in the study of chemical environments of microscopic matrices; therefore, its application to other areas of knowledge would significate another advantage as consequence of its development.



# Capítulo 1

## Introducción, Teoría, Hipótesis y Objetivos



## 1. Introducción

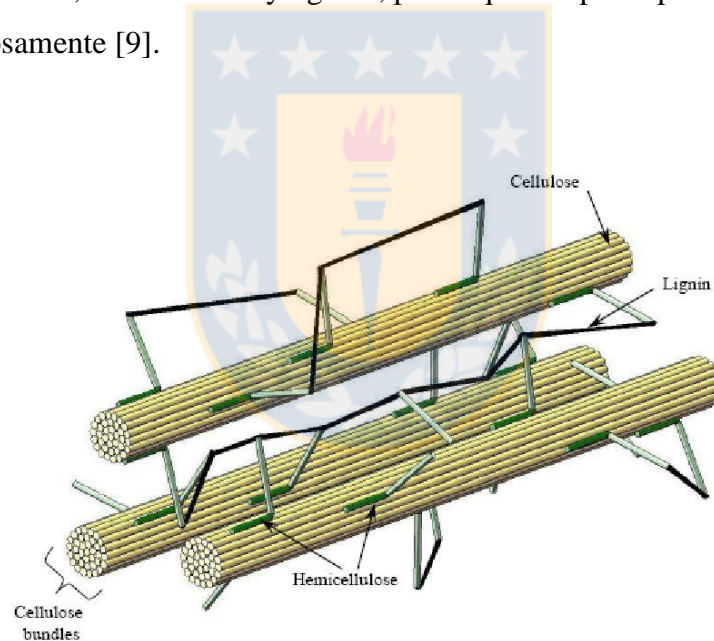
### 1.1 Biomasa lignocelulósica y producción de bioetanol

En nuestro país la demanda de energía se ha incrementado en forma sostenida en años recientes. Para el año 2017 se ha proyectado una demanda de 68.661 GWh de energía y un incremento anual de 5,5% para los próximos 8 años, alcanzándose para el año 2025 una demanda de 105.560 GWh [1] . Esta creciente demanda energética ha sido abastecida principalmente mediante la importación de petróleo crudo y gas natural, por lo que en los últimos años ha crecido el interés por fuentes energéticas locales, renovables y alternativas a los combustibles fósiles [2]. Una corriente importante de este interés se ha dirigido a la producción de biocombustibles a partir de biomasa lignocelulósica debido a la abundancia de este recurso en nuestro país. La biomasa lignocelulósica es la materia orgánica más abundante y renovable de la biosfera, representando el 50% de la biomasa del planeta [3]. En Chile, las exportaciones de productos forestales ocupan el tercer lugar con 13% del producto bruto interno, después de la minería y la industria alimenticia [4]. La región del Bío-Bío particularmente concentra la mayor cantidad de superficie con plantaciones forestales, con casi 1 millón de hectáreas de especies alóctonas para uso industrial, principalmente de *Pinus radiata* y *Eucalyptus globulus* [5].

La conversión de biomasa lignocelulósica, tales como residuos agrícolas y forestales, en etanol y otros productos químicos, ofrece el potencial de ser una fuente de bajo costo, bajo impacto ambiental y proporcionar múltiples beneficios económicos al dar valor agregado a residuos que clásicamente sólo son quemados para generar energía [2]. En relación con su impacto ambiental, la producción de etanol desde biomasa celulósica no genera ningún aporte neto en las emisiones de CO<sub>2</sub> derivadas de la actividad productiva, puesto que la fermentación producida durante la producción de etanol es parte del ciclo del carbono [6]. Si bien existen otras fuentes de biomasa para la obtención de biocombustibles, los residuos forestales y de agricultura tendrían la ventaja sobre las tecnologías basadas, por ejemplo, en producción de etanol desde maíz, ya que no compiten con la demanda alimenticia mundial, condición que limitaría la cantidad de materia prima disponible.

Aunque el potencial de la biomasa forestal como materia prima para la producción de etanol y otros compuestos químicos de interés es alto, hay factores que limitan su competitividad

con fuentes combustibles tradicionales. A diferencia de lo que inicialmente se puede pensar, el contenido total de celulosa de una pulpa de madera no es necesariamente proporcional a la cantidad efectiva de azúcares disponibles luego de una sacarificación enzimática. Las células vegetales de la madera han desarrollado una organización estructural muy compleja y orquestada, lo que le confiere una resistencia mecánica natural a los ataques físicos y biológicos para asegurar su supervivencia [7]. Existen barreras físicas y químicas compuestas por la asociación de los principales componentes de la biomasa lignocelulósica: celulosa, hemicelulosas y lignina (ver Figura 1.1), los que obstaculizan la hidrólisis de celulosa y hemicelulosas en azúcares fermentables [8]. Por lo tanto, se ha desarrollado una serie de pretratamientos para incrementar la accesibilidad a estos azúcares y así conseguir mejorar la digestibilidad de la celulosa. Cada tipo de pretratamiento tiene un efecto específico en los fragmentos de celulosa, hemicelulosa y lignina, por lo que el tipo de pretratamiento debe ser escogido cuidadosamente [9].



**Figura 1.1: Modelo conceptual simplificado de la estructura de biomasa lignocelulósica [10]**

## 1.2 Factores que afectan la recalcitrancia de la fibra lignocelulósica

Existen variados estudios que buscan correlacionar la recalcitrancia de la biomasa lignocelulósica con alguna característica específica de ésta. Algunos autores han atribuido la efectividad de un pretratamiento para mejorar la digestibilidad enzimática de biomasa, al incremento de la porosidad y la superficie de la fibra disponible [11, 12]. Otros estudios establecen que la remoción o redistribución de lignina o hemicelulosas tendrían un papel importante en la recalcitrancia, ya que su presencia obstaculizaría la hidrólisis enzimática [13-15]. Los fenómenos mencionados tienen relación con un mismo factor, que es la accesibilidad de las enzimas a la celulosa. Otros factores apuntan al grado de polimerización y cristalinidad de la celulosa, la que afectaría de manera negativa en la digestibilidad enzimática [16]. Sin embargo, debido a la complejidad de la estructura de la pared celular, los cambios en la cristalinidad absoluta de celulosa son difíciles de determinar de manera precisa [17] (ver Figura 1.1). La estructura, composición y proporción relativa de los biopolímeros presentes en madera puede variar no solo entre tejidos y diferentes tipos celulares, sino que también dentro de una misma célula vegetal. Así, diferentes tipos de fibras tiene una composición diferente de lignina, celulosa y otros constituyentes, y no puede ser considerada igual a otras en términos de cantidad y distribución de sus diferentes componentes [18]. Algunas interacciones que pudieran existir entre otros factores químicos y físicos hace aún más difícil de determinar las variables aisladas que tendrían un mayor impacto en la recalcitrancia [15, 18]. Los factores mencionados han sido resumidos en un trabajo de C.E. Wyman [8] y son resumidos en la Tabla 1.1.

**Tabla 1.1:** Resumen de características primarias del sustrato y su hipotético impacto en la accesibilidad de celulasas en base a un ranking de impacto [10]

Características del Sustrato	Accesibilidad de celulosa a celulasas (Ranking de Impacto <sup>a</sup> )	Efectividad de Celulasas (Rankig de Impacto)
Grupos Acetilo	Efecto pequeño pero notorio (02)	Sí (04)
Cristalinidad de Celulosa	Sí <sup>b</sup>	Sí <sup>b</sup>
Grado de Despolimerización de Celulosa	No concluyente	No concluyente, pero algo de impacto (01)
Contenido de xilanos	Impacto significativo (10)	Impacto mayor (06)
Remoción de Lignina	Aparentemente despreciable (0.5)	Impacto significativo (08)

<sup>a</sup> El ranking fue basado de 0 a 10, donde 10 es el impacto más alto notado en la característica y 0 corresponde a un impacto despreciable

<sup>b</sup> El ranking no fue dado debido a una falta de resolución convincente en literatura

Además de la complejidad intrínseca presente en la matriz de fibra lignocelulósica, la sacarificación enzimática es una reacción heterogénea que requiere de múltiples procesos químicos y fisicoquímicos de interacción entre la enzima y su sustrato, así como de procesos bioquímicos relacionados con la propia actividad enzimática [15], variables que pudiesen tener una incidencia importante en la evaluación de la recalcitrancia a través de rendimientos enzimáticos.

Debido a las variables mencionadas, el estudio de los reales factores que gobiernan la accesibilidad enzimática y el rendimiento en la sacarificación para la obtención de productos de valor agregado es ambiguo y no comprendido del todo [8, 19], por lo que se haría necesario un cambio en la perspectiva en que se está enfrentando la pregunta de investigación, variando hacia metodologías que permitan una aproximación más exhaustiva a los fenómenos microscópicos que están ocurriendo en el proceso de sacarificación de la celulosa.

### 1.3 Pretratamiento de biomasa lignocelulósica

Como ya se mencionó, la sacarificación enzimática de la celulosa está obstaculizada por factores estructurales, químicos y físico-químicos, los que deben necesariamente sortearse. El pretratamiento de la biomasa es por lo tanto un paso esencial para obtener azúcares potencialmente fermentables para la producción de etanol. La premisa principal del pretratamiento es la remoción o ruptura de lignina y la disrupción de la estructura cristalina de la celulosa para mejorar la accesibilidad de las enzimas celulasa y  $\beta$ -glucosidasa [20].

Además de ser considerada una etapa crucial para la conversión de biomasa a etanol y otros subproductos, el pretratamiento representa uno de los mayores costos económicos en el proceso. De hecho es el elemento más caro en el proceso de conversión de lignocelulosa, si no se considera el costo de la materia prima [20].

En primer lugar, el pretratamiento dependerá si la madera corresponde a madera dura o madera blanda, puesto que la química de ambas difiere en algunos aspectos esenciales. En la madera blanda, las hemicelulosas se componen principalmente de glucomananos y galactoglucomananos con bajo grado de acetilación, lo que dificulta la remoción. También hay presencia de xilanos, aunque se encuentran en menor proporción. En el caso de madera



dura, las hemicelulosas están compuestas mayoritariamente por azúcares de 5 carbonos, principalmente heteroxilanos altamente acetilados, los que pueden ser removidos con menor dificultad. También hay presencia de hexosanos aunque en menor cantidad, compuestos principalmente por glucomananos. La lignina de madera blanda se encuentra compuesta principalmente por unidades G (guaiacilo). En madera dura, la lignina se compone de unidades G y S (guaiacilo y siringilo). Las unidades siringilos contienen un grupo metoxilo más que las unidades guaiacilo, por lo que la lignina de madera dura está menos condensada y requiere de menores temperaturas en el proceso de delignificación [21]. Estas características hacen que en madera dura, los procesos de remoción y relocalización de lignina sean más sencillos en comparación a los de madera blanda, y la celulosa se encuentra más accesible luego de los pretratamientos aplicados [21].

Existen diferentes pretratamientos para madera, los que se basan en principios mecánicos, biológicos, físicos, químicos; y físico-químicos, cada uno con un efecto específico diferente en celulosa, hemicelulosas y lignina, por lo que el tipo de pretratamiento debe ser escogido cuidadosamente [9]. Alvira *et al.* publicó en 2010 un completo trabajo sobre el efecto de cada pretratamiento en la biomasa y en la hidrólisis de ésta [9].

Entre los métodos existentes, los de naturaleza química y termoquímica son los más efectivos y en la actualidad son las tecnologías más promisorias para su aplicación industrial [9]. Alvira *et al* [9] concluyen que es necesario poner más atención al estudio de los pretratamientos a un nivel más fundamental, a niveles celulares, de ultraestructura e incluso a nivel molecular, para así contribuir al entendimiento de las diversas reacciones catalíticas que ocurren en la biomasa, así como las consecuencia de los diferentes pretratamientos, lo que llevaría a la obtención de un proceso de conversión de biomasa en etanol más eficiente e integrado. De este modo se podría obtener una comprensión más acabada de los mecanismos fundamentales para un pretratamiento más eficaz, lo ayudaría a acelerar el desarrollo de una tecnología de menor costo para la producción de etanol desde biomasa forestal, al dilucidar los pasos necesarios para vencer la recalcitrancia natural de la fibra vegetal a la disrupción biológica [22].

#### 1.4 Métodos analíticos para el estudio de materiales lignocelulósicos

Los métodos analíticos más usados en la industria para el estudio de materiales lignocelulósicos corresponden a métodos de “química húmeda” los que requieren digestión, separación, y/o purificación de los componentes de la madera para una posterior caracterización mediante métodos analíticos como gravimetría, espectroscopía UV y cromatografía [23-28]. Otros métodos menos destructivos, como la espectroscopía en el infrarrojo, también han sido usados para la caracterización de los macro componentes de la madera [29]. Estas estrategias proporcionan información de la composición química global de la muestra, pero por sí solos no proporcionan información sobre la localización y distribución de sus componentes en la ultraestructura de la muestra, es decir, a nivel de superficie de las fibras lignocelulósicas.

Más recientemente, otras técnicas analíticas han sido utilizadas para investigar las características físicas y químicas de material lignocelulósico que pudiesen afectar su recalcitrancia. En particular, técnicas microscópicas permiten la visualización y descripción de la topografía de las fibras lignocelulósicas a micro-escala. Por ejemplo, la microscopía electrónica de barrido (*scanning electronic microscopy* o SEM, por sus siglas en inglés), la microscopía electrónica de transmisión (*transmission electronic microscopy* ó TEM) la microscopía confocal laser de barrido (*confocal laser scanning microscopy* ó CLSM) y la microscopía de fuerza atómica (*atomic force microscopy* ó AFM) permiten observar características físicas de una muestra con una alta resolución espacial, del orden de los nanómetros, y han sido usadas para observar el efecto de pretratamientos en las características morfológicas superficiales de material lignocelulósico a nivel microscópico [14, 30-34]. Sin embargo, la información química que estas técnicas proporcionan es muy limitada.

Una técnica promisoría para el estudio en simultáneo de las características microscópicas y la composición química del material lignocelulósico, es la micro-espectroscopía infrarroja con transformada de Fourier (llamada también microscopía FT-IR ó FTIR de imágenes). Esta técnica proporciona imágenes microscópicas con información química de alta calidad. Aunque el potencial de esta técnica en el estudio de material lignocelulósico es alta, existen relativamente pocos estudios publicados al respecto, y los trabajos que sí han sido publicados

se basan en análisis espectrales monovariados [35], o bien con técnicas multivariadas preliminares [36, 37], por lo que no explotan el verdadero potencial de esta técnica para la descripción de la “topografía química” a nivel microscópico de biomasa lignocelulósica.

### 1.5 Micro-espectroscopía infrarroja y estudio químico in situ

Las técnicas de micro-espectroscopía infrarroja presenta la oportunidad única de obtener imágenes microscópicas con información química detallada. En este contexto el estudio de imágenes químicas representa una nueva disciplina que combina las capacidades analíticas de la espectroscopía con la habilidad de observar la distribución espacial de los constituyentes de una muestra. Los espectros son recogidos a través de un microscopio, en conjunto con imágenes de alta calidad visual en el punto exacto en que el espectro fue adquirido, logrando visualizar grupos químicos funcionales en sustancias de estructura compleja, a través de sus características vibracionales intrínsecas [38].

Para examinar áreas específicas en una muestra, es necesaria una unidad óptica condensadora que dirija el haz de radiación producido por el interferómetro. Los primeros intentos de implementar microscopía FT-IR siguieron un proceso de dos etapas: primero, las zonas de la muestra se identificaban mediante una inspección visual con microscopio. Luego, se seleccionaba una zona de interés y se incidía luz infrarroja, restringiendo físicamente la radiación que absorbía el área de interés de la muestra. Los espectros eran luego medidos por atenuación de la radiación incidente usando un detector de un solo elemento. Puesto que el análisis debía hacerse punto por punto, el proceso tomaba muchas horas y requería de mucha precisión, por lo que el análisis se limitaba a zonas vecinas limitadas. Por esta razón, se denomina “mapeo” (*mapping*) a este tipo de micro-espectroscopía infrarroja [38].

Más adelante tuvo lugar una segunda aproximación hacia una mejora en la resolución espacial, la que involucraba la segmentación de la radiación en el plano de detección. En contraste al mapeo, en que una pequeña área de la muestra era incidida, se ilumina todo el campo visual. Las contribuciones de las diferentes zonas de la muestra son luego separadas mediante un arreglo, ya sea planar (*focal planar array*, o FPA por sus siglas en inglés) o lineal, de múltiples detectores sensibles a luz infrarroja que barren la muestra, entregando

un espectro IR desde cada punto o pixel de la imagen [39]. Este tipo de técnica es denominada FT-IR de imágenes (*Imaging*), puesto que se obtienen verdaderas imágenes espectrales. Es superior al mapeo en resolución espacial y tiempo de análisis, aunque ambas pueden ser definidas como micro-espectroscopía infrarroja.

Debido a la gran cantidad de información que se obtiene con detectores múltiples, el desarrollo de FPA ha tenido que estar a la par con la capacidad computacional. Por cada interferograma, la señal es integrada en el FPA, transferida al computador y almacenada digitalmente. Con tantos detectores presentes en FPA, el período de tiempo requerido para integrar y almacenar la información excede significativamente el tiempo real de recolección de la señal (milisegundos y microsegundos, respectivamente). Para solucionar este problema, se ha buscado calibrar el interferómetro a bajas velocidades o estabilizar y retardar el espejo, por lo que no se puede sacar todo el partido al uso de esta herramienta [39]. Una nueva aproximación ha sido recientemente implementada, en la cual se emplea un arreglo lineal de detectores para recoger la imagen de un área correspondiente a un espacio rectangular de la muestra. Como el área no es suficientemente representativa de la distribución de las especies químicas de la muestra, el arreglo lineal se mueve secuencialmente a la zona de la muestra que no fue analizada. El proceso es conceptualmente similar al mapeo pero tiene la ventaja de la detección por multicanal. En este caso el número menor de detectores comparado a FPA permite que el interferómetro opere en escaneo continuo, a una frecuencia que favorece la relación señal ruido [39].

Se han incorporado varios tipos de materiales sensibles a IR a los arreglos de detectores; antimoniuro de indio (InSb), telururo de mercurio-cadmio (MCT), arseniuro de silicio ( $\text{Si}_3\text{As}_4$ ), y bario estroncio titanio (BST). La micro-espectroscopía en el rango del infrarrojo medio usando arreglos de detectores MCT ha sido la más popular en términos de número de estudios realizados, debido a su disponibilidad y bajo costo, además de acceder a la zona de huella digital del espectro. Es además el detector más sensible, lo que se hace necesario debido a la escasa cantidad de fotones que llega al detector, producto de las pequeñas aperturas requeridas en IR [37].

En conclusión, la micro-espectroscopía vibracional es una herramienta idónea para el estudio de la química de la madera, y en particular para el estudio de la topo-química de la superficie

de la fibra lignocelulósica, ya que es posible visualizar su composición química su distribución en la anatomía de la pared celular a niveles microscópicos, con poca necesidad de preparación de muestra y con alta información tanto visual como molecular.

### 1.6 Otras técnicas microscópicas: microscopía electrónica de barrido y microscopía de fluorescencia

A fin de dilucidar los complejos mecanismos que subyacen en la interacción enzima-substrato lignocelulósico, así como los factores que determinan la digestibilidad de la celulosa para su aprovechamiento como materia prima para la producción de etanol, es necesario un estudio a nivel microscópico, *in situ*, con la mínima modificación posible de la microestructura de la fibra lignocelulósica [9]. Además de la micro-espectroscopía infrarroja, existen otras técnicas microscópicas, que si bien no entregan información química tan detallada, proporcionan valiosa información morfológica con una resolución espacial muy superior.

La microscopía electrónica de barrido (SEM), es una técnica que utiliza un haz de electrones en vez de luz, los que viajan por la superficie de una muestra previamente cubierta con una fina capa de un conductor (usualmente un metal, como el oro). Permite una visualización con alto nivel de detalle de la superficie de la muestra a estudiar, con una resolución espacial que llega incluso a los nanómetros. Ha sido utilizada en distintos trabajos para el estudio de la recalcitrancia de lignocelulósicos [13, 14, 30-34]. Mediante SEM se ha descrito, por ejemplo, la presencia de gotitas de lignina en la superficie de fibras lignocelulósicas pre-tratadas, las que pueden tener tamaños del orden de los nanómetros hasta varias micras, y que pudiesen tener una influencia en la recalcitrancia, aunque hay versiones encontradas sobre si esta influencia es positiva o negativa [13, 14, 34].

La microscopía confocal laser de barrido (CLSM) combina la microscopía de fluorescencia y la capacidad de seleccionar planos focales, para así obtener imágenes de alta resolución, las que pueden usarse para generar una representación tridimensional del objeto de estudio, mediante la superposición de varias imágenes confocales. Si la fluorescencia puede registrarse a diferentes longitudes de onda, entonces puede obtenerse un espectro, ya sea de

absorción o emisión, del o los fluoróforos en estudio, técnica que se conoce como microscopía confocal espectral. Para evitar confusiones en la nomenclatura, en el presente trabajo se denominará CLSM a la microscopía confocal que muestra la señal de fluorescencia total (1 solo canal), mientras que la microscopía confocal que entrega un espectro de emisión de fluorescencia (varios canales) se denominará CLSM espectral.

La técnica de CLSM ha sido muy utilizada en el estudio de materiales lignocelulósicos [33, 40-42], particularmente por la fuerte autofluorescencia (fluorescencia intrínseca, sin adición de fluoróforos) de la lignina [43]. La lignina tiene la propiedad de auto fluorescer cuando es excitada por luz en el rango UV-Vis [44, 45]. La naturaleza de los cromóforos fluorescentes de la lignina ha sido ampliamente estudiada. Su estructura química compleja aromática y heterogénea presenta muchos elementos estructurales que pueden ser fuentes de fluorescencia: alcoholes coniferílicos, bifenilos, estilbenos, fenilcumarona y estructuras similares a aldehídos coniferílicos son algunos de ellos [45, 46].

Por otro lado, el rol de la celulosa como fuente fluorescente ha sido controversial [43]. La fluorescencia ocurre típicamente en moléculas aromáticas policíclicas altamente conjugadas, sin embargo, ha sido posible observar fluorescencia en muestras de celulosa donde la lignina ha sido removida por pulpaje y blanqueamiento [47]. Aunque se ha postulado que esta fluorescencia proviene de lignina residual u otras impurezas, como proteínas, también es posible observar fluorescencia en muestras de micro-celulosa y en línteres de algodón, donde la presencia de lignina es muy poco probable [48]. Se han propuesto también algunos cromóforos fenólicos y quinoides residuales como los responsables de la fluorescencia en materiales celulósicos [49], sin embargo, la estructura química exacta de los fluoróforos de la celulosa aún no está clara [46].

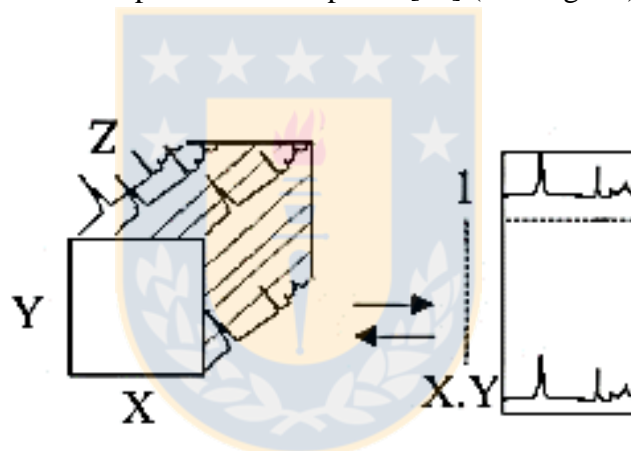
Pese a las demostradas propiedades fluorescentes de la celulosa, la mayoría de los estudios en materiales lignocelulósicos usa tinciones específicas [44] o sondas fluorescentes [50, 51] para la identificación de celulosa en imágenes confocales, por lo que la auto-fluorescencia de celulosa como herramienta analítica, a conocimiento del autor, no ha sido aún explorada. Independiente de la naturaleza de los cromóforos presentes en celulosa, la espectroscopía de fluorescencia permitiría la posibilidad de analizar la estructura molecular de fibras lignocelulósicas (celulosa y lignina) en la misma micro-imagen, sin la necesidad de tinciones

ni sondas, y con una resolución espacial muy superior a la técnica de micro-espectroscopía FT-IR, aunque con una selectividad química mucho menor.

## 2. Teoría

### 2.1 Análisis multivariado de imágenes espectrales

En imágenes híper-espectrales, cada píxel representa un espectro tomado en un rango amplio de longitud de onda, por lo que la cantidad de información que existe es muy elevada (un espectro por cada píxel). La estructura de los datos obtenidos es un arreglo tridimensional conocido como cubo híper-espectral, cuyas dimensiones corresponden a las 2 coordenadas espaciales de la imagen más el espectro en cada punto [52] (Ver Fig.1.2).



**Figura 1.2: Representación esquemática del desdoblamiento del hiper-cubo (imagen espectral) de dimensiones X, Y, Z en una matriz bidimensional (modificado de Budevskva *et al* [53])**

Usualmente, no todos los píxeles o todas las variables contienen información significativa para el análisis. Si todas las variables y los píxeles fuesen analizados individualmente el ejercicio tomaría demasiado tiempo y sería inefectivo. Puesto que un porcentaje de esa información no es relevante para el estudio, los espectros pueden ser analizados por métodos multivariados (conocidos también como análisis quimiométrico) para obtener información descriptiva depurada. Para ello, las imágenes espectrales obtenidas deben ser desdobladas y

tratadas como una matriz bidimensional, píxel-espectro, a fin de aplicar procesamientos matemáticos para extraer información (ver Fig. 1.2) [36].

Después del análisis multivariado los datos son replegados y la imagen es reconstruida en base a la técnica quimiométrica utilizada, obteniéndose así una imagen que contiene sólo información relevante para el análisis, la que puede ser cualitativa y/o cuantitativa. Al permitir el manejo sistemático de grandes cantidades de información y ofrecer una manera de visualizar el análisis de resultados, el uso de la quimiometría ha llegado a ser fundamental en el estudio de imágenes híper-espectrales [54-57].

## 2.2 Pretratamientos de los datos espectrales

Mediante el uso de softwares de análisis matemáticos los datos son desdoblados a la matriz de 2D y luego los espectros en una imagen pueden ser transformados y pre-procesados, como en espectroscopía tradicional, antes del análisis quimiométrico. El pretratamiento aplicado a los datos puede mejorar significativamente los resultados obtenidos al aplicar herramientas de análisis multivariado. Algunos problemas típicos encontrados en imágenes espectrales pueden ser altos niveles de ruido, desplazamiento irregular de línea base (presencia de fondo o presencia de pendiente) y espectros anómalos (“*outliers*”) que influyan los resultados de la imagen analizada [58]. El suavizado espectral puede ser usado para minimizar el ruido. La primera derivada puede compensar el desplazamiento de la señal para remover el fondo. La segunda derivada puede ser usada tanto para eliminar el desplazamiento del fondo como la pendiente de la línea base. Otras transformaciones usadas para remover el fondo e incrementar la información química son la corrección multiplicativa de señal (MSC) y la variable normal estándar (SNV).

Las transformaciones mencionadas, ocurren individualmente para cada muestra, en la dirección del espectro (longitud de onda). El pre-procesamiento, por otro lado, es una operación específica del set de muestras. El centrado en la media, es el pre-procesamiento más usado en datos continuos, como los datos espectrales. En ella se calcula la media para cada variable en un set de datos y luego esta media es sustraída del espectro de cada variable



individual. De este modo las relaciones entre muestras son mejor visualizadas al acentuarse sus diferencias espectrales.

### 2.3 Métodos Quimiométricos

**Análisis de Componentes Principales (PCA):** Los métodos estadísticos multivariados apuntan a descubrir variaciones correlacionadas ocultas en los datos. El análisis de componentes principales, es una de las técnicas quimiométricas más ampliamente utilizada. Tiene como objetivo descubrir la dirección de mayor variación en los datos a través de la descomposición de los datos crudos, en una combinación lineal de factores independientes. En síntesis, consiste en combinaciones lineales de los componentes puros “verdaderos” del espectro [53]. La matriz de componentes principales describe vectores ortogonales independientes que explican la mayor cantidad de varianza de los datos espectrales originales ( $X$ ) y entrega como resultado una matriz de *loading* ( $V$ ) y una de *scores* ( $U$ ) que representa la contribución de cada componente principal encontrado en los datos originales [53]. La matriz de *loadings* puede ser comprendida como los “pesos” o importancia de cada variable original cuando se calcula el componente principal. La matriz de *scores* muestra cuán relacionada está una muestra con un componente principal en particular. Así, la matriz de datos originales puede ser reconstruida a partir de la multiplicación de la matriz de *scores* y la matriz de *loadings*, como se muestra en la siguiente ecuación:

$$X = U V^T + E$$

De este modo describe en unos pocos “componentes principales” (menos que las variables originales de la matriz), la mayoría de la varianza o información contenida en la imagen espectral, depurando el error residual ( $E$ ) el que debería contener un porcentaje de varianza mínimo. No obstante la utilidad de este análisis, el PCA corresponde a una representación matemática abstracta, y no representa necesariamente compuestos químicos puntuales, sino más bien diferencias que pueden atribuirse a factores tanto químicos, físicos o ambientales [53]. Esta última característica es un inconveniente si el objetivo del estudio es una descripción química localizada del objeto del cual se toma la imagen, por lo que PCA serviría sólo como un análisis preliminar, más que un análisis químico *in situ*.

**Resolución Multivariada de Curvas - Mínimos Cuadrados Alternantes (MCR-ALS):** La resolución multivariada de curvas con mínimos cuadrados alternantes pertenece a una familia de técnicas de análisis multivariado iterativo no supervisado desarrollada específicamente para el estudio de mezclas, cuando no existe información acerca de su naturaleza y composición [59]. Fue diseñada en un inicio para resolver picos solapados en métodos cromatográficos, pero su aplicación ha llegado a ser amplia, siendo especialmente idónea en el análisis de imágenes espectrales [53, 57]. En términos sencillo, la meta final de MCR-ALS es encontrar el espectro puro ( $S$ ) de un compuesto y su concentración relativa ( $C$ ) directamente de los datos originales ( $D$ ), como se muestra en la siguiente ecuación:

$$D = C S^T + E$$

$E$  es la matriz de residuos, los que no poseen información química relevante. El proceso de deconvolución se divide básicamente en tres etapas: primero, se debe definir el número real de componentes que proporcionan información química independiente de aquella técnica, y con una sensibilidad suficiente. En algunos casos, la cantidad de componentes presentes en la muestra es conocida de antemano (por ejemplo en muestras artificiales como preparados farmacéuticos) por lo que en el análisis se busca resolver esa cantidad de componentes. Sin embargo, cuando no existe información previa acerca de la cantidad de componentes presentes en una muestra, se puede usar la descomposición de valores singulares (SVD) para determinarlo (uno de los algoritmos más usados para PCA).

En segundo lugar, se necesita una estimación inicial, ya sea del espectro o de la concentración pura del componente, sobre la cual comienza el proceso iterativo de optimización vía ALS. Si se cuenta con información de referencia de los componentes presentes en la mezcla, por ejemplo un espectro puro, esa información es ideal para ser usada como estimación inicial. Ya que MCR-ALS fue diseñado para situaciones donde existe poca o nula información de la muestra, existen algunas aproximaciones matemáticas que proporcionan un estimado inicial en estos casos. Para el caso de imágenes espectrales, la alternativa más usada es el algoritmo “SIMPLISMA” (del inglés SIMPLE-to-use Interactive Self-modeling Mixture Analysis, recientemente ha cambiado su nombre a “PURE” ya que el nombre fue patentado y protegido por derechos de autor, aunque el código es público [60]). Este algoritmo supone que hay variables en los datos en la que contribuye solo uno de los componentes presentes en la

muestra. Por lo que descubre estos componentes puros buscando el mayor ratio entre la desviación estándar de la media para una misma variable, y la variable misma.

El último paso consiste en la optimización por ALS, usando restricciones apropiadas. Puesto que matemáticamente existen infinitas combinaciones lineales entre concentración ( $C$ ) y espectro ( $S^T$ ) que expliquen a la matriz original (problema conocido como ambigüedad rotacional) es necesario la aplicación de restricciones al algoritmo, con marcado criterio químico, para modelar los perfiles entregados hacia un resultado único [61]. Existe una serie de restricciones matemáticas que pueden ser aplicadas dependiendo de la naturaleza de la técnica analítica usada[62]. En datos obtenidos mediante espectroscopía vibracional, la restricción de no negatividad en concentración y en espectro usualmente basta para encontrar una solución única. Finalmente, el proceso iterativo continúa a través de los cálculos de  $C$  y  $S$ , hasta la convergencia del modelo a un error mínimo o a un número de iteraciones preestablecido.

En síntesis, la aplicación de MCR-ALS a imágenes espectrales es una herramienta para responder simultáneamente las siguientes preguntas: cuántos componentes son detectados, cuál es su identidad química (espectro puro) y cómo se distribuyen en el área examinada (concentración relativa) [53].

**Regresión por Mínimos Cuadrados Parciales (PLS):** PLS es el algoritmo más comúnmente usado para en análisis multivariado cuantitativo. Consiste en una regresión lineal, salvo que en vez de usar las variables originales para la regresión, está basado en una reducción de éstas mediante el uso de “componentes principales”. A diferencia de los componentes en PCA, los cuales son ortogonales entre sí, estos componentes son modelados buscando la mayor correlación con la propiedad que se está calibrando, la que en la mayoría de los casos es la concentración. Por esta razón no son llamados “componentes” sino más bien variables latentes.

PLS trabaja construyendo un modelo inverso de calibración  $Y = XB$ , donde  $Y$  es la matriz de concentración,  $X$  es la matriz de variables experimentales y  $B$  es el coeficiente. Para la estimación de los coeficientes de regresión en la matriz  $B$ , ambas matrices  $X$  e  $Y$  son descompuestas independientemente. Una de las características principales de PLS es que el algoritmo modela la información de la matriz experimental ó  $X$  (en el caso de medidas

espectrales, el espectro) al mismo tiempo que modela la concentración (o matriz  $Y$ ), por lo que busca la variable que maximiza el producto de ambos, llamado covarianza. En términos físicos, PLS supone que hay errores en ambas matrices, las que son de igual importancia. La descomposición de la matriz  $X$  e  $Y$  se realiza según las siguientes ecuaciones:

$$\mathbf{X} = \mathbf{T} \mathbf{P}^T + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{T} \mathbf{q}^T + \mathbf{f}$$

Donde  $\mathbf{T}$  corresponden a los *scores*,  $\mathbf{P}$  y  $\mathbf{q}$  son las matrices de pesos de las variables (similar a la matriz *loadings* del PCA, aunque no son ortogonales) y  $\mathbf{E}$  y  $\mathbf{f}$  corresponden a las matrices de residuos de cada bloque. El modelo de regresión es construido buscando minimizar el valor de los residuos (suma de cuadrados del error residual de predicción, o PRESS por sus siglas en inglés), los que son calculados mediante la ecuación:

$$PRESS = \sum_{i=1}^n (y - \hat{y})^2$$

A partir del error residual se calcula un error de calibración (raíz cuadrada media del error de calibración, o RMSEC) mediante la ecuación.

$$RMSEC = \sqrt{\frac{PRESS}{n - k}}$$

Donde  $n$  es el número de muestras y  $k$  el número de variables latentes. A medida que el número de variables latentes consideradas aumenta, el RMSEC va disminuyendo. Se debe seleccionar el menor número de variables latentes posibles para un buen modelo, incluyendo sólo la cantidad de variables con información química significativa para una buena predicción y excluyendo el resto que contiene información no relevante (ruido). Como parámetro para la selección del número de variables latentes a utilizar, se realiza una predicción, ya sea mediante el uso de un set externo de validación, o bien, mediante validación cruzada, el que consiste en eliminar sistemáticamente una o más muestras del set de calibración y prediciendo su valor en un “nuevo” modelo construido sin la presencia de ellas. El error de predicción (raíz cuadrada media del error de predicción RMSEP si se trabaja con un set de validación externa, o raíz cuadrada media del error de validación cruzada

RMSECV si se trabaja con validación cruzada) en ambos casos se calcula mediante la ecuación:

$$RMSEP = \sqrt{\frac{PRESS}{n}}$$

Este parámetro permite además diagnosticar la calidad predictiva del modelo.

La principal diferencia entre usar métodos de calibración en datos espectrales convencionales versus imágenes espectrales es la concentración de referencia que se usa para la calibración. En imágenes, se debería tener valores de referencia para cada pixel de la imagen, lo que es muy poco probable. La situación más común es tener el valor de concentración global de una muestra como referencia. Se han propuesto estrategias que consisten en tomar el espectro promedio o espectro mediana de una imagen como el valor representativo de una imagen y realizar la calibración usando la concentración global de cada muestra [63].

**Desmezclado lineal (*linear unmixing*):** *linear unmixing* es una técnica usada principalmente para deconvolución de espectros de fluorescencia. Se basa en la suposición de que la señal total detectada “S” para cada canal  $\lambda$  puede ser expresada como una combinación lineal de n fluoróforos (Fluo):

$$S(\lambda) = A_1 \times \text{Fluo1}(\lambda) + A_2 \times \text{Fluo2}(\lambda) \dots A_n \times \text{Fluo n}(\lambda)$$

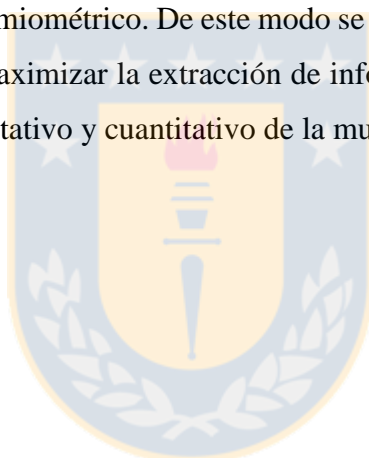
Donde  $A_x$  representa la contribución de un fluoróforo específico. También puede expresarse como:

$$S(\lambda) = A \times R$$

Donde R es un espectro de emisión de referencia de los fluoróforos. Con la señal S detectada y el espectro R de referencia conocido, las contribuciones de los coeficientes “A” de los fluoróforos en la muestra son determinadas calculando el valor de contribución que más se ajuste a las señales en los canales. Esto puede hacerse mediante un ajuste por mínimos cuadrados que minimice la diferencia de cuadrados entre los valores calculados y los medidos. Las ecuaciones lineales son resueltas usualmente mediante el método SVD. Luego del cálculo de los coeficientes A, se puede crear una representación separada de cada fluoróforo.

**Otros Métodos Quimiométricos Aplicados a Imágenes Espectrales:** Otras técnicas de análisis multivariado han sido utilizadas para el estudio de imágenes espectrales, tales como HCA [59], CLS [52], OPLS [55], ICA [64, 65], además de una serie de variaciones y combinaciones de éstas [57], cada uno con ventajas y desventajas, por lo que la técnica a utilizar depende en gran medida del problema particular que se desea resolver. Su aplicación en general permite depurar y maximizar la información que es posible obtener por imágenes espectrales en el estudio químico microscópico de una matriz.

Esta tesis tiene como propósito desarrollar una plataforma analítica que permita el estudio del efecto de diferentes pretratamientos en la composición química de biomasa lignocelulósica así como la localización y distribución de sus componentes a nivel microscópico, mediante el uso de técnicas microscópicas-espectrales (microscopía FT-IR y CLSM espectral) y análisis quimiométrico. De este modo se pretende mejorar la selectividad de las técnicas existentes, al maximizar la extracción de información química relevante que permita un análisis *in situ* cualitativo y cuantitativo de la muestra.



### 3. Hipótesis

La identidad y distribución de los componentes químicos específicos de madera pre-tratada son determinados a través de técnicas microscópicas y métodos quimiométricos aplicados a imágenes espectrales. Además, variaciones en los patrones de distribución de estos componentes afectan el rendimiento en los procesos de conversión enzimática.

### 4. Objetivo general

Desarrollar estrategias de análisis multivariado de imágenes espectrales que permitan identificar y localizar a nivel microscópico los componentes químicos principales presentes en madera pre-tratada, y evaluar el efecto de éstos en los procesos de conversión enzimática para la producción de biocombustibles

### 5. Objetivos específicos

- 1- Desarrollar una estrategia para el análisis quimiométrico de imágenes espectrales de fibras lignocelulósicas obtenidos por FT-IR de imágenes.
- 2- Evaluar la técnica de FT-IR de imágenes para análisis cuantitativo *in situ* en material lignocelulósico pre-tratado.
- 3- Desarrollar una estrategia para el análisis de lignina en fibras lignocelulósicas, mediante SEM, CLSM y CLSM espectral.
- 4- Analizar la distribución y composición química específica a nivel microscópico de material lignocelulósico pre-tratado, mediante las estrategias analíticas desarrolladas y comparar con el rendimiento en los procesos de bio-conversión enzimática.

## Referencias

1. Pedro Maldonado, R.M., Alfredo Muñoz, Francisco Domenech, Iris Silva, Jorge Pontt, Gerd Reinke, Patricio Nuñez; Roberto Leiva, Germán Ubilla, Jaime Espinoza, Rafael Bolocco, Carlos Pontt, Mario Orellana, Nicolás Faúndez, Daniel Díaz, Pedro Velásquez. *Aporte potencial de energías renovables no convencionales y eficiencia energética a la mariz eléctrica, 2008-2025*. 2008 23 June 2017]; Available from: <http://www.neim.utfsm.cl/arch/ERNC.pdf>.
2. Reyes, R., *Leña: una oportunidad para la conservación de los bosques templados del sur de Chile*. Bosque nativo, 2005. **37**: p. 16-23.
3. Ward, O.P. and A. Singh, *Bioethanol technology: developments and perspectives*. Advances in applied microbiology, 2002. **51**: p. 53-80.
4. Donoso, P.J. and L.A. Otero, *Hacia una definición de país forestal: ¿ Dónde se sitúa Chile?* Bosque (Valdivia), 2005. **26**(3): p. 5-18.
5. INFOR. *Mapa Forestal Inventario Continuo*. 2017 23 Junio 2017]; Available from: [http://mapaforestal.infor.cl/index.php?option=com\\_content&view=article&id=54&Itemid=72](http://mapaforestal.infor.cl/index.php?option=com_content&view=article&id=54&Itemid=72).
6. Wyman, C.E., *Ethanol from lignocellulosic biomass: Technology, economics, and opportunities*. Bioresource Technology, 1994. **50**(1): p. 3-15.
7. Xu, F., et al., *Anatomy, ultrastructure and lignin distribution in cell wall of Caragana Korshinskii*. Industrial Crops and Products, 2006. **24**(2): p. 186-193.
8. Wyman, C., *Key features of pretreated lignocelluloses biomass solids and their impact on hydrolysis*. Bioalcoholproduction, 2010: p. 73.
9. Alvira, P., et al., *Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review*. Bioresource Technology, 2010. **101**(13): p. 4851-4861.
10. Wyman, C., *Key features of pretreated lignocelluloses biomass solids and their impact on hydrolysis*. Bioalcohol production: biochemical conversion of lignocellulosic biomass, 2010: p. 73.
11. Ishizawa, C.I., et al., *Porosity and its effect on the digestibility of dilute sulfuric acid pretreated corn stover*. Journal of Agricultural and Food Chemistry, 2007. **55**(7): p. 2575-2581.
12. Chandra, R.P., A.R. Esteghlalian, and J.N. Saddler, *Assessing substrate accessibility to enzymatic hydrolysis by cellulases*. Characterization of Lignocellulosic Materials, 2009: p. 60-80.
13. Koo, B.-W., et al., *Structural changes in lignin during organosolv pretreatment of Liriodendron tulipifera and the effect on enzymatic hydrolysis*. Biomass and Bioenergy, 2012. **42**: p. 24-32.
14. Donohoe, B.S., et al., *Visualizing lignin coalescence and migration through maize cell walls following thermochemical pretreatment*. Biotechnology and Bioengineering, 2008. **101**(5): p. 913-925.
15. Kumar, R. and C.E. Wyman, *Effects of cellulase and xylanase enzymes on the deconstruction of solids from pretreatment of poplar by leading technologies*. Biotechnology Progress, 2009. **25**(2): p. 302-314.



16. Yoshida, M., et al., *Effects of cellulose crystallinity, hemicellulose, and lignin on the enzymatic hydrolysis of Miscanthus sinensis to monosaccharides*. *Bioscience, biotechnology, and biochemistry*, 2008. **72**(3): p. 805-810.
17. Chang, V.S. and M.T. Holtzapfle. *Fundamental factors affecting biomass enzymatic reactivity*. in *Twenty-first symposium on biotechnology for fuels and chemicals*. 2000. Springer.
18. Zhao, X., L. Zhang, and D. Liu, *Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose*. *Biofuels, Bioproducts and Biorefining*, 2012. **6**(4): p. 465-482.
19. Meng, X. and A.J. Ragauskas, *Recent advances in understanding the role of cellulose accessibility in enzymatic hydrolysis of lignocellulosic substrates*. *Current opinion in biotechnology*, 2014. **27**: p. 150-158.
20. Mosier, N., et al., *Features of promising technologies for pretreatment of lignocellulosic biomass*. *Bioresource technology*, 2005. **96**(6): p. 673-686.
21. Brown, R.C. and T.R. Brown, *Biorenewable resources: engineering new products from agriculture*. 2013: John Wiley & Sons.
22. Wyman, C.E., *What is (and is not) vital to advancing cellulosic ethanol*. *Trends in Biotechnology*, 2007. **25**(4): p. 153-157.
23. Sithole, B.B., *Pulp and paper*. *Analytical chemistry*, 1995. **67**(12): p. 87R-95R.
24. Dietrich, F. and W. Gerd, *Wood: chemistry, ultrastructure, reactions*. *J. Polym. Sci. Polym. Lett. Ed*, 1989. **23**: p. 601.
25. Lin, S., *Ultraviolet spectrophotometry*, in *Methods in lignin chemistry*. 1992, Springer. p. 217-232.
26. Nimz, H., *Eighth international symposium on wood and pulping chemistry (ISWPC), June 6-9, 1995, Helsinki, Finland*. 1996, WALTER DE GRUYTER & CO GENTHINER STRASSE 13, D-10785 BERLIN, GERMANY.
27. Moore, W.E. and D.B. Johnson, *Procedures for the chemical analysis of wood and wood products (as used at the US Forest Products Laboratory)*. 1967.
28. Morohoshi, N., *Chemical characterization of wood and its components*. 1991: Marcel Dekker, Inc., New York.
29. Castillo, R.d.P., et al., *Infrared spectroscopy as alternative to wet chemical analysis to characterize Eucalyptus globulus pulps and predict their ethanol yield for a simultaneous saccharification and fermentation process*. *Applied biochemistry and biotechnology*, 2012. **168**(7): p. 2028-2042.
30. Selig, M.J., et al., *Deposition of lignin droplets produced during dilute acid pretreatment of maize stems retards enzymatic hydrolysis of cellulose*. *Biotechnology progress*, 2007. **23**(6): p. 1333-1339.
31. Rezende, C.A., et al., *Chemical and morphological characterization of sugarcane bagasse submitted to a delignification process for enhanced enzymatic digestibility*. *Biotechnology for Biofuels*, 2011. **4**(1): p. 1.
32. Fromm, J., et al., *Lignin distribution in wood cell walls determined by TEM and backscattered SEM techniques*. *Journal of structural biology*, 2003. **143**(1): p. 77-84.
33. Sant'Anna, C. and W. de Souza, *Microscopy as a tool to follow deconstruction of lignocellulosic biomass*. *Formatex Research Center, Espanha*, 2012: p. 639-645.

34. Li, H., et al., *Investigation of lignin deposition on cellulose during hydrothermal pretreatment, its effect on cellulose hydrolysis, and underlying mechanisms*. Biotechnology and bioengineering, 2014. **111**(3): p. 485-492.
35. Gou, J.-Y., et al., *Compositional characterization and imaging of “wall-bound” acylesters of Populus trichocarpa reveal differential accumulation of acyl molecules in normal and reactive woods*. Planta, 2008. **229**(1): p. 15-24.
36. Fackler, K., et al., *Localisation and characterisation of incipient brown-rot decay within spruce wood cell walls using FT-IR imaging microscopy*. Enzyme and microbial technology, 2010. **47**(6): p. 257-267.
37. Fackler, K. and L.G. Thygesen, *Microspectroscopy as applied to the study of wood molecular structure*. Wood science and technology, 2013. **47**(1): p. 203-222.
38. Lewis, E.N., et al., *Fourier transform spectroscopic imaging using an infrared focal-plane array detector*. Analytical chemistry, 1995. **67**(19): p. 3377-3381.
39. Levin, I.W. and R. Bhargava, *FOURIER TRANSFORM INFRARED VIBRATIONAL SPECTROSCOPIC IMAGING: Integrating Microscopy and Molecular Recognition\**. Annu. Rev. Phys. Chem., 2005. **56**: p. 429-474.
40. Travis, A.J., et al., *Measurement of Cell Wall Volume using Confocal Microscopy and its Application to Studies of Forage Degradation*. Annals of Botany, 1997. **80**(1): p. 1-11.
41. Donaldson, L.A., *Lignification and lignin topochemistry — an ultrastructural view*. Phytochemistry, 2001. **57**(6): p. 859-873.
42. Bond, J., et al., *Safranine fluorescent staining of wood cell walls*. Biotechnic & Histochemistry, 2008. **83**(3-4): p. 161-171.
43. Olmstead, J. and D. Gray, *Fluorescence spectroscopy of cellulose, lignin and mechanical pulps: a review*. Journal of pulp and paper science, 1997. **23**(12): p. J571-J581.
44. Hernández-Hernández, H.M., et al., *Evaluation of agave fiber delignification by means of microscopy techniques and image analysis*. Microscopy and Microanalysis, 2014. **20**(5): p. 1436-1446.
45. Albinsson, B., et al., *The origin of lignin fluorescence*. Journal of Molecular Structure, 1999. **508**(1): p. 19-27.
46. Wondraczek, H., et al., *Photoactive polysaccharides*. Carbohydrate polymers, 2011. **83**(3): p. 1048-1061.
47. Castellan, A., V. Trichet, and J. Pommier, *Studied by UV/Vis Diffuse Reflectance and Fluorescence Spectroscopy*. Journal of pulp and paper science, 1995. **21**(9).
48. Castellan, A., et al., *Studies on fluorescence of cellulose*. Holzforschung, 2007. **61**(5): p. 504-508.
49. Rosenau, T., et al., *Isolation and identification of residual chromophores in cellulosic materials*. Polymer, 2004. **45**(19): p. 6437-6443.
50. Porter, S.E., et al., *Microscopic analysis of corn fiber using starch- and cellulose-specific molecular probes*. Biotechnology and bioengineering, 2007. **98**(1): p. 123-131.
51. Paës, G., *Fluorescent probes for exploring plant cell wall deconstruction: a review*. Molecules, 2014. **19**(7): p. 9380-9402.
52. Amigo, J.M. and C. Ravn, *Direct quantification and distribution assessment of major and minor components in pharmaceutical tablets by NIR-chemical imaging*. European Journal of Pharmaceutical Sciences, 2009. **37**(2): p. 76-82.

53. Budevska, B.O., S.T. Sum, and T.J. Jones, *Application of Multivariate Curve Resolution for Analysis of FT-IR Microspectroscopic Images of *i* in Situ*. Applied spectroscopy, 2003. **57**(2): p. 124-131.
54. Geladi, P., H. Grahn, and M. Manley, *Data analysis and chemometrics for hyperspectral imaging*. Raman, Infrared, and Near-Infrared Chemical Imaging, 2010: p. 93-107.
55. Gorzsás, A., et al., *Cell-specific chemotyping and multivariate imaging by combined FT-IR microspectroscopy and orthogonal projections to latent structures (OPLS) analysis reveals the chemical landscape of secondary xylem*. The Plant Journal, 2011. **66**(5): p. 903-914.
56. Monti, F., et al., *A multivariate statistical analysis approach to highlight molecular processes in plant cell walls through ATR FT-IR microspectroscopy: The role of the  $\alpha$ -expansin PhEXPA1 in *Petunia hybrida**. Vibrational Spectroscopy, 2013. **65**: p. 36-43.
57. Ruckebusch, C. and L. Blanchet, *Multivariate curve resolution: A review of advanced and tailored applications and challenges*. Analytica chimica acta, 2013. **765**: p. 28-36.
58. Salzer, R. and H.W. Siesler, *Infrared and Raman spectroscopic imaging*. 2009: John Wiley & Sons.
59. Patel, I.I., et al., *High contrast images of uterine tissue derived using Raman microspectroscopy with the empty modelling approach of multivariate curve resolution-alternating least squares*. Analyst, 2011. **136**(23): p. 4950-4959.
60. Windig, W. and J. Guilment, *Interactive self-modeling mixture analysis*. Analytical chemistry, 1991. **63**(14): p. 1425-1432.
61. Jaumot, J., et al., *A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB*. Chemometrics and Intelligent Laboratory Systems, 2005. **76**(1): p. 101-110.
62. Jaumot, J., A. de Juan, and R. Tauler, *MCR-ALS GUI 2.0: new features and applications*. Chemometrics and Intelligent Laboratory Systems, 2015. **140**: p. 1-12.
63. Burger, J. and P. Geladi, *Hyperspectral NIR image regression part II: dataset preprocessing diagnostics*. Journal of Chemometrics, 2006. **20**(3-4): p. 106-119.
64. Parastar, H., M. Jalali-Heravi, and R. Tauler, *Is independent component analysis appropriate for multivariate resolution in analytical chemistry?* TrAC Trends in Analytical Chemistry, 2012. **31**: p. 134-143.
65. Wang, G., Q. Ding, and Z. Hou, *Independent component analysis and its applications in signal processing for analytical chemistry*. TrAC Trends in Analytical Chemistry, 2008. **27**(4): p. 368-376.

# Capítulo 2

## Estrategia Analítica

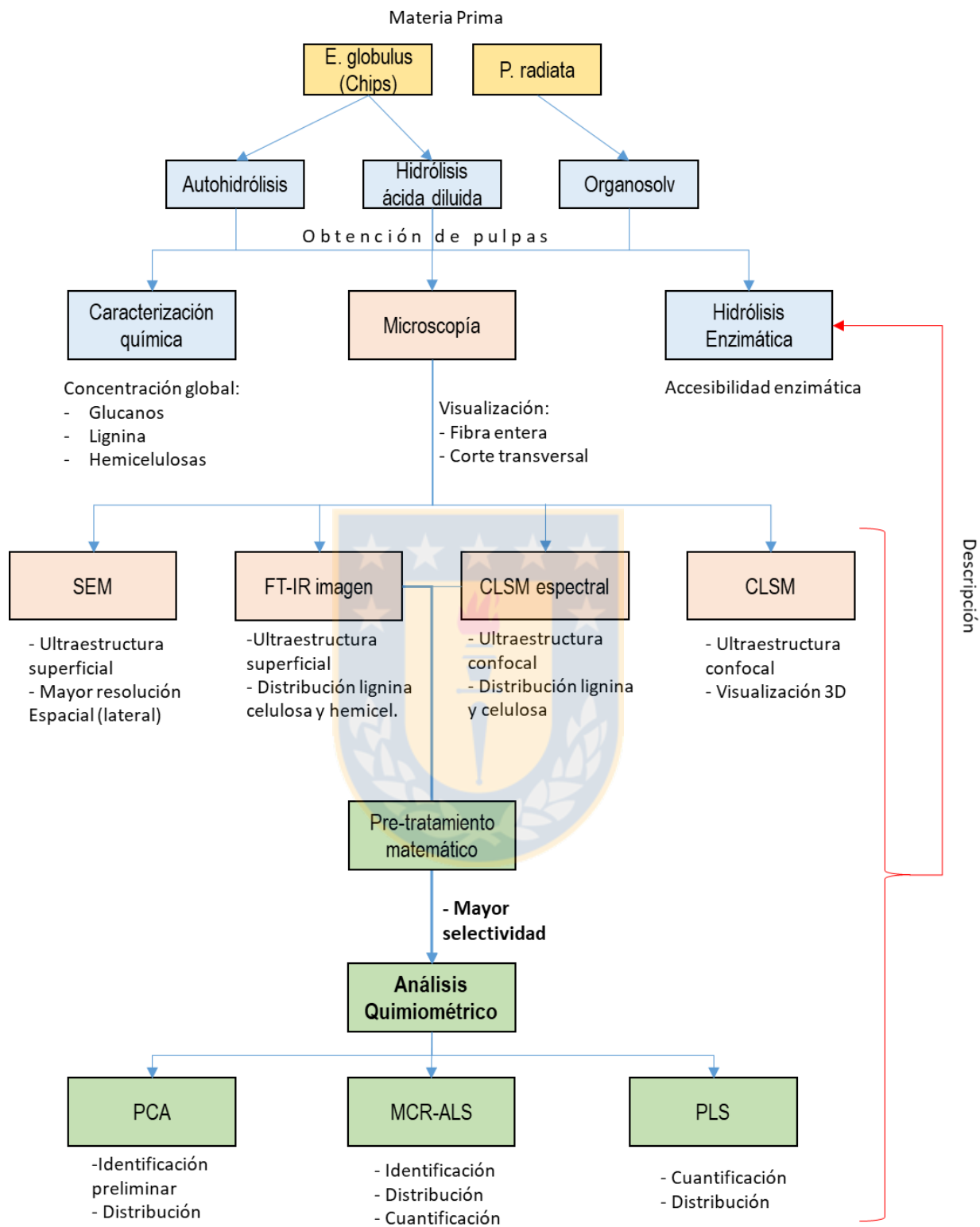


De acuerdo a los antecedentes expuestos en el capítulo 1, los métodos analíticos usados para el estudio de la química de la madera son en su mayoría destructivos y por lo tanto permiten conocer su composición química, pero no permiten saber cómo se encuentran distribuidos sus macrocomponentes en la fibra misma. El conocer en simultáneo la identidad y localización de los componentes químicos de la madera a escala microscópica, otorgaría una herramienta poderosa para el estudio de fenómenos como la recalcitrancia de la madera pre-tratada al ataque enzimático.

Este trabajo de tesis plantea una nueva plataforma analítica que permite el análisis químico cualitativo y cuantitativo en la superficie misma de la muestra lignocelulósica, a micro-escala y de forma no destructiva, a través del uso de imágenes espectrales y el refinamiento de la información química que éstas contienen mediante la aplicación de herramientas quimiométricas. Además, la aplicación de esta plataforma permite describir mediante imágenes los cambios químicos y estructurales que ocurren en la pared celular vegetal durante el proceso de pre-tratamiento de material lignocelulósico para su posterior hidrólisis enzimática.

La estrategia analítica empleada también se vale de las herramientas de química analítica clásicamente usadas en el estudio de la química de la madera a fin de caracterizar a nivel macroscópico las muestras de trabajo. Posteriormente se realiza el análisis químico localizado usando diferentes técnicas de imagen espectral (FT-IR y fluorescencia espectral) y análisis multivariado, lo que permitiría describir los patrones de distribución de celulosa, lignina y hemicelulosas en la fibra de madera, como se muestra en la Figura 2.1.





**Figura 2.1:** Esquema de la estrategia analítica para la caracterización química *in situ* de fibras lignocelulósicas (propiedad del autor)

La desarrollo de la estrategia analítica tuvo por lo tanto los siguientes pasos:

1. Obtención de pulpas lignocelulósicas:

Se usaron chips de madera de dos especies diferentes como materia prima para los análisis, *Eucalyptus globulus* (madera dura) y *Pinus radiata* (madera blanda). Se aplicaron tres técnicas de pulpage diferentes: las muestras de *E. globulus* fueron pre-tratadas mediante hidrólisis con ácido diluido (HAD) y por auto-hidrólisis (AH), mientras que las muestras de *P. radiata* fueron pre-tratadas mediante Organosolv. Las condiciones para cada pretratamiento fueron seleccionadas mediante diseño experimental, aunque en este caso no se midieron variables de respuesta ni se realizó una optimización. El objetivo fue ordenar una serie de condiciones experimentales con el fin de obtener una alta variabilidad en el contenido de celulosa, lignina y hemicelulosa, y sus patrones de distribución en la superficie de la fibra lignocelulósica. La severidad del pre-tratamiento fue expresada como  $S_0$  que corresponde al logaritmo del factor de severidad ( $R_0$ ), calculados según formula 1 y 2.

$$R_0 = \int_0^{t_f} \exp^{(T_{exp}-100/14.75)} \cdot dt \quad (\text{Formula 1})$$

$$S_0 = \text{Log } R_0 \quad (\text{Formula 2})$$

En la Tabla 2.1 se presentan los once experimentos del pre-tratamiento HAD con sus respectivas condiciones de pretratamiento, porcentaje de recuperación de sólidos (%Rs) y severidad.

**Tabla 2.1:** Condiciones de pretratamiento por hidrólisis con ácido diluido ( $\text{H}_2\text{SO}_4$ ) para *Eucalyptus globulus*. (propiedad del autor)

N° Exp	Pretratamiento	%Rs	S <sub>0</sub>
1	120° C/ 0.10%/ 1:4	97.9	3.22
2	170° C/ 0.10%/ 1:4	83.3	4.72
3	120° C/ 0.75%/ 1:4	97.6	3.26
4	170° C/ 0.75%/ 1:4	65.9	4.74
5	120° C/ 0.10%/ 1:8	97.4	3.33
6	170° C/ 0.10%/ 1:8	82.6	4.71
7	120° C/ 0.75%/ 1:8	94.9	3.24
8	170° C/ 0.75%/ 1:8	73.7	4.83
9	145° C/ 0.43%/ 1:6	94.1	4.02
10	113° C/ 0.43%/ 1:6	96.8	2.76
11	177° C/ 0.43%/ 1:6	71.0	5.03

En la Tabla 2.2 se presentan los seis experimentos del pre-tratamiento AU con sus respectivas condiciones de pretratamiento, porcentaje de recuperación de sólidos (%Rs) y severidad.

**Tabla 2.2:** Condiciones de pretratamiento por autohidrólisis para *Eucalyptus globulus* (propiedad del autor)

N° Exp	Pretratamiento	%Rs	S <sub>0</sub>
1	190° C/ 70 min	66.1	5.89
2	175° C/ 81 min	64.1	5.46
3	160° C/ 70 min	81.0	4.93
4	190° C/ 15 min	70.9	5.33
5	160° C/ 15 min	94.1	4.40
6	175° C/ 4 min	86.4	4.39

De las muestras pre-tratadas por HAD y AU, se seleccionó una de alta severidad y otra de baja severidad para la obtención de imágenes, a fin de observar el impacto de la severidad en la ultraestructura de la fibra de madera.

En la Tabla 2.3 se muestran los experimentos del pre-tratamiento por Organosolv. En el caso del pre-tratamiento por Organosolv, la selección de muestras para el análisis microscópico se hizo buscando muestras que tuviesen contenidos similares de glucanos y lignina, pero diferencias significativas en los rendimientos de hidrólisis enzimática.



**Tabla 2.3:** Condiciones de pre-tratamiento por Organosolv para *Pinus radiata* (propiedad del autor)

Pulpa	%H <sub>2</sub> SO <sub>4</sub>	% EtOH/H <sub>2</sub> O	T° (°C)	Tiempo (min)
PR1	0.7	50/50	180	30
PR2	0.7	50/50	190	30
PR3	0.35	50/50	190	30
PR4	0.35	25/75	190	30
PR5	0.35	50/50	190	30
PR6	0.7	25/75	190	30
PR7	0.7	25/75	180	30
PR8	0.35	25/75	180	30
PR9	0.35	50/50	180	30

### 2. Caracterización química de pulpas

La madera pre-tratada fue caracterizada por métodos convencionales. La hidrólisis fue realizada en dos etapas, una con 4% de H<sub>2</sub>SO<sub>4</sub> a temperatura ambiente y la otra con 72% H<sub>2</sub>SO<sub>4</sub> a 121°C por 30 min. La cuantificación de carbohidratos (glucosa, celobiosa, y hemicelulosas) en la fracción soluble fue realizada por HPLC en un cromatógrafo Merck Hitachi con una columna Aminex HPX-87H a 45°C con un flujo de 0,6 mL min<sup>-1</sup> con 5mM de H<sub>2</sub>SO<sub>4</sub>, usando un detector de índice de refracción. Se realizó una conversión matemática de los monómeros de azúcares en anhidro-monómeros usando un factor de 0.9 para glucosa-glucano y 0.95 para celobiosa-glucano. El contenido total de glucanos fue determinado mediante la suma de ambas fracciones de anhidro-monómeros. Las hemicelulosas fueron medidas desde un solo pico cromatográfico, el que contenía una mezcla de xilosa, manosa y galactosa, y fueron convertidos en unidades monoméricas aplicando un factor de 0.88.

El contenido de lignina se determinó por gravimetría para la fracción insoluble, la que fue recolectada usando un filtro Gosh. La lignina soluble se determinó mediante espectroscopía en el UV midiendo la absorbancia a 205 nm y considerando un coeficiente de extinción de 110 g L<sup>-1</sup> cm<sup>-1</sup>. El cálculo de la lignina total es la suma de ambas fracciones. Los glucanos, lignina, y hemicelulosas fueron expresados como porcentaje respecto del peso seco de las muestras (% p/p).

### 3. Hidrólisis enzimática

La accesibilidad enzimática de las muestras pre-tratadas fue evaluada en función de rendimiento de hidrólisis enzimática (%H.E.).

Para la H.E. se pesaron las pulpas húmedas equivalentes a 3g de peso seco en matraces Erlenmeyer. La H.E. se realizó con una preparación comercial de complejo enzimático de celulasas ((NS-22128 CCN03128; 71 FPU mL<sup>-1</sup>, Novozymes A/S, Bagsværd, Dinamarca) suplementado con  $\beta$ -glucosidasa (NS-22128 DCN00216; 265 CBU mL<sup>-1</sup>, Novozymes A/S, Bagsværd, Dinamarca). La cantidad de enzima usada fue de 20 FPU de celulasa y 40 CBU de  $\beta$ -glucosidasa por gramo de muestra seca. La incubación se realizó a 60° en agitación constante. Se recolectaron muestras cada 24, 36, 48 y 72 h para estimar el rendimiento máximo, el que se produjo a las 72 h. El rendimiento de H.E. fue calculado en función del contenido de glucosa, el que fue medido según lo descrito en el punto (ii) de este capítulo. El %H.E fue reportado como la media de medidas en triplicado.

#### 4. Obtención de imágenes microscópicas

Para SEM, las muestras fueron secadas por punto crítico y cubiertas con una capa de oro. Las imágenes fueron obtenidas con un equipo Jeol JSM-6380LV con un detector de electrones secundarios, usando un voltaje de aceleración entre 15-20 kV.

Las imágenes por CLSM (obtenidas como fluorescencia total a 490-530 nm) fueron adquiridas usando un microscopio confocal espectral Zeiss LSM780 con un láser de Ar a 488 nm de excitación y un objetivo de 20x0.50 Plan-Neofluar.

Para el análisis de cortes transversales las muestras fueron congeladas a -80°C en medio de congelación NEG50<sup>TM</sup> y cortadas a un grosor de 10  $\mu$ m usando un criostato HM525 NX Thermo Scientific en un ambiente controlado a -20°C. Los cortes fueron depositados en un cubre objeto (el grosor del vidrio tiene efecto en las señales de fluorescencia).

Para la adquisición de imágenes espectrales de emisión, los cubre objeto con las muestra cortadas fueron montados en un microscopio confocal espectral Zeiss LSM780 y se utilizó el láser a 405 nm para la excitación, con un aumento de 63X (el láser es diferente al usado para las imágenes obtenidas por fluorescencia total). El espectro de emisión fue registrado entre 484-689 nm con una resolución espectral de 4,4 nm. La información espectral de las imágenes para el análisis de datos fue extraída usando el software ImageJ y Matlab R2013a (Mathworks).

Se adquirieron imágenes por microscopía FT-IR de la superficie de fibras así como de cortes transversales. Para ello un microscopio FT-IR Spectrum Frontier/Spotlight 400

(PerkinElmer) con un arreglo lineal de 16 detectores MCT. Las muestras fueron previamente secadas por 8 h a 45°C. Los espectros fueron tomados por reflectancia total atenuada (ATR), usando un cristal de germanio. El tamaño de píxel escogido fue de 1,56  $\mu\text{m}$  x 1,56  $\mu\text{m}$ , la resolución espectral de 8  $\text{cm}^{-1}$ , un rango espectral de 4000-748  $\text{cm}^{-1}$  y 16 scans por espectro. Para la mayoría de los análisis la zona entre 4000-2000  $\text{cm}^{-1}$  fue eliminada enfocar el análisis en la región de huella dactilar del infrarrojo.

##### 5. Análisis quimiométrico de las imágenes microscópicas obtenidas

Todos los pre-tratamientos y análisis matemáticos fueron realizados usando Matlab® R2013a (Mathworks, inc), exceptuando PLS, el que fue realizado en Pirouette® (Infometrics, inc). Se aplicaron diferentes transformaciones a las imágenes espectrales, por ejemplo MSC, suavizado y primera derivada por Savitzky-Golay, normalización, etc. dependiendo de las características de los espectros observados en cada una. Los espectros usados para los modelos PLS fueron pre-procesados por centrado en la media, luego de haber sido transformados. Antes del pre-tratamiento espectral, las imágenes espectrales fueron desdobladas, y analizadas mediante métodos quimiométricos, como muestra la Fig. 2.1. Las imágenes obtenidas por microscopía FT-IR fueron analizadas y reconstruidas por las tres técnicas quimiométricas descritas (PCA, PLS y MCR-ALS), mientras que las imágenes obtenidas por CLSM espectral fueron analizadas y reconstruidas usando MCR-ALS solamente.

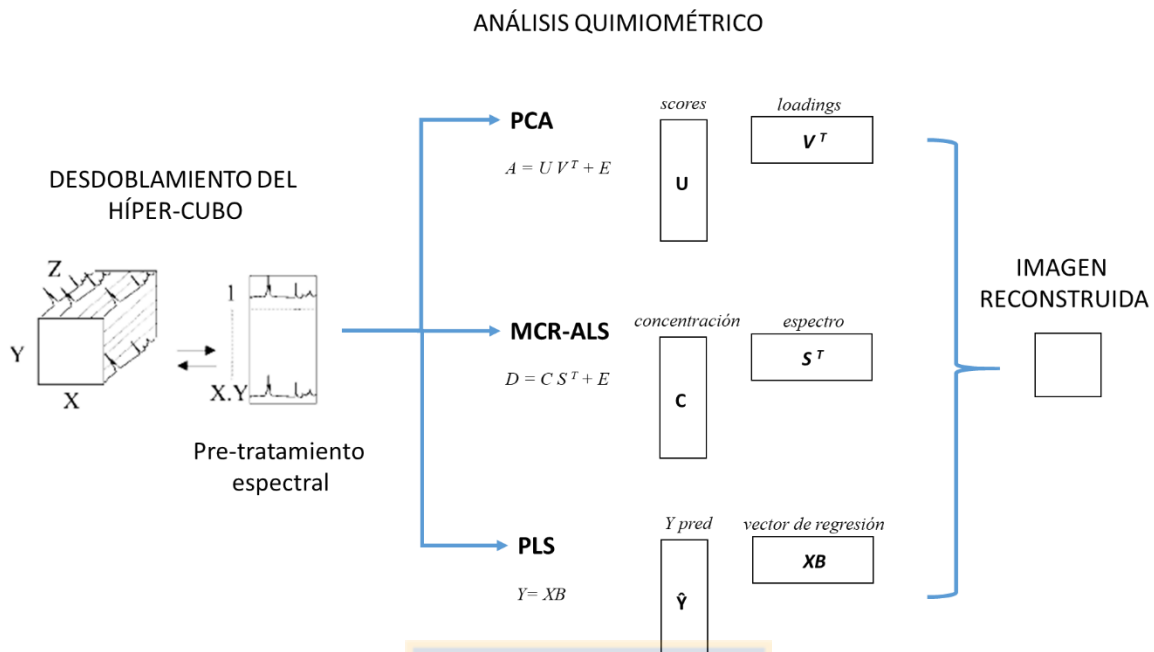


Figura 2.2: Esquema de análisis de imágenes espectrales mediante métodos quimiométricos (propiedad del autor)



## Capítulo 3

### Resultados: Enfoque Cualitativo



El presente capítulo agrupa los resultados de diferentes análisis que tienen un carácter cualitativo, puesto que entregan información de la identidad química de los componentes presentes en las fibras lignocelulósicas en un contexto espacial, así como la morfología y ultraestructura de éstas. En particular este capítulo responde a los objetivos específicos 1, 3 y 4.

El capítulo se encuentra dividido en dos secciones. La sección 1 muestra resultados de análisis enfocados en estudiar a nivel microscópico la topografía de la superficie de fibras lignocelulósicas. Algunos autores afirman que la correlación entre lignina e hidrólisis enzimática no se explica por la cantidad de lignina, sino por otros cambios en el proceso de remoción de lignina [1] y que el concepto más importante en el pre tratamiento es que la lignina puede ser “removida” desde la mayoría de la superficie de la fibra, sin ser necesariamente removida de la biomasa [2]. A fin de poner a prueba estas afirmaciones y al mismo tiempo comparar la distribución microscópica de la lignina y los otros componentes presentes en la superficie de la fibra, se usaron las fibras pre-tratadas por Organosolv con diferente severidad en muestras de *P. radiata*. La ultraestructura de las fibras fue visualizada mediante CLSM y SEM. Mediante microscopía FT-IR y análisis quimiométrico (PCA y MCR-ALS) se determinó la identidad química de los principales componentes químicos presentes en la superficie. A fin de facilitar la identificación de los espectros infrarrojo resueltos por MCR-ALS, en los resultados se muestran espectros infrarrojo de celulosa y lignina obtenidos desde patrones puros

Los resultados incluyen también la caracterización química de las pulpas de *Pinus* así como los rendimientos en hidrólisis enzimática obtenidos para cada una, calculado como el porcentaje de glucosa obtenido post-hidrólisis enzimática con respecto a la glucosa potencial de la pulpa. Si bien en la sección 1 se informa la caracterización química de todas las pulpas de *P. radiata* obtenidas, el análisis de imágenes microscópicas y la discusión de ellas se realizaron basándose sólo en 2 muestras contrastantes de interés.

La sección 2 muestra la aplicación de técnicas analíticas para estudiar el interior de la fibra lignocelulósica, es decir, en la pared misma de la célula, a fin de determinar qué ocurre con la lignina en esas zonas producto de los diferentes pre-tratamientos aplicados. Para ello se utilizaron pulpas de *P. radiata* pre-tratadas por Organosolv, y pulpas de *E. globulus* pre-

tratadas por autohidrólisis e hidrólisis ácida diluida. Para acceder al interior de las fibras se realizaron cortes transversales con micrótopo de fibras embebidas en resina epoxi, y cortes con criostato usando un medio de congelación a  $-80^{\circ}\text{C}$ . Los cortes fueron analizados por CLSM espectral y microscopía FT-IR, y las imágenes obtenidas fueron analizadas por técnicas quimiométricas (MCR-ALS) para localizar la posición de los componentes químicos principales de la pulpa. Para el análisis se seleccionaron dos muestras por cada pretratamiento, una proveniente de condiciones de alta severidad y otra de baja a fin de describir los cambios producidos por dichos pretratamientos en la distribución de lignina, celulosa y hemicelulosa.



## Sección 1

### **Análisis químico superficial a nivel microscópico de fibras pre-tratadas de *Pinus radiata* mediante SEM, CLSM y microscopía FT-IR.**

La Tabla 2.3 (en el capítulo 2) muestra las condiciones usadas para el pulpaje por Organosolv de nueve muestras de *P. radiata*, las que se obtuvieron mediante el uso de un diseño experimental. El objetivo del diseño no fue estudiar las condiciones de reacción que propiciarán mayor o menor contenido de lignina total, sino la obtención de diferentes fibras con similar contenido de lignina y celulosa, pero un patrón de distribución diferente, con el propósito de estudiar si hay alguna correlación con el rendimiento de hidrólisis enzimática. Al respecto, las fibras PR2 y PR3 reúnen las características para este objetivo, puesto que la cantidad de lignina y celulosa totales cuantificadas en ambas no tienen diferencias estadísticamente significativas (ver Tabla 3.1), sin embargo el rendimiento en la hidrólisis enzimática presenta variaciones drásticas ( $92,59 \pm 2,9\%$  para PR2 y  $43,61 \pm 1,7\%$  para PR3). Por lo tanto, el análisis microscópico y espectroscópico se enfocó en estas dos muestras.

**Tabla 3.1:** Cuantificación del porcentaje de lignina y celulosa total, obtenida mediante método Klason, de dos pulpas Organosolv, y cálculo del rendimiento de la hidrólisis enzimática (H.E.), mediante cuantificación por HPLC de azúcares post hidrólisis. La diferencia en el porcentaje de rendimiento de hidrólisis entre PR2 y PR3 es estadísticamente significativa (\*\*\*) =  $p < 0,001$ . (Propiedad del autor).

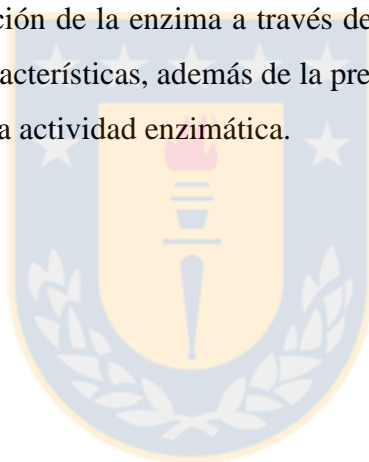
<b>Pulpa</b>	<b>% Lignina</b>	<b>% Glucanos</b>	<b>% Rendimiento H.E. 72 h</b>
PR2	$21,7 \pm 0,6$	$80,5 \pm 3,9$	$92,6 \pm 2,9$ ***
PR3	$23,8 \pm 2,7$	$73,5 \pm 3,3$	$43,6 \pm 1,7$

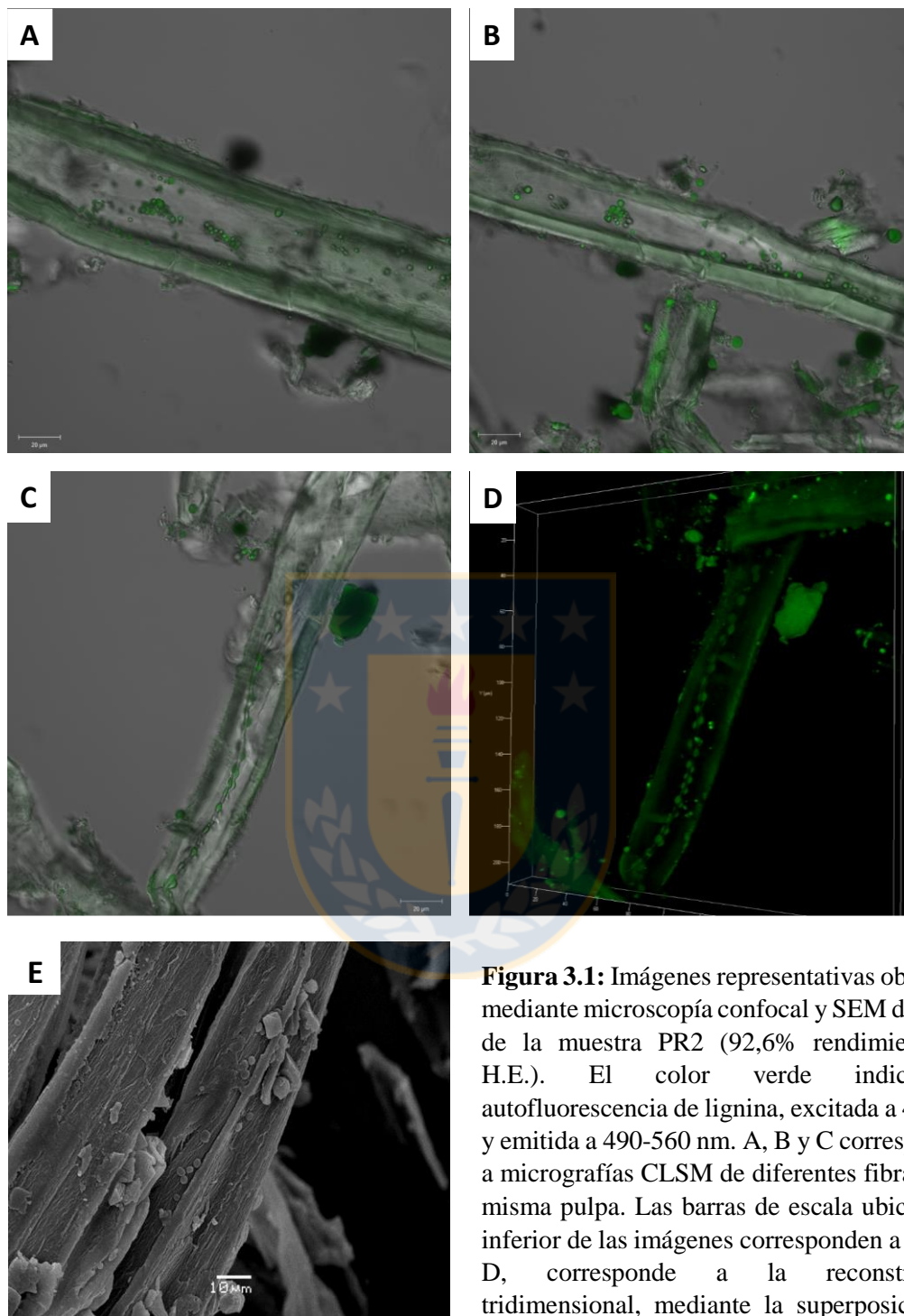
La autofluorescencia de la lignina permite su visualización y distribución en la fibra, mediante microscopía confocal. Las Figuras 3.1 y 3.2 muestran fibras representativas presentes en ambas pulpas. Para la muestra PR2, de alto rendimiento enzimático, se puede observar la frecuente presencia de cúmulos y cuerpos esféricos de lignina, tanto al interior como al exterior cuyo tamaño varía entre 3-5 micras de diámetro para las esferas y  $>5$  micras para los cúmulos. En el caso de la muestra PR3, la presencia de gotitas de lignina es menos



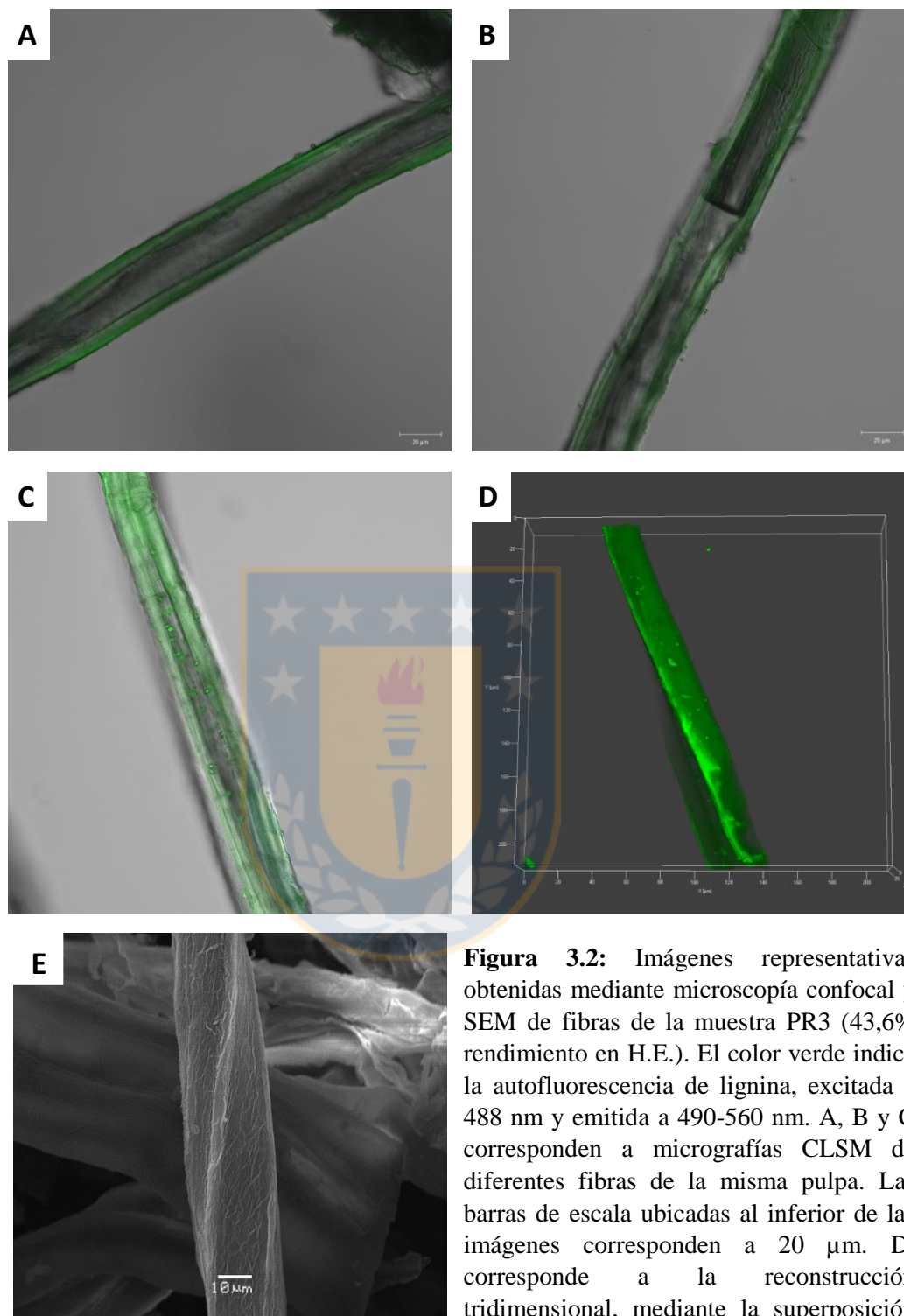
frecuente, y se observan que su tamaño varía entre 1-2 micras. Esta característica microscópica podría explicar, al menos en parte, la diferencia en el rendimiento enzimático de ambas pulpas. Sin embargo, la presencia de estas esferas parece concentrarse en el interior de las fibras. Tanto en la Figura 3.1 D como en 3.2 D se observan menos esferas en la superficie de las fibras, por lo que la presencia de éstas sería más común en el interior.

Estudios realizados por Donohoe *et al* [2] demostraron, mediante diferentes técnicas microscópicas, que las regiones adyacentes a estas gotas coalescentes de lignina se abren, propiciando un incremento en la superficie disponible y mejorando la accesibilidad enzimática. Si bien más autores han corroborado que la presencia de gotitas de lignina está relacionada con un mayor rendimiento enzimático [1] otros estudios afirman que la deposición de gotitas en la superficie de la fibra inhibe la actividad enzimática al suponer un impedimento para la movilización de la enzima a través de la fibra de celulosa [3]. Por lo tanto, pudiesen existir otras características, además de la presencia de gotitas de lignina, que determinen las variaciones en la actividad enzimática.





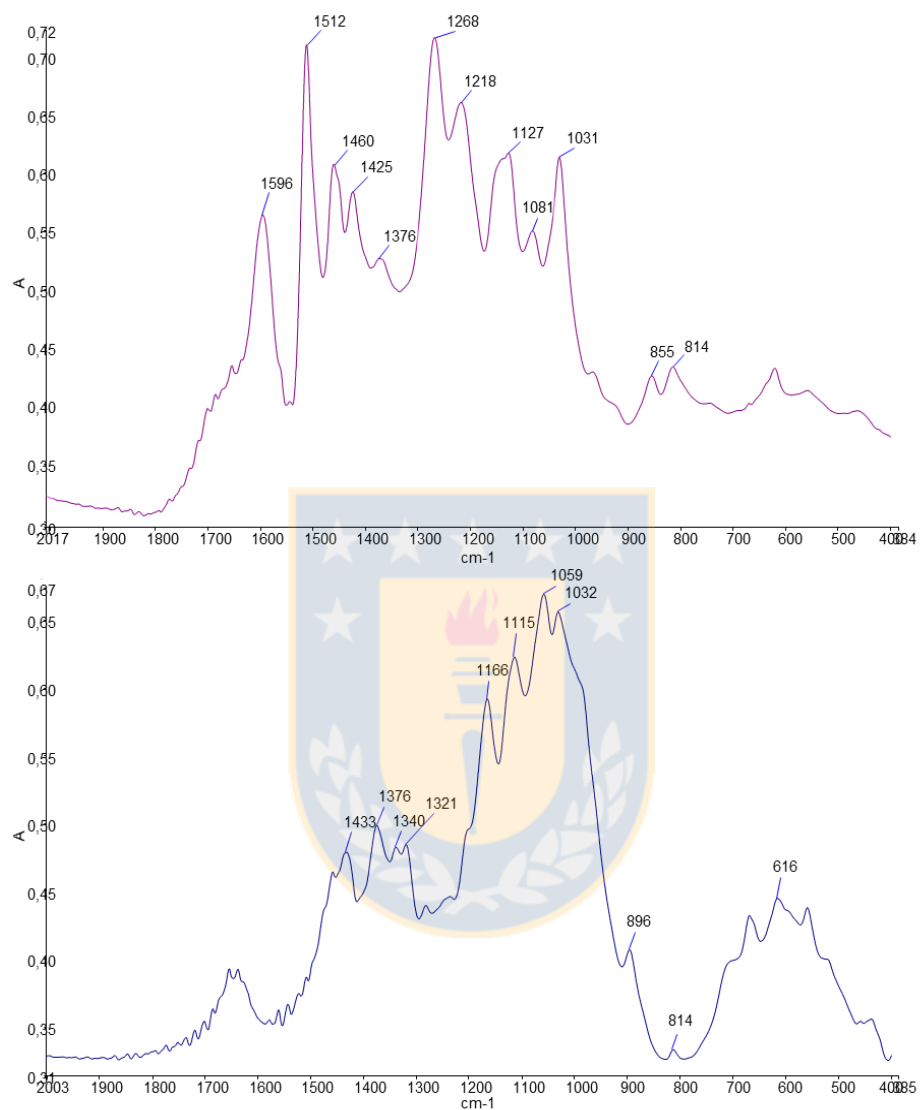
**Figura 3.1:** Imágenes representativas obtenidas mediante microscopía confocal y SEM de fibras de la muestra PR2 (92,6% rendimiento en H.E.). El color verde indica la autofluorescencia de lignina, excitada a 488 nm y emitida a 490-560 nm. A, B y C corresponden a micrografías CLSM de diferentes fibras de la misma pulpa. Las barras de escala ubicadas al inferior de las imágenes corresponden a 20  $\mu$ m. D, corresponde a la reconstrucción tridimensional, mediante la superposición de diferentes planos focales, de la fibra mostrada en C. E, micrografía SEM. (Propiedad del autor).



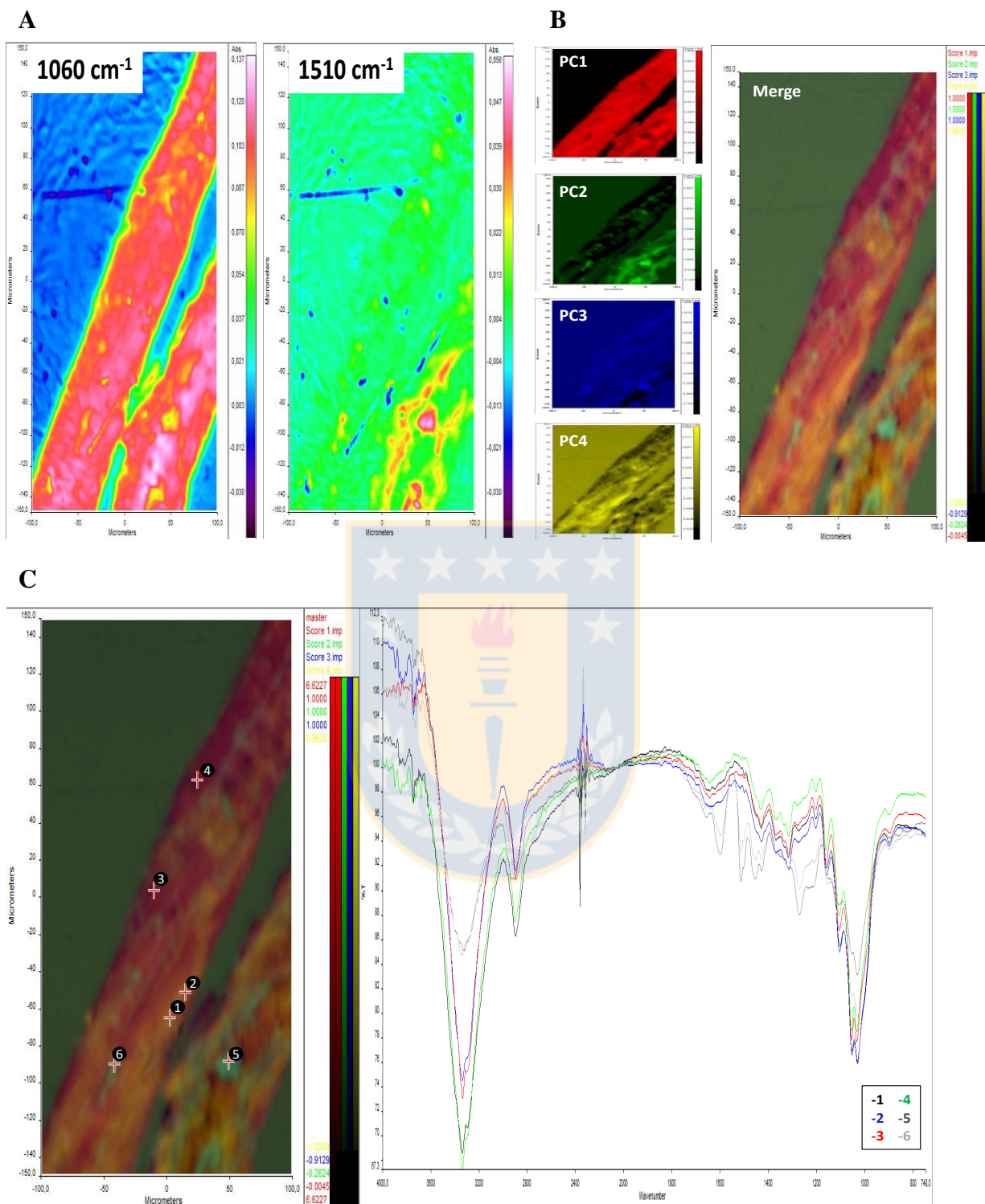
**Figura 3.2:** Imágenes representativas obtenidas mediante microscopía confocal y SEM de fibras de la muestra PR3 (43,6% rendimiento en H.E.). El color verde indica la autofluorescencia de lignina, excitada a 488 nm y emitida a 490-560 nm. A, B y C corresponden a micrografías CLSM de diferentes fibras de la misma pulpa. Las barras de escala ubicadas al inferior de las imágenes corresponden a 20  $\mu\text{m}$ . D, corresponde a la reconstrucción tridimensional, mediante la superposición de diferentes planos focales, de la fibra mostrada en C. E, micrografía SEM. (Propiedad del autor).

El análisis por microscopía FT-IR tiene la ventaja de localizar en la topografía de la fibra sus diferentes componentes químicos, en base a su espectro infrarrojo. A fin de facilitar la interpretación espectral de las imágenes a analizar, se tomaron espectros infrarrojos mediante espectroscopía convencional (sin microscopía) de muestras comerciales de celulosa y lignina Kraft. Los espectros se muestran en la Figura 3.3 y sólo son referenciales, pues la posición de las bandas varía en cada pulpa dependiendo del tipo de madera y del efecto del pretratamiento, especialmente para lignina.

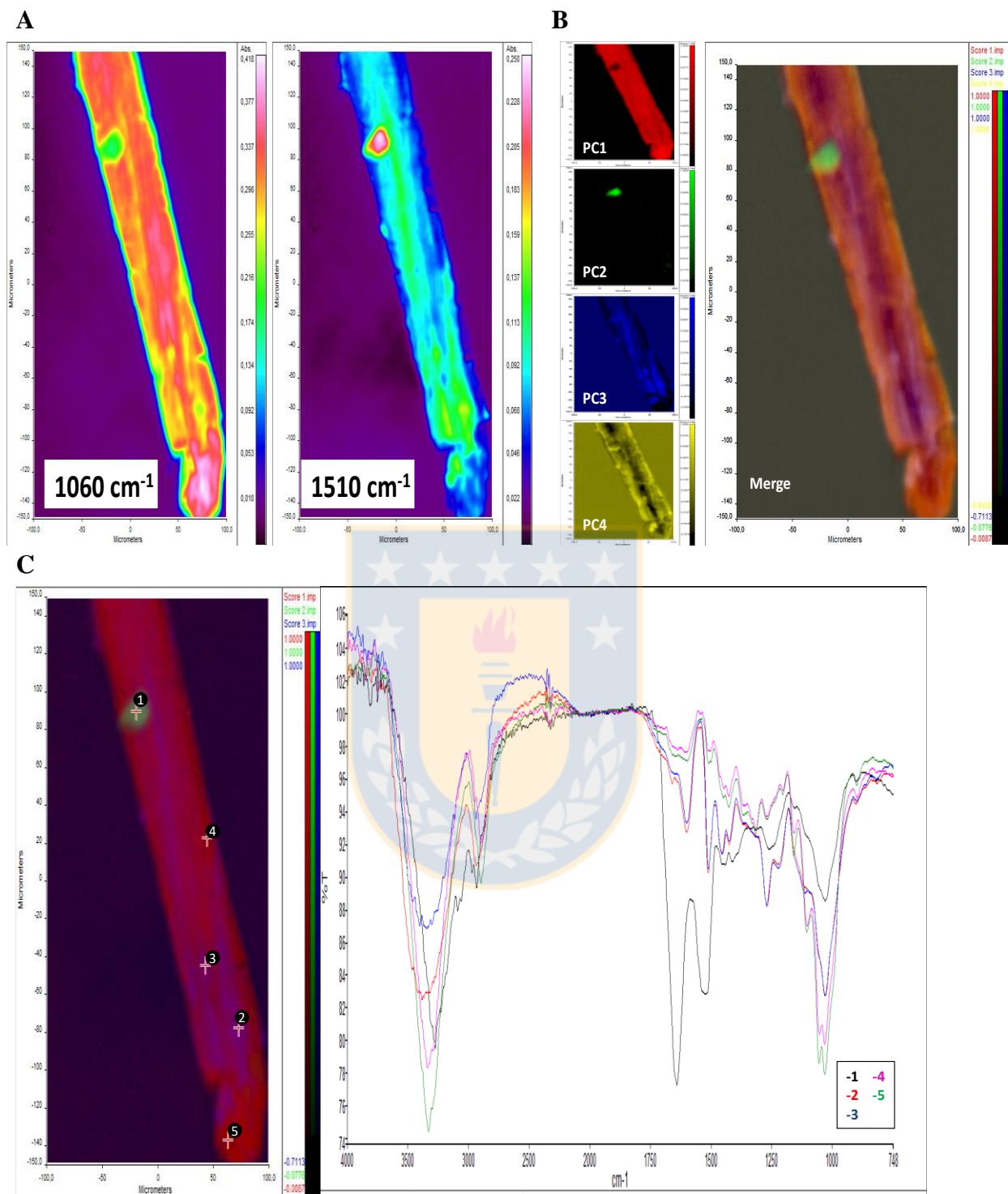
Las Figuras 3.4 y 3.5 muestran imágenes micro-espectrales de las fibras PR2 y PR3, respectivamente. El en panel A se observa la imagen de la fibra a un número de onda determinado. La banda de  $1510\text{ cm}^{-1}$  está relacionada a la lignina y permite visualizar las zonas en las que ésta se estaría acumulando más. Con respecto a esta banda, la muestra PR3 (Fig. 3.5) tiene una absorbancia mucho mayor que la muestra PR2 (Fig 3.4) lo que indicaría que el contenido de lignina en superficie es mayor. La banda de  $1060\text{ cm}^{-1}$  está relacionada con los polisacáridos y permite diferenciar la fibra del fondo de la imagen. El panel B muestra la reconstrucción de las imágenes por sus componentes principales (PCA). Este procesamiento de la imagen es entregado por el software de control de microscopio FT-IR y permite visualizar diferentes zonas con características espectroscópicas similares. El panel C, muestra espectros en diferentes puntos de la imagen, obtenidos desde las distintas zonas diferenciadas por la reconstrucción de la imagen por componentes principales. Los resultados muestran que la reconstrucción por PCA de la imagen diferencia zonas ricas en ligninas de zonas con menor lignina, donde se observa casi sólo celulosa. Esta visualización coincide con las imágenes obtenidas por microscopía confocal, pues es posible apreciar la presencia de cúmulos de lignina en la fibra PR2. Por otro lado, la muestra PR3 muestra una distribución más homogénea de la lignina en la superficie de la fibra, lo que pudiese tener relación con la mayor recalcitrancia, ya que al recubrir más uniformemente la fibra habría una menor superficie de contacto con la enzima. Este tipo de reconstrucción por componentes principales tiene la ventaja, por sobre el análisis monovariado, de permitir una visualización preliminar más completa de los componentes que están presentes en la superficie de la fibra.



**Figura 3.3:** Espectros FT-IR de patrones de lignina alcalina Kraft (superior) y celulosa (inferior). (Propiedad del autor).



**Figura 3.4:** Microscopía FT-IR de fibra PR2, de alto rendimiento en hidrólisis enzimática. A, Imagen infrarrojo de zonas de absorción característica para celulosa (1060 cm<sup>-1</sup>) y lignina (1510 cm<sup>-1</sup>). B, Reconstrucción de imagen por PCA, identificándose los primeros 4 componentes principales (PC). C, Espectros de transmitancia obtenidos de 6 zonas específicas de la imagen reconstruida por PCA. (Propiedad del autor).

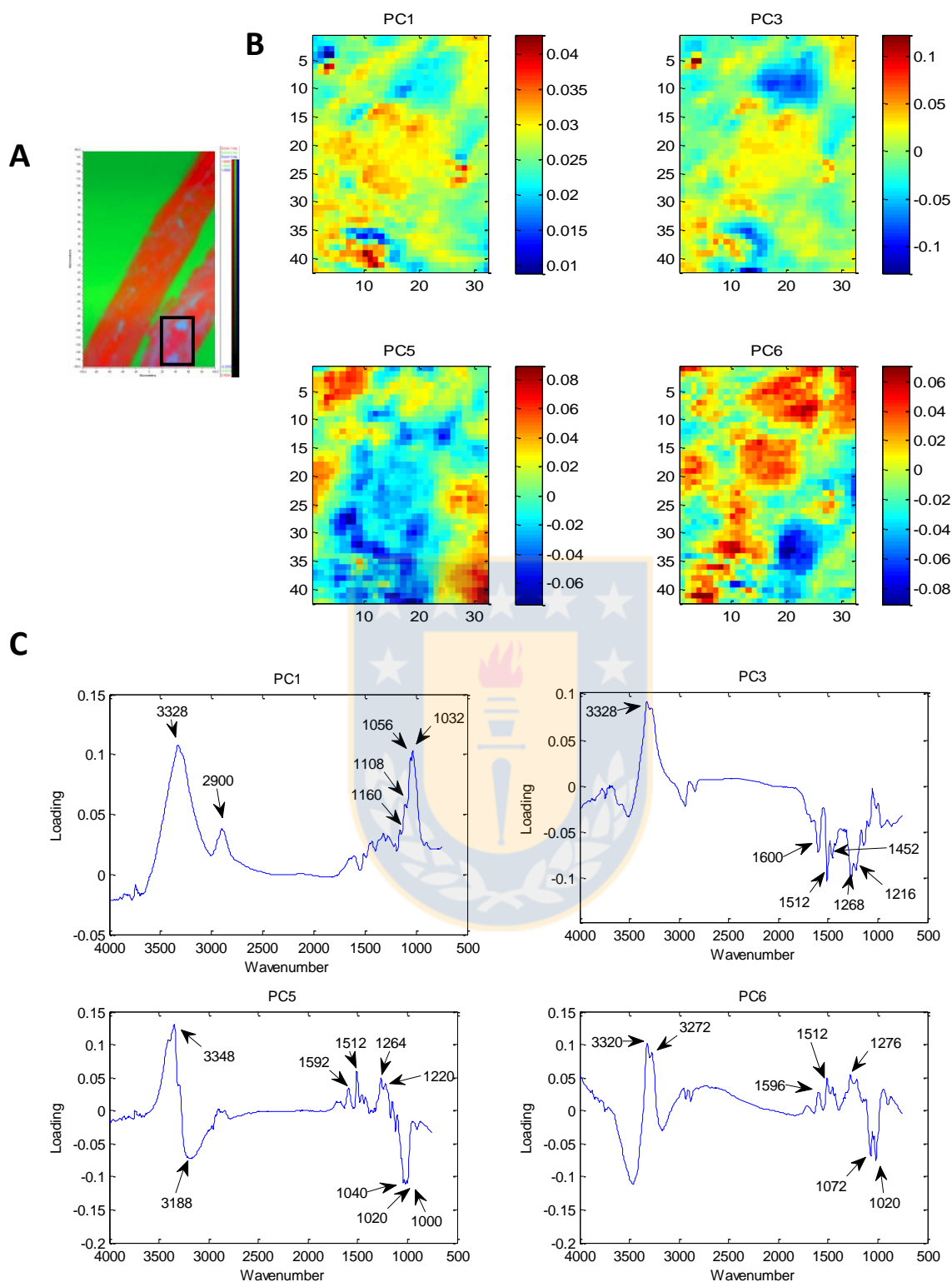


**Figura 3.5:** Microscopía FT-IR de fibra PR3, de bajo rendimiento en hidrólisis enzimática. A, Imagen infrarrojo de zonas de absorción característica para celulosa ( $1060\text{ cm}^{-1}$ ) y lignina ( $1510\text{ cm}^{-1}$ ). B, Reconstrucción de imagen por PCA, identificándose los primeros 4 componentes principales. C, Espectros de transmitancia obtenidos desde 5 zonas específicas de la imagen reconstruida por PCA. (Propiedad del autor).

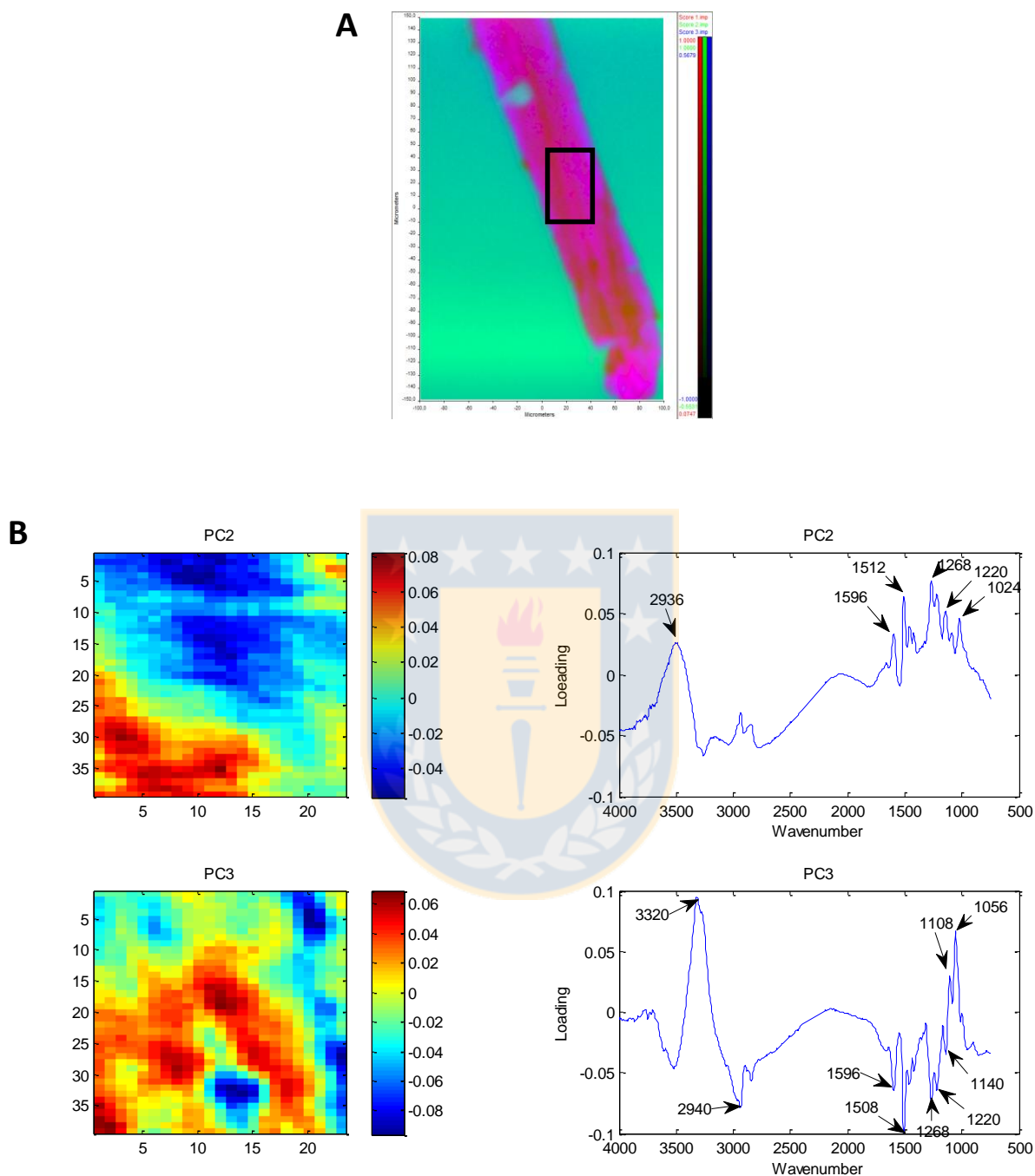
Las Figuras 3.6 y 3.7 muestran un acercamiento y análisis por componentes principales a zonas de interés de la fibra. En la Figura 3.6 A, se observan la reconstrucción por scores de PCA de la imagen. El panel B permite identificar las bandas que mayor peso tienen en la diferenciación de estos componentes. Se puede observar, por ejemplo, que en el componente principal (PC) 1 predominan las bandas asociadas a celulosa, mientras que en el PC3 existe una fuerte influencia de las bandas asociadas a lignina, pero en un sentido negativo, es decir, la zona se caracteriza por la ausencia de estas bandas. Por otro lado, los PC5 y PC6 tienen un aporte en sentido positivo de las bandas asociadas a lignina, pero un aporte mayor de las bandas asociadas a celulosa, pero en sentido negativo. Esto quiere decir que estas zonas o componentes se encuentran descritos por presencia de lignina y ausencia de celulosa. Los otros componentes principales (PC2 y PC4) no son mostrados en la figura puesto que sus loadings no contienen información química relevante (por ejemplo, muestran desplazamiento de línea base). En la Figura 3.7 se observa algo similar, existen una contraposición de las bandas asociadas a lignina, con las bandas asociadas a celulosa, lo que permite clasificar estas regiones de la imagen en base a la presencia de un compuesto y la ausencia del otro.

Este tipo de resultados tiene ventajas y desventajas. Por un lado, es posible la visualización rápida de los principales compuestos presentes en la muestra y su distribución. Por otro lado, no es posible identificar pequeñas variaciones espectrales entre componentes químicos, ya que éstas estarían explicadas por una pequeña cantidad de la varianza. El PCA otorga mayor peso a las diferencias más grandes (que describen mayor una varianza), por lo tanto los primeros componentes principales estarían dominados por las bandas principales de celulosa y lignina, los cuales son los compuestos mayoritarios de la pulpa. Esto último podría ser un problema mayor, puesto que distintos autores han descrito modificaciones en la identidad de celulosa o lignina e incluso asociaciones entre ellos producto del pre-tratamiento, el que sería mayor a medida que aumenta la temperatura o la concentración de ácido [4-6].





**Figura 3.6:** A, Imagen reconstruida por PCA de la fibra PR2. El marco negro indica una región rica en lignina seleccionada para un análisis más detallado. B y C, muestran los scores y loading respectivamente del PCA de la zona aumentada. El loading del PC1 tiene gran similitud con el espectro de celulosa, mientras que los loading del PC3, PC5 y PC6 tienen bandas dominantes, con valores positivos y/o negativos, en las regiones asociadas a celulosa y lignina. (Propiedad del autor).

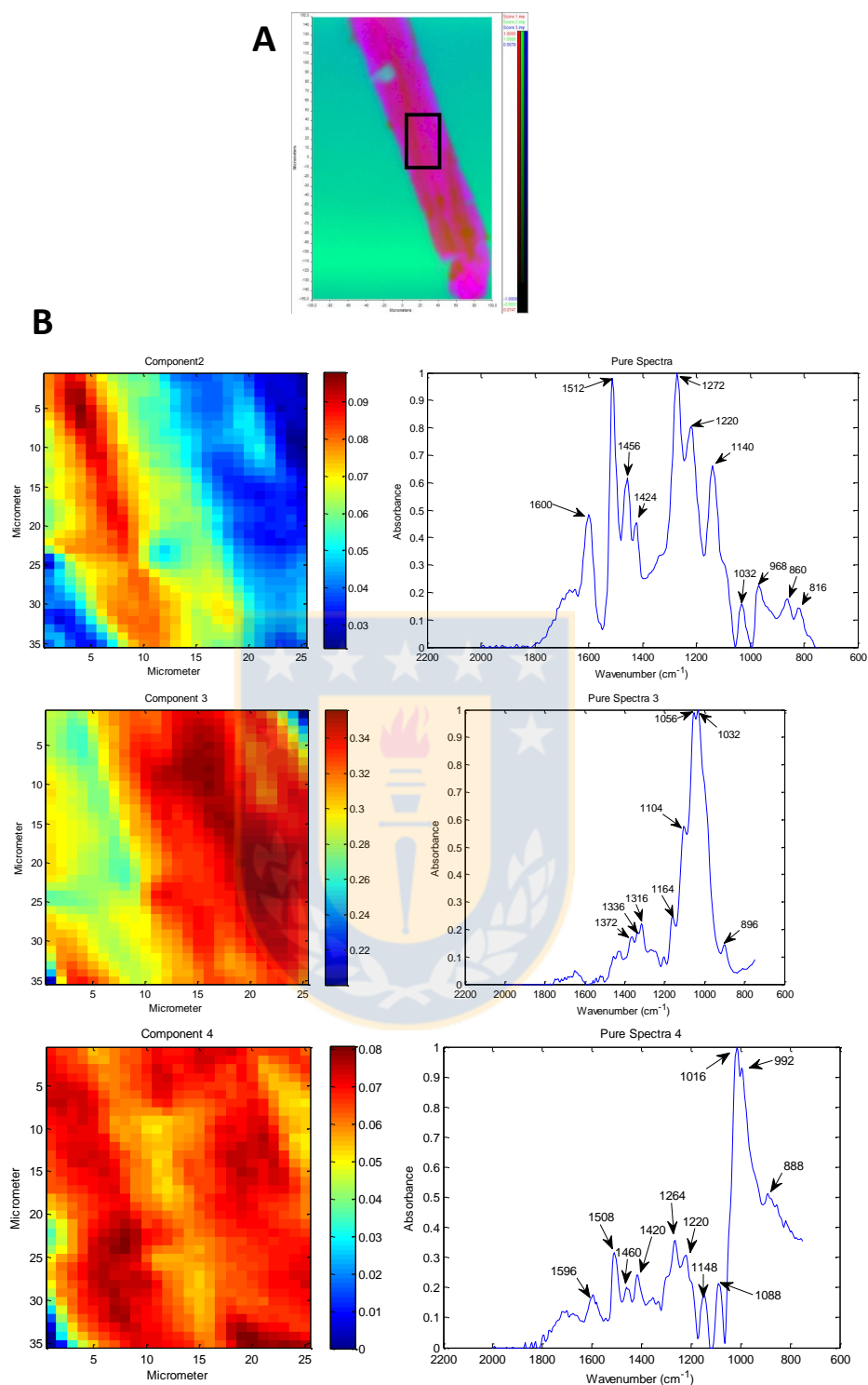


**Figura 3.7:** A, Imagen reconstruida por PCA de la fibra PR3. El marco negro indica una región representativa seleccionada para un análisis más detallado. B, muestran los scores y loading respectivamente del PCA de la zona aumentada. El loading del PC2 tiene gran similitud con el espectro de lignina, mientras que el loading del PC3 tiene bandas dominantes con valores positivos en las regiones asociadas a celulosa y bandas negativas en regiones asociadas a lignina. (Propiedad del autor).

MCR-ALS tiene como gran ventaja la entrega de los espectros puros presentes en la mezcla, en base a la reconstrucción matemática mediante la aplicación de algoritmos iterativos. Esta técnica quimiométrica permitiría, a diferencia de PCA, la visualización del espectro puro de cada componente, consiguiéndose describir incluso pequeñas variaciones espectrales. En la Figura 3.8 A se observa un acercamiento a una zona rica en lignina de la fibra PR2. En la Figura 3.8 B, la reconstrucción de componentes puros muestra que el componente 1 presenta un espectro muy similar al de la lignina (compare con Figura 3.3) y el componente 3 muestra un espectro muy similar a celulosa (compare con Figura 3.3). La similitud entre el espectro reconstruido y el espectro puro es casi completa. El componente 2 no es mostrado puesto que está asociado a zonas de mal contacto con el cristal ATR (fondo, en las esquinas de la imagen). El componente 4 presenta bandas asociadas en su mayoría a celulosa, pero también se observa la presencia de bandas asociadas a lignina (ver Figura 3.8 B). Esto sugiere que existiría otro componente, además de celulosa y lignina, que estaría compuesto por un tipo de asociación entre ambos. Esto tiene precedentes en bibliografía, donde se ha descrito que procesos de pretratamiento hidrotérmico de fibras gatilla ciclos de despolimerización y repolimerización de lignina, produciéndose interacciones lignina-carbohidratos covalentes de mecanismo desconocido [7].

La Figura 3.9, corresponde al análisis por MCR-ALS de la imagen de la fibra PR3. El componente 1 no se mostró ya que corresponde a zonas de mal contacto con el cristal en las esquinas de la imagen. Los componentes 2 y 3 muestran un espectro muy similar a lignina y celulosa, respectivamente, cuya distribución se contraponen en la imagen. El otro componente tiene características espectrales de ambos componentes, por lo que pudiese corresponder a un complejo lignina-carbohidrato. En esta fibra, la lignina se encuentra distribuida más homogéneamente en la superficie de la fibra, lo que pudiese tener una relación con el bajo rendimiento en la hidrólisis enzimática, debido a la menor accesibilidad de la enzima. Esto se condice con la Figura 3.2, donde se observa en la imagen de microscopía confocal que la fibra PR3 presenta menor cantidad de gotitas coalescentes en la superficie, por lo que la mayoría de la lignina estaría depositada de manera uniforme en la superficie de la fibra, actuando como un aislante.



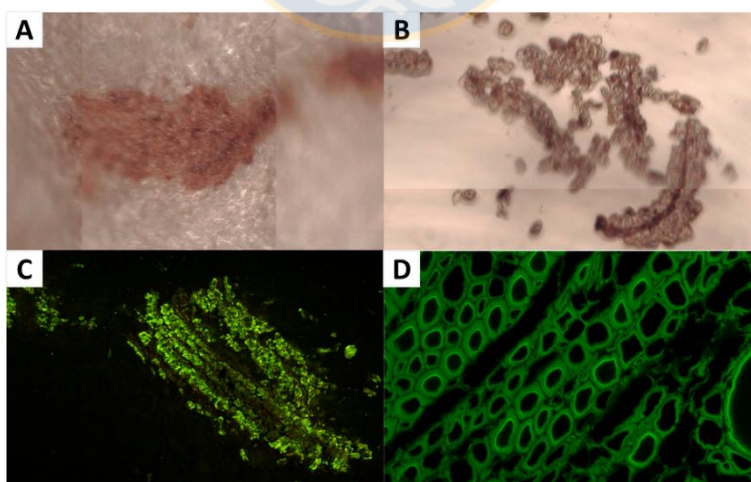


**Figura 3.9:** Análisis por MCR-ALS de la fibra PR2. A, Imagen reconstruida por PCA de la fibra PR3. El marco negro indica una región representativa de la fibra seleccionada para el análisis por MCR-ALS. B, Concentraciones representadas como mapa de calor y espectros puros de los componentes presentes en la fibra PR2, resueltos mediante MCR-ALS. Los componentes 2 y 4 tienen gran similitud con el espectro puro de lignina, mientras que el componente 3 tiene gran similitud con el espectro de celulosa. (Propiedad del autor).

## Sección 2

### Efecto de diferentes técnicas de pre-tratamiento de material lignocelulósico en la relocalización de lignina al interior de la pared celular

La Figura 3.10 muestra imágenes obtenidas por microscopía óptica y CLSM, de cortes transversales hechos con micrótopo en fibras embebidas en resina epoxi (A y C), y de cortes obtenidos mediante criostato (B y D). En ambos casos se utilizaron fibras provenientes de la misma pulpa, por lo que las diferencias observadas no son consecuencia de la severidad del pre-tratamiento. En los cortes hechos en resina (A y C) se pierde la integridad estructural de las células vegetales y no es posible identificar el lumen, la pared ni el contorno entre diferentes células. El uso de resina epoxi ha sido reportado en varios estudios de análisis microscópico de madera [8-10], por lo que las dificultades encontradas en su uso por parte del autor pudiesen ser explicadas en que el cuchillo de diamante del micrótopo no se encontraba adecuadamente afilado. Por otro lado, la resina tiene un espectro infrarrojo cuyas bandas se solapan con bandas importantes en la madera (datos no mostrados), por lo tanto el uso de la resina entorpecía el análisis por esta técnica. En contraste, en los cortes realizados por criostato (Figura 3.10 B y D) se observa que la integridad estructural ha sido preservada, siendo posible observar la nítidamente la pared celular. En consecuencia, el estudio se realizó usando esta última técnica de corte.



**Figura 3.10:** Cortes transversales de fibras pre-tratadas. A y C, imágenes en microscopía óptica y CLSM (respectivamente) de muestras embebidas en resina epoxi cortadas con micrótopo. B y D, imágenes en microscopía óptica y CLSM de muestras cortadas con criostato en medio de congelación. (Propiedad del autor).

A través de cortes transversales de fibras se estudió el efecto de tres pre-tratamientos a diferentes severidades en los patrones de distribución de lignina en la pared celular. Muestras de *P. radiata* fueron pre-tratadas por Organosolv, mientras que muestras de *E. globulus* fueron pre-tratadas por hidrólisis con ácido diluido y auto hidrólisis. Para cada pre-tratamiento se seleccionaron dos muestras de interés, una a alta y otra a baja severidad y/o que contrastaban en sus porcentajes de rendimiento en hidrólisis enzimática (ver Tabla 3.2).

**Tabla 3.2:** Composición química y % de rendimiento en hidrólisis enzimática (%R.H.E.) de muestras a alta y baja severidad pre-tratadas por autohidrólisis, hidrólisis con ácido diluido y Organosolv. (Propiedad del autor).

Pre-tratamiento	Muestras	*Condiciones	Severidad (So)	(g/100g de pulpa seca)				%R.H.E
				Lignina	Glucanos	Hemicelulosas	Acetilos	
Autohidrólisis	AU-1	190°C/70min	5.89	36.1	61.1	1.5	0.1	75.9
	AU-5	160°C/15min	4.4	27.8	50.2	13.7	2.7	13.5
Hidrólisis con ácido diluido	HAD-5	120°C/0.1%/1:4/20min	3.33	33.2	47.1	35.7	4	6.2
	HAD-11	177°C/0.43%/1:6/20min	5.03	24.5	58.8	4.2	0.5	80.0
Organosolv	PR2	0.7%/50:50/190°C/30min	5.46	21.7	80.5	-	-	92.6
	PR3	0.35%/50:50/190°C/30min	5.45	23.7	73.5	-	-	43.6

\*Autohidrólisis: T°/tiempo de reacción

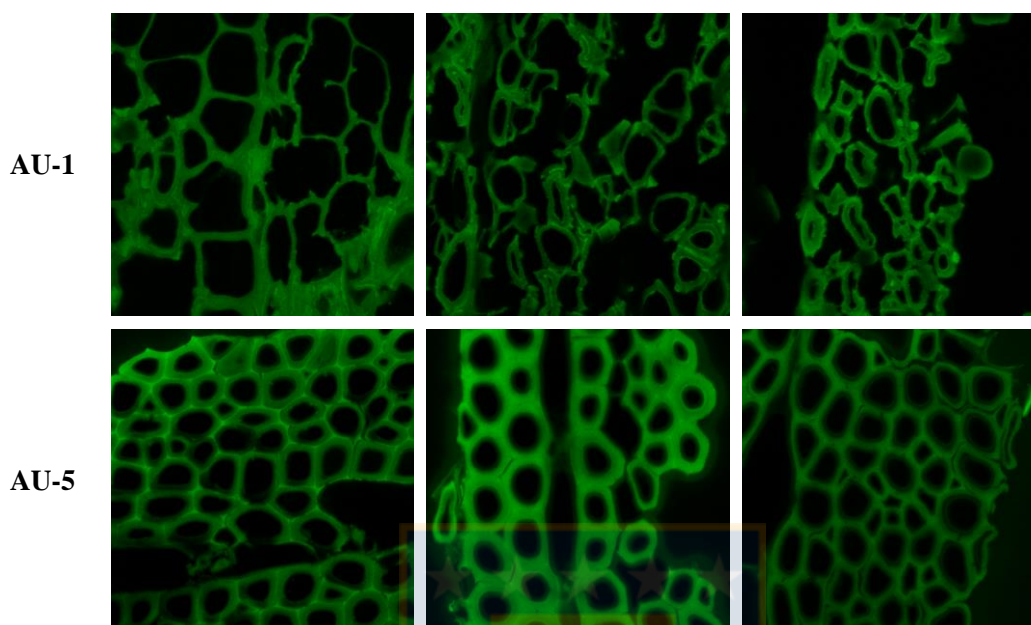
Hidrólisis ácida diluida: T°/% H<sub>2</sub>SO<sub>4</sub>/razón madera:licor/tiempo

Organosolv: % H<sub>2</sub>SO<sub>4</sub>/%etanol:agua/T°/tiempo de reacción

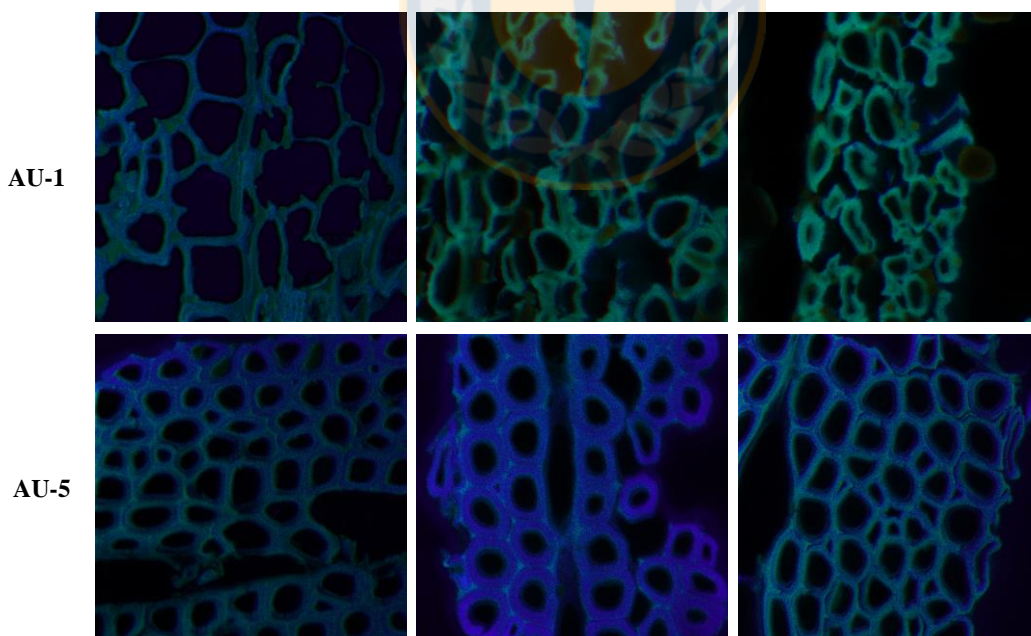
### Efecto del pre-tratamiento por autohidrólisis

La Figura 3.11 muestra imágenes obtenidas por CLSM de tres regiones diferentes de la muestra AU-1, de alto rendimiento en hidrólisis enzimática, y de la muestra AU-5, de bajo rendimiento. Se puede observar que la muestra de alto rendimiento tiene en general una estructura más colapsada y desorganizada comparada con la fibra de bajo rendimiento. Esto es coherente, puesto que la muestra AU-1 fue sometida a mayor temperatura y por una mayor cantidad de tiempo que la muestra AU-5 (ver Tabla 3.2). El colapso de la pared celular puede

explicar el mayor rendimiento en hidrólisis enzimática, puesto que la celulosa se encontraría más expuesta.



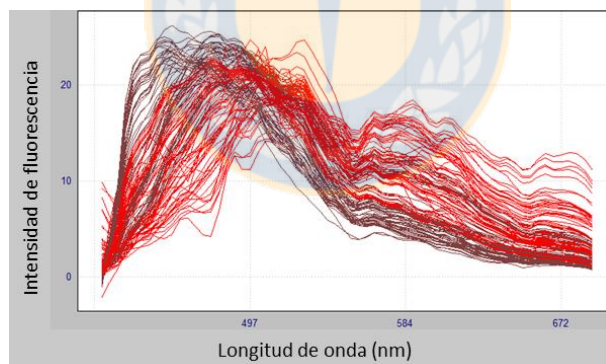
**Figura 3.11:** Imágenes obtenidas por CLSM de fibras provenientes de pulpas pre-tratadas por autohidrólisis. La muestra AU-1 fue sometida a condiciones más severas de pre-tratamiento y produjo mayores rendimientos de glucosa en hidrólisis enzimática que AU-5. (Propiedad del autor).



**Figura 3.12:** Imágenes obtenidas por CLSM espectral, obteniendo el espectro de emisión de fluorescencia, de fibras provenientes de pulpas pre-tratadas por autohidrólisis. La muestra AU-1 fue sometida a condiciones más severas de pre-tratamiento y produjo mayores rendimientos de glucosa en hidrólisis enzimática que AU-5. (Propiedad del autor).



La Figura 3.12 muestra las mismas imágenes de la Figura 3.11 pero esta vez obteniendo el espectro de emisión de fluorescencia, mediante CLSM espectral. Para esto, no se usaron fluoróforos adicionales a los presentes en la muestra, por lo que las señales provienen exclusivamente de auto-fluorescencia. El color de la imagen espectral corresponde a la superposición de todos los canales (longitudes de onda) de fluorescencia en que fue obtenido el espectro, por lo que diferencias en la coloración de distintas zonas de la muestra corresponderían a diferencias en el espectro de emisión de fluorescencia. En la muestra AU-1 (Figura 3.12) se observan zonas de coloración verde oscuro, que debido a su distribución y ubicación pudiesen ser atribuidas a lignina y zonas de coloración verde claro/azul, las que podrían corresponder a celulosa, y cuya separación se encuentra bien definida. En la muestra AU-5 también se observan 2 regiones de diferente color (calipso para zonas atribuidas a lignina, azul para zonas atribuidas a celulosa), aunque en este caso la distribución es más homogénea que en AU-1. En AU-5, las zonas de diferente color coinciden más claramente con estructuras celulares de la fibra. El color calipso (lignina) se encuentra dispuesto en la zona entre células, llamada lamela media, mientras que el color azul se encuentra ubicado en la pared celular (ver imagen central).

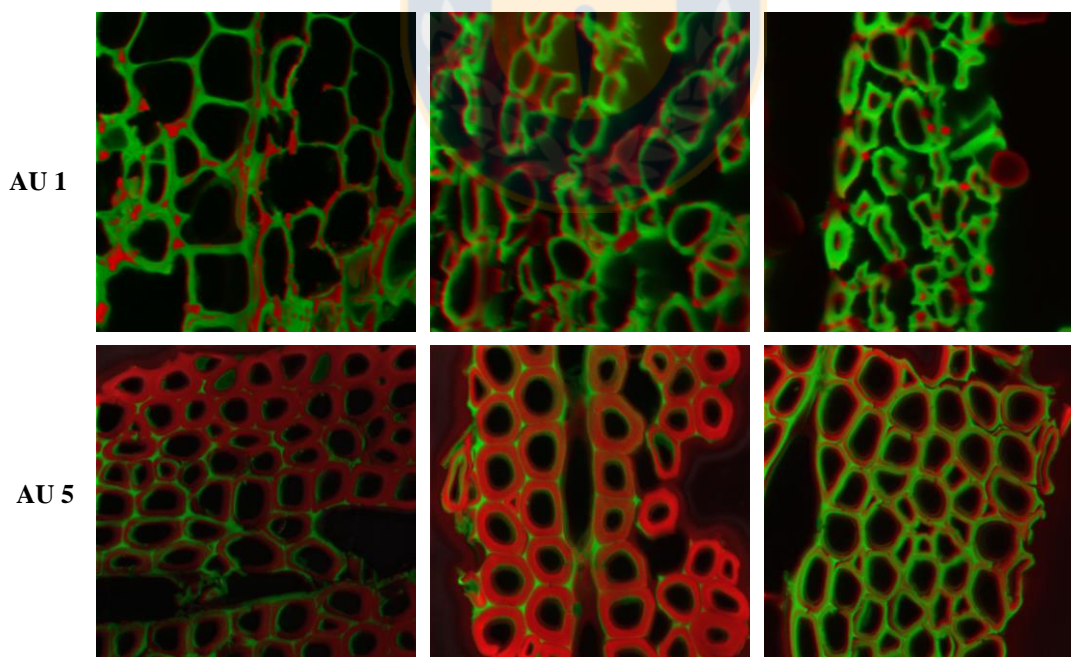


**Figura 3.13:** Espectros de emisión de fluorescencia reconstruidos por PCA (8 componentes) obtenidos de dos regiones de coloración diferente en imágenes obtenidas por CLSM espectral de las muestras AU-1 y AU-5. (Propiedad del autor).

La Figura 3.13 muestra espectros reconstruidos después de un PCA con 8 componentes principales, obtenidos desde puntos de la imagen seleccionados en base a su coloración. Los espectros en rojo provienen de las zonas atribuidas a lignina, mientras que los espectros en café provienen de las zonas atribuidas a celulosa. Se puede observar que los espectros en rojo

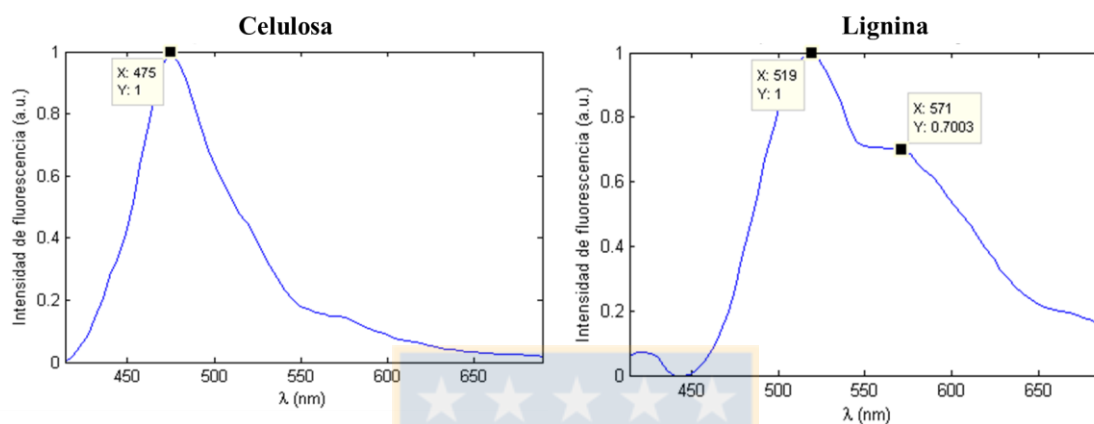
se encuentran más desplazados hacia la derecha que los espectros en café, y a diferencia de éstos tienen un segundo pico de menor intensidad cercano a 584 nm.

El software ZEN® (Carl Zeiss, Alemania) que controla el microscopio confocal laser espectral utilizado tiene la opción de realizar un desmezclado espectral lineal (en inglés, *linear unmixing*) que permite separar dos señales fluorescentes. *Linear unmixing* se basa en la suposición de que la señal total detectada para cada canal de fluorescencia (a una longitud de onda determinada) puede ser expresada como una combinación lineal de cada fluoróforo contribuyente [11, 12]. Usando el software del equipo, se realizó el *linear unmixing* proporcionando como referencia espectros representativos obtenidos desde las dos zonas de coloración diferente ya descritas. En la Figura 3.14 se observa la reconstrucción de la imagen mostrada en la Figura 3.12 realizada por el software ZEN® mediante *linear unmixing*. Para AU-1, las zonas en rojo corresponden a las zonas atribuidas a lignina de la imagen original (Figura 3.12), mientras que para AU-5 las zonas atribuidas a lignina se encuentran representadas en verde.



**Figura 3.14:** Imágenes obtenidas por CLSM espectral (emisión de fluorescencia) y reconstruidas mediante desmezclado lineal, de las muestras AU-1 y AU-5 pre-tratadas por autohidrólisis a distintas severidades. (Propiedad del autor).

A fin de comprobar la identidad de los fluoróforos responsables de la discriminación por color de la imagen fluorescente espectral, se tomaron espectros de emisión de fluorescencia de patrones de celulosa micro-cristalina y de lignina Kraft comercial (ver Figura 3.15). Los espectros coinciden con los mostrados en la Figura 3.13, por lo que lignina y celulosa serían los responsables de la auto-fluorescencia.

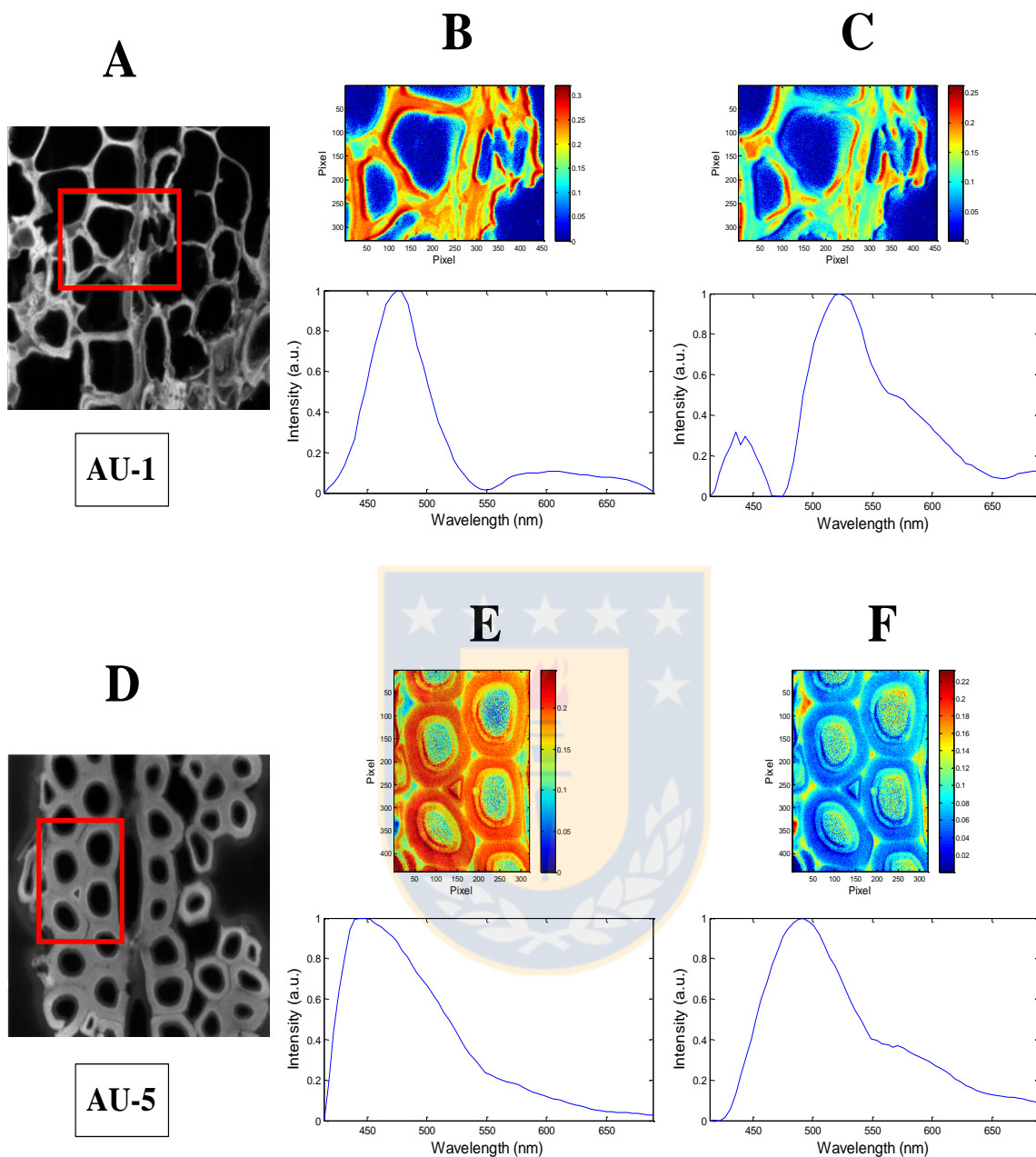


**Figura 3.15:** Espectros de emisión de fluorescencia de celulosa y lignina. (Propiedad del autor).

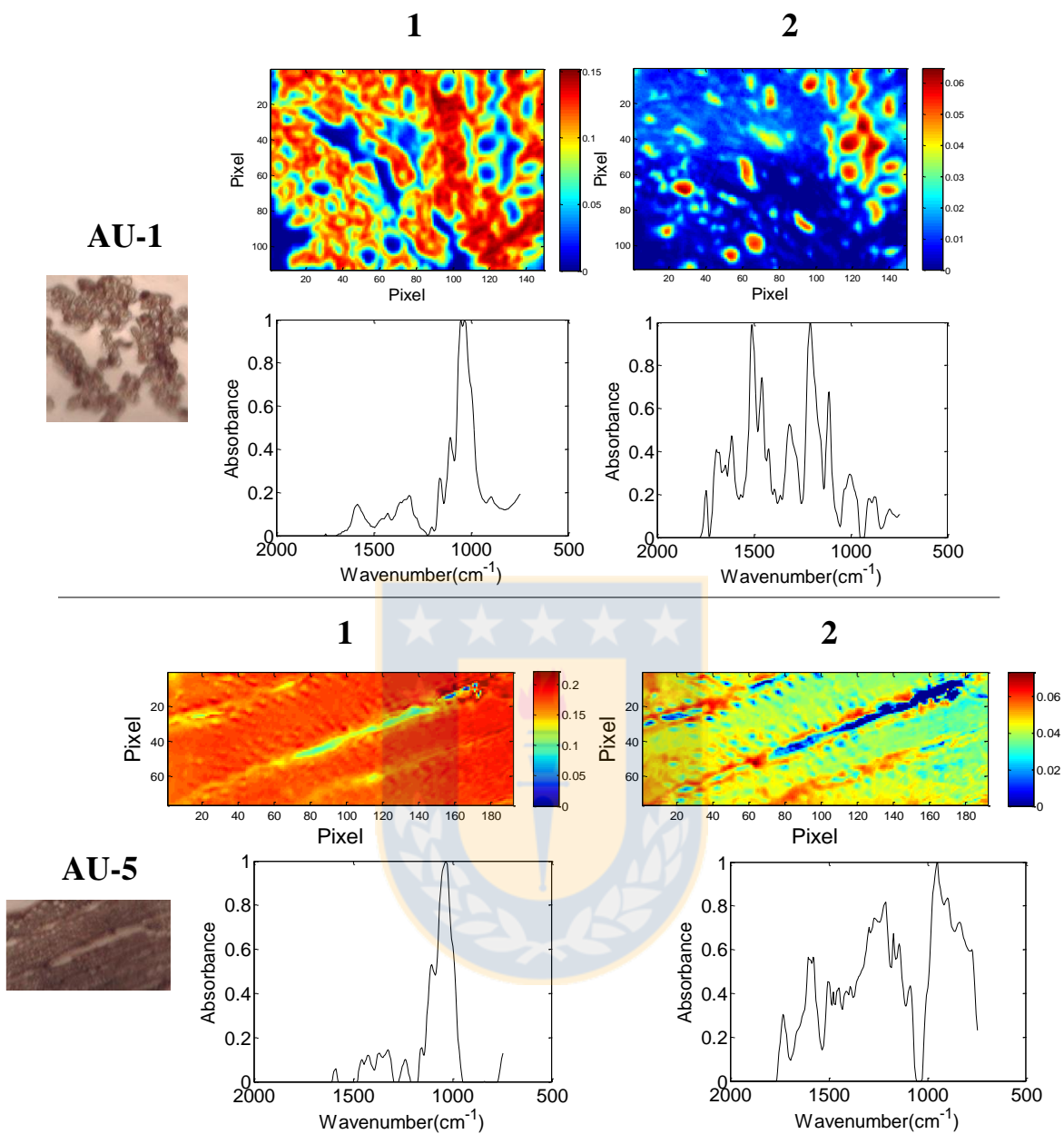
Como se mencionó en el capítulo 1 del presente trabajo, la auto-fluorescencia de lignina es un fenómeno muy estudiado[13], sin embargo, las propiedades auto-fluorescentes de la celulosa son un fenómeno que ha generado debate, aunque diferentes autores han descrito su existencia en el pasado [14-16]. El espectro de emisión de fluorescencia mostrado en la Figura 3.15 demuestra que la celulosa es efectivamente responsable de una fracción de la auto-fluorescencia total de la imagen obtenida por CLSM, por lo que proporcionaría mayor información química al análisis espectral por esta técnica, que clásicamente es interpretada sólo en función de la auto-fluorescencia de lignina. Habiendo comprobado la identidad de los fluoróforos presentes en la imagen, es posible hacer un análisis más detallado de las Figuras 3.12 y 3.14. Como resultado del pre-tratamiento por autohidrólisis, la lignina que originalmente se encuentra en la lamela media (ver muestra AU-5 en Figura 3.14) es desplazada hacia el exterior de la pared celular, ya sea en forma de gotitas (ver imagen superior derecha de muestra AU-1 en Figura 3.14) o bien en forma de conglomerados amorfos (ver imagen superior izquierda y central de muestra AU-1 en Figura 3.14).

La Figura 3.16 muestra un acercamiento a la zona de la pared celular de una imagen obtenida por CLSM espectral de las muestras AU-1 y AU-5, las que fueron analizadas por MCR-ALS a fin de resolver el espectro puro y la concentración relativa en cada píxel de la imagen. Este análisis se realizó para corroborar los resultados obtenidos por *linear unmixing*, los que por haber sido realizados mediante un software comercial de código oculto, no entregaban la confianza suficiente al autor. Por otro lado, a diferencia del análisis por *linear unmixing*, el análisis por MCR-ALS no requiere necesariamente de un espectro de referencia proporcionado por el operador, y es capaz de resolver los espectros de cada componente presente en la imagen, los que al ser comparados con los obtenidos desde patrones (ver Figura 3.15) proporcionarían una especie de validación a esta aproximación experimental. Los resultados muestran que los espectros resueltos por MCR-ALS y la distribución de cada componente en el mapa de concentración reconstruido (ver Figura 3.16) coinciden con los espectros de emisión de fluorescencia obtenidos para los patrones de celulosa y lignina (ver Figura 3.15) y con la distribución obtenida para celulosa y lignina de la Figura 3.14. En la Figura 3.16 también es posible observar que la muestra AU-1 se encuentra más colapsada que AU-5 y que la lignina de la lamela media ha comenzado a aglomerarse y salir hacia el exterior de la pared celular (B y C), mientras que en AU-5 la lignina se encuentra concentrada principalmente en la lamela media (F).

Los cortes obtenidos para las muestras AU-1 y AU-5 fueron también analizados por microscopía FT-IR y las imágenes obtenidas fueron resueltas por MCR-ALS. Si bien, la microscopía FT-IR tiene una resolución espacial mucho menor que la obtenida por CLSM (tamaño de píxel de 1,56  $\mu\text{m}$  en microscopía FT-IR, versus 88 nm en CLSM), la información química contenida en los espectros es de mucho mayor selectividad. El algoritmo SVD ayuda a estimar la cantidad de componentes químicos presentes en una muestra en base a la cantidad de varianza explicada por cada uno, sin embargo, la presencia en la muestra de medio de congelación y zonas de no contacto con el cristal ATR (fondo) hizo necesario seleccionar una cantidad mayor de componentes para encontrar los compuestos de interés. Generalmente los primeros cuatro componentes resueltos correspondían a celulosa, lignina, medio de congelación y fondo (en cualquier orden). La información de los componentes correspondientes al medio de congelación y el fondo no es mostrada por no ser relevante.



**Figura 3.16:** Análisis por MCR-ALS de imágenes obtenidas por CLSM espectral (emisión de fluorescencia) de fibras pre-tratadas por autohidrólisis. A y D, zonas de acercamiento de imagen de muestra AU-1 y AU-5 respectivamente. B y C, imagen reconstruida y espectro puro resuelto de celulosa y lignina, respectivamente, para la muestra AU-1. E y F, imagen reconstruida y espectro puro resuelto de celulosa y lignina, respectivamente, para la muestra AU-5. (Propiedad del autor).



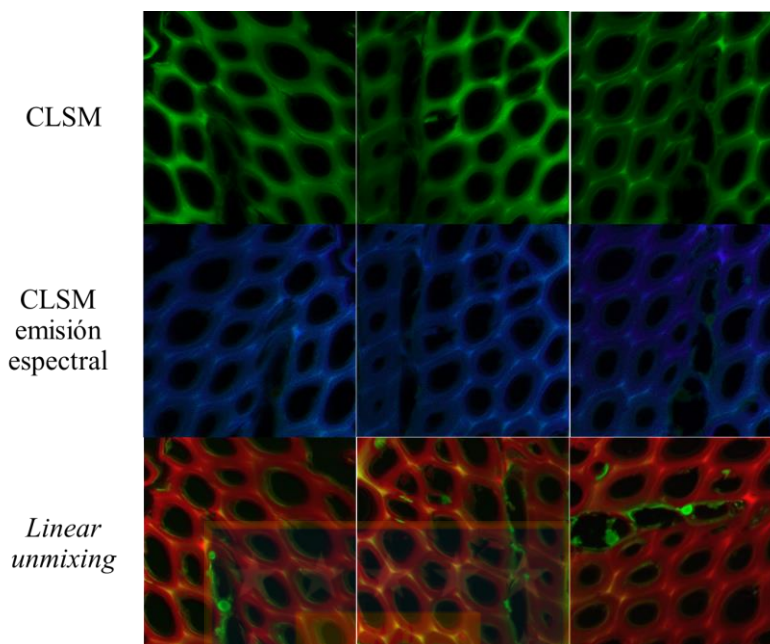
**Figura 3.17:** Análisis por MCR-ALS (mapa de concentración relativa y espectro puro) de imágenes obtenidas por microscopía FT-IR de cortes transversales de las muestras AU-1 y AU-5. En AU-1 el componente 1 correspondería a celulosa y el componente 2 a lignina. En AU-5 el componente 1 correspondería a celulosa y el componente 2 a un complejo lignina-carbohidrato. (Propiedad del autor).

Los resultados mostrados en la Figura 3.17 son coherentes con los resultados de los análisis por CLSM. El componente 1 de la muestra AU-1 tiene bandas con intensidades típicas de celulosa en  $1032$  y  $1055\text{ cm}^{-1}$  (estiramiento C-O),  $1108\text{ cm}^{-1}$  (vibración asimétrica anillo glucosa) y  $1160\text{ cm}^{-1}$  (estiramiento asimétrico C-O-C); y el componente 2 tiene bandas típicas de lignina a  $1212\text{ cm}^{-1}$  (C-O grupo G),  $1508$  y  $1610\text{ cm}^{-1}$  (vibración del esqueleto aromático) y  $1468\text{ cm}^{-1}$  (deformación C-H de metilo y metileno). La distribución de celulosa concuerda con la estructura de la pared celular, mientras que la lignina se encuentra aglomerada en esferas o cúmulos, del mismo modo descrito en la Figura 3.14. En la muestra AU-5 el primer componente muestra bandas típicas de celulosa ( $1032$ ,  $1108$ ,  $1160\text{ cm}^{-1}$ ) y el segundo componente tiene características espectrales de lignina (bandas a  $1504$ ,  $1604$ ,  $1212\text{ cm}^{-1}$ ), además de presentar una banda a  $1732\text{ cm}^{-1}$ , relacionada con una vibración de C=O de hemicelulosa, por lo que correspondería a un complejo lignina-carbohidrato. La distribución de este último componente es relativamente homogénea en toda la muestra. Estas dos características son lógicas si se considera que la muestra AU-5 fue pre-tratada bajo condiciones menos severas que AU-1, por lo que el pre-tratamiento no fue capaz de romper las uniones lignina-carbohidrato para conseguir la relocalización de lignina.

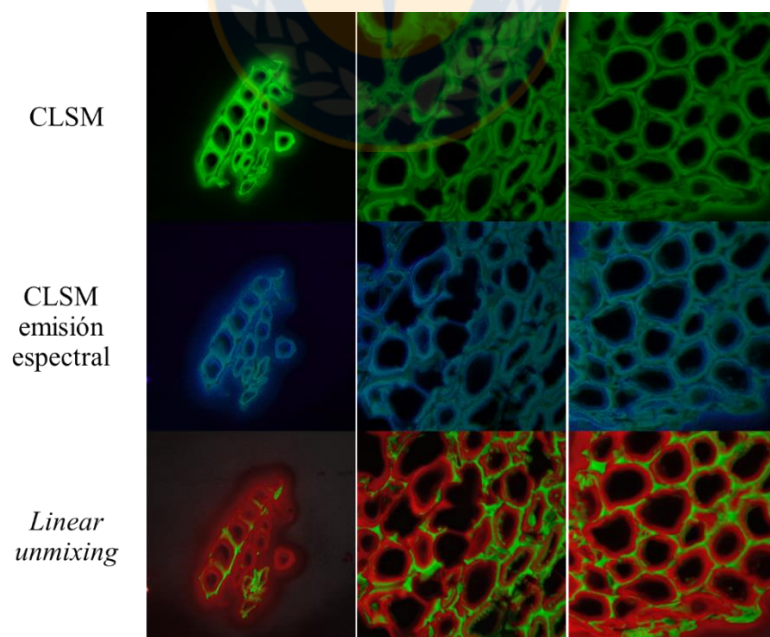
### *Efectos del pre-tratamiento por hidrólisis con ácido diluido*

Los análisis usados para describir el efecto del pre-tratamiento por hidrólisis con ácido diluido fueron los mismos que los usados para autohidrólisis. La muestra HAD-5 corresponde a una muestra de baja severidad y bajo rendimiento de glucosa en hidrólisis enzimática y la muestra HAD-11 corresponde a una muestra de mayor severidad y alto rendimiento de hidrólisis enzimática. La Figura 3.18 muestra tres imágenes diferentes de la muestra HAD-5 obtenidas por CLSM (visualizadas por fluorescencia total, emisión espectral de fluorescencia y *linear unmixing*). La estructura de la pared celular se encuentra íntegra debido a la baja severidad del tratamiento y se puede observar la aparición de algunas gotas de lignina en zonas cercanas a la lamela media. La muestra HAD-11 en cambio, muestra marcadas diferencias morfológicas (ver Figura 3.19). Producto del pre-tratamiento la estructura celular se encuentra más colapsada, la lignina parece haber sido retirada de la pared, migrando hacia la zona de la lamela media, lo que parece haber afectado la integridad de la pared. A diferencia del mecanismo observado en la autohidrólisis, la lignina no parece

migrar ni coalescer hacia el exterior de la fibra, sino que su retirada hacia la lamela media dejaría disponible la celulosa de la pared celular para el ataque enzimático.



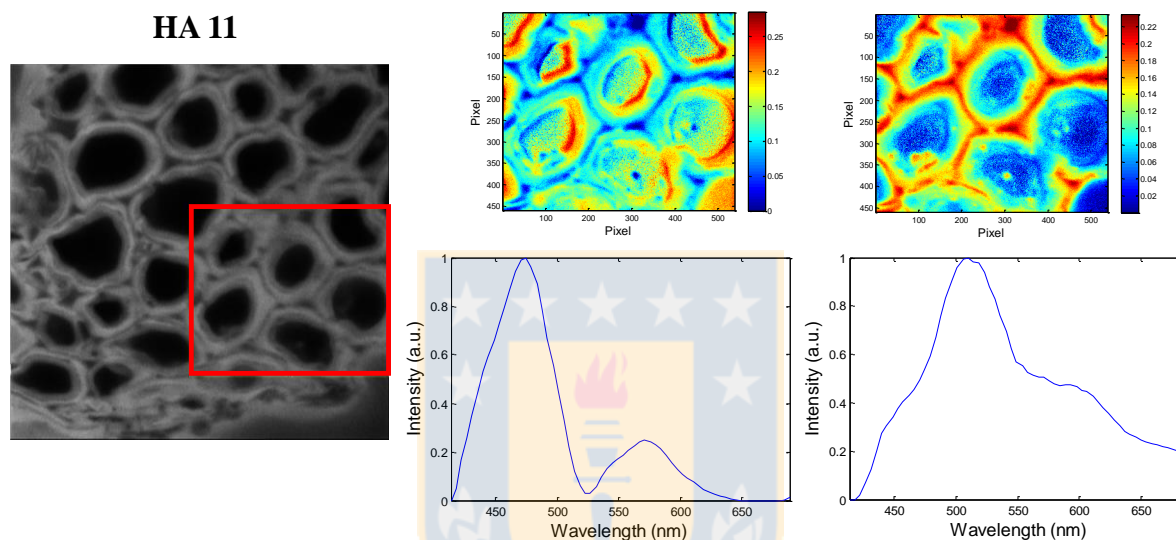
**Figura 3.18:** Imágenes obtenidas de la muestra HAD-5 de bajo rendimiento en hidrólisis enzimática, tratada por hidrólisis ácida diluida a baja severidad. (Propiedad del autor).



**Figura 3.19:** Imágenes obtenidas de la muestra HAD-11 de alto rendimiento en hidrólisis enzimática, tratada por hidrólisis ácida diluida a alta severidad. (Propiedad del autor).

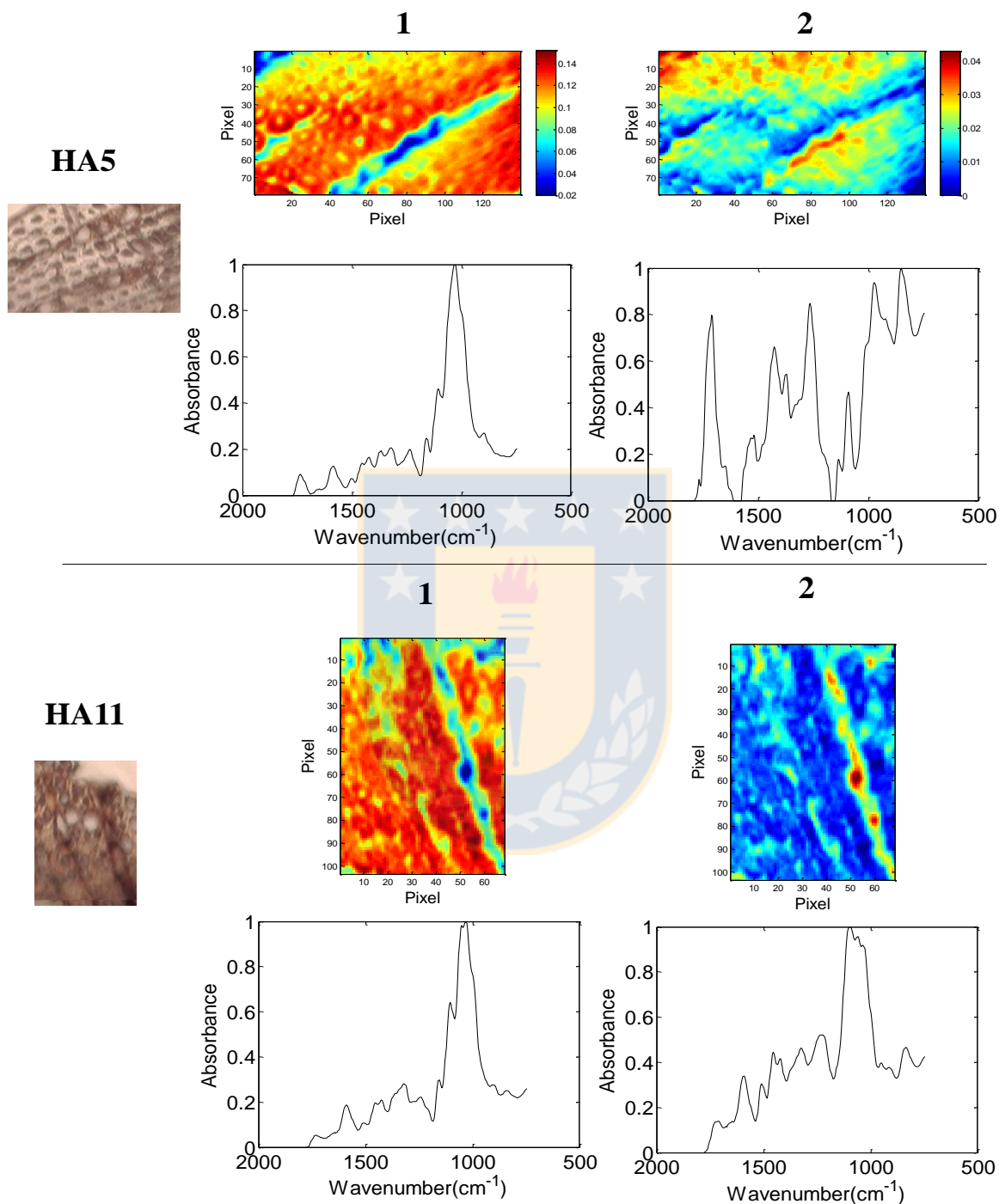


La Figura 3.20 muestra un acercamiento a la pared celular de la imagen obtenida por fluorescencia espectral de la muestra HAD-11. En esta imagen es posible observar en mayor detalle que la lignina se ha desplazado casi totalmente desde la pared secundaria de la célula, desplazándose hacia la lamela media. La celulosa por otro lado se observa más libre en la zona de pared secundaria.



**Figura 3.20:** Análisis por MCR-ALS de imágenes obtenidas por CLSM espectral (emisión de fluorescencia) de la muestra HAD-11. Los mapas de concentración y espectros puros de emisión resueltos corresponden a celulosa (izquierda) y lignina (derecha). (Propiedad del autor).

El análisis por microscopía FT-IR muestra un comportamiento semejante (ver Figura 3.21). En HAD-5, el componente 1 correspondería a celulosa (bandas en 1032, 1108 y 1160  $\text{cm}^{-1}$ ) y el componente 2 tiene bandas asociadas a lignina en 1264  $\text{cm}^{-1}$  (C-O grupo G), 1428  $\text{cm}^{-1}$  (deformación asimétrica de C-H) y 1712  $\text{cm}^{-1}$  (estiramiento C=O en lignina). Lignina se encuentra más homogéneamente distribuida en la fibra a lo largo de toda la sección transversal. En HAD-11 el componente 1 es celulosa (bandas en 1032, 1052, 1104 y 1156  $\text{cm}^{-1}$ ) y el componente 2 es lignina (bandas a 1510 y 1592  $\text{cm}^{-1}$ ). En esta muestra la lignina se encuentra concentrada en la región de la lamela media, por lo que los resultados serían coherentes entre ambas técnicas de microscopía espectral usadas.



**Figura 3.21:** Análisis por MCR-ALS (mapa de concentración relativa y espectro puro) de imágenes obtenidas por microscopía FT-IR de cortes transversales de las muestras HAD-5 y HAD-11. En ambos casos el componente 1 corresponde a celulosa y el componente 2 corresponde a lignina. (Propiedad del autor).

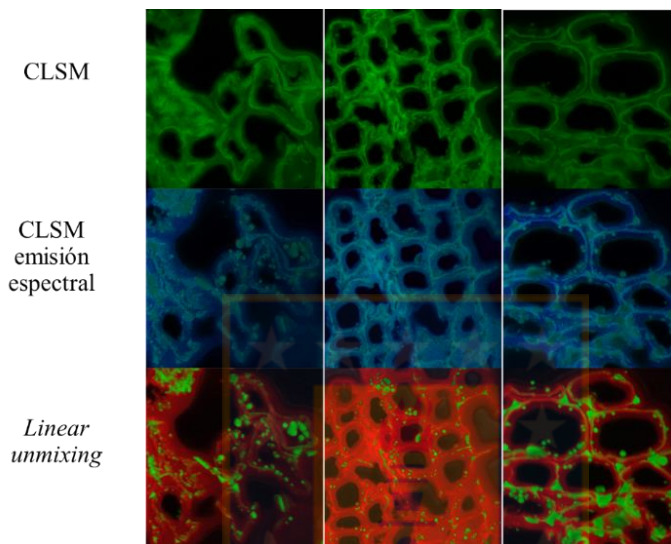
### Efectos del pre-tratamiento por Organosolv

Las muestras analizadas para estudiar el efecto de Organosolv en la distribución de lignina corresponden a las mismas pulpas usadas en la sección 1, donde se realizaron análisis a nivel superficial. En este caso, el análisis entrega información de lo que ocurre al interior de la fibra. La Figura 3.22 muestra el interior de fibras de la muestra PR2, de mayor rendimiento enzimático. Mediante este análisis se corroboró lo observado en la sección 1. La lignina es relocalizada formando conglomerados esféricos y amorfos de tamaño variable, los que se localizan tanto al interior como al exterior de la pared celular, ya sea en el lumen celular como en la lamela media y zona externa de la fibra (ver Figura 3.22, Figura 3.24 y Figura 3.25). La estructura de la pared se encuentra consecuentemente debilitada y colapsada. Todas estas características ayudarían a explicar el alto rendimiento de hidrólisis enzimática obtenido en esta muestra, la cual es mayor que todas las otras muestras estudiadas (ver Tabla 3.2).

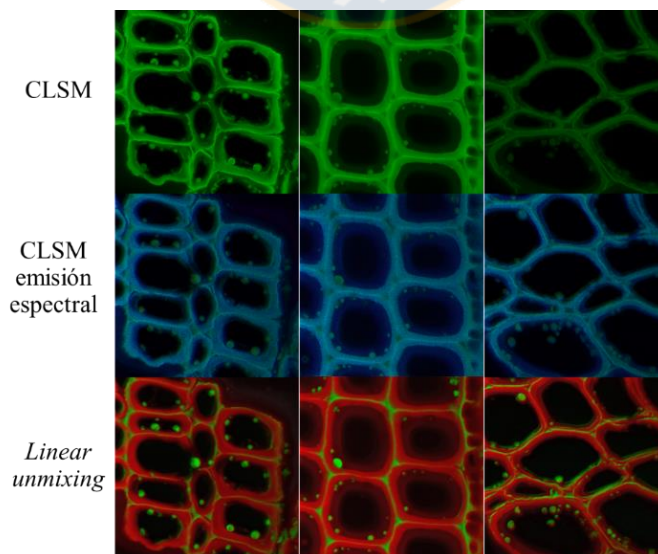
La Figura 3.23 muestra imágenes de la muestra PR3, la cual mostró menor rendimiento en hidrólisis enzimática. La muestra PR3 contiene la misma cantidad global de glucanos y lignina que la muestra PR2 (ver Tabla 3.1), sin embargo fue pre-tratada con la mitad de concentración de ácido sulfúrico (ver Tabla 3.2) lo que explicaría las diferencias morfológicas entre ambas fibras. Por ejemplo, la pared celular de PR3 no ha sufrido cambios estructurales tan notables como PR2, salvo por la presencia de varias esferas de lignina, las que son de menor tamaño que las vistas para la muestra PR2. La presencia de estas esferas en PR3 parece ser más numerosa hacia el lumen celular y no se observan hacia el exterior de las fibras o dentro de la pared celular (ver Figura 3.23, Figura 3.24).

Las imágenes por FT-IR (Figura 3.25) corroboran lo visto por CLSM. El componente 1 de PR2 y PR3 corresponde a celulosa y muestra las bandas típicas ya descritas en las otras muestras. La lignina (componente 2) en PR2 muestra bandas típicas en 1268, 1460, 1512 y 1592  $\text{cm}^{-1}$ , su distribución se ve segmentada, coherente con la presencia de esferas y aglomerados. El componente 2 de PR3 muestra bandas asociadas a lignina (1264, 1216, 1136 [deformación en el plano de CH de unidad G] y 1508  $\text{cm}^{-1}$ ) y también muestra intensidades puntuales, aunque se ubica más uniformemente en la lamela media.

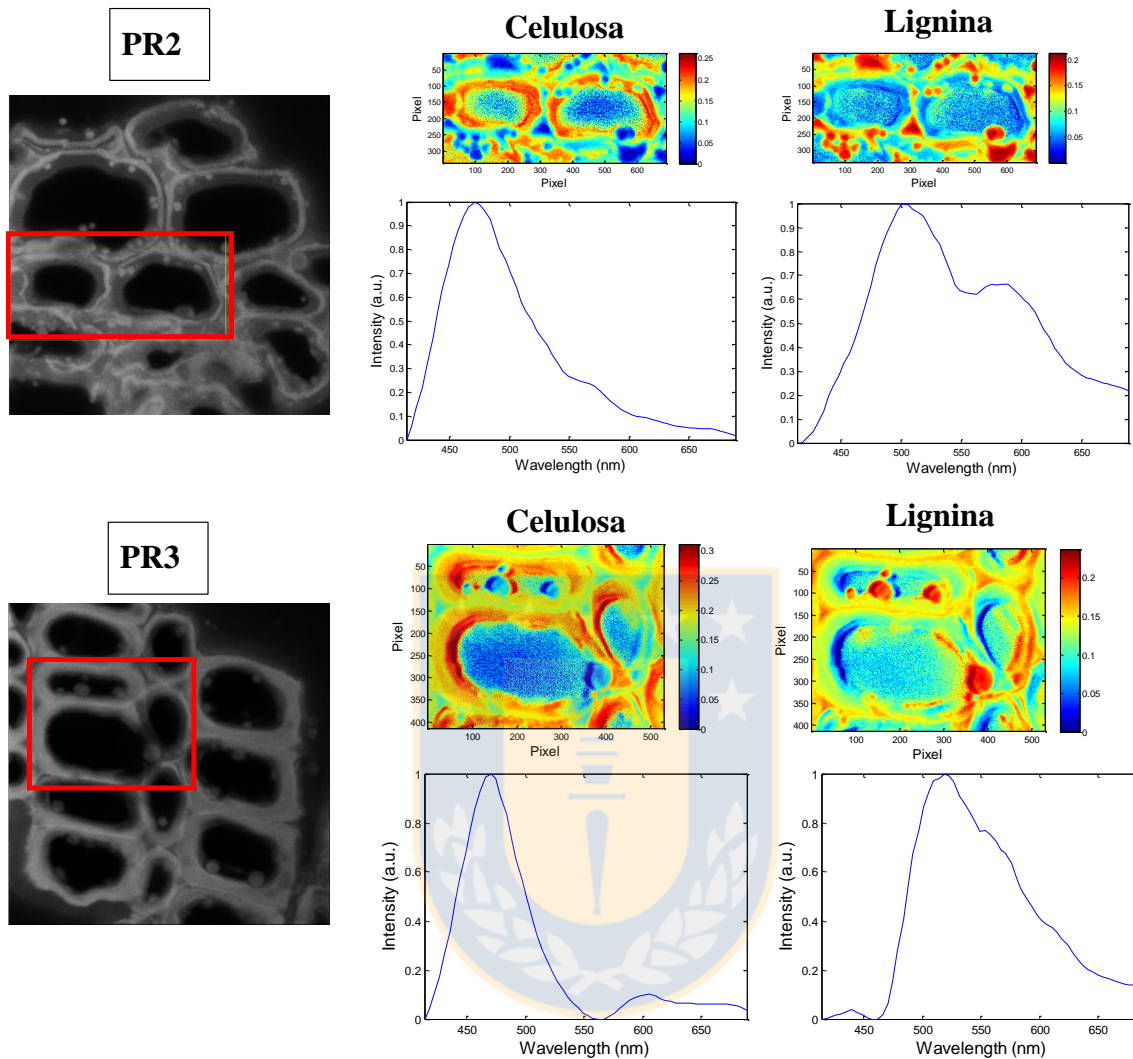
Estas características observadas en el corte transversal de la fibra son coherentes con las observaciones hechas en la superficie de la fibra en la sección 1 de este capítulo. Así, se puede asumir que el efecto principal del pre-tratamiento por Organosolv es la relocalización de lignina en esferas de diversos tamaños, lo que dejaría libre la celulosa de la pared celular para el ataque enzimático. Además, a mayor número y tamaño de estas esferas, mayor sería el rendimiento de hidrólisis enzimática.



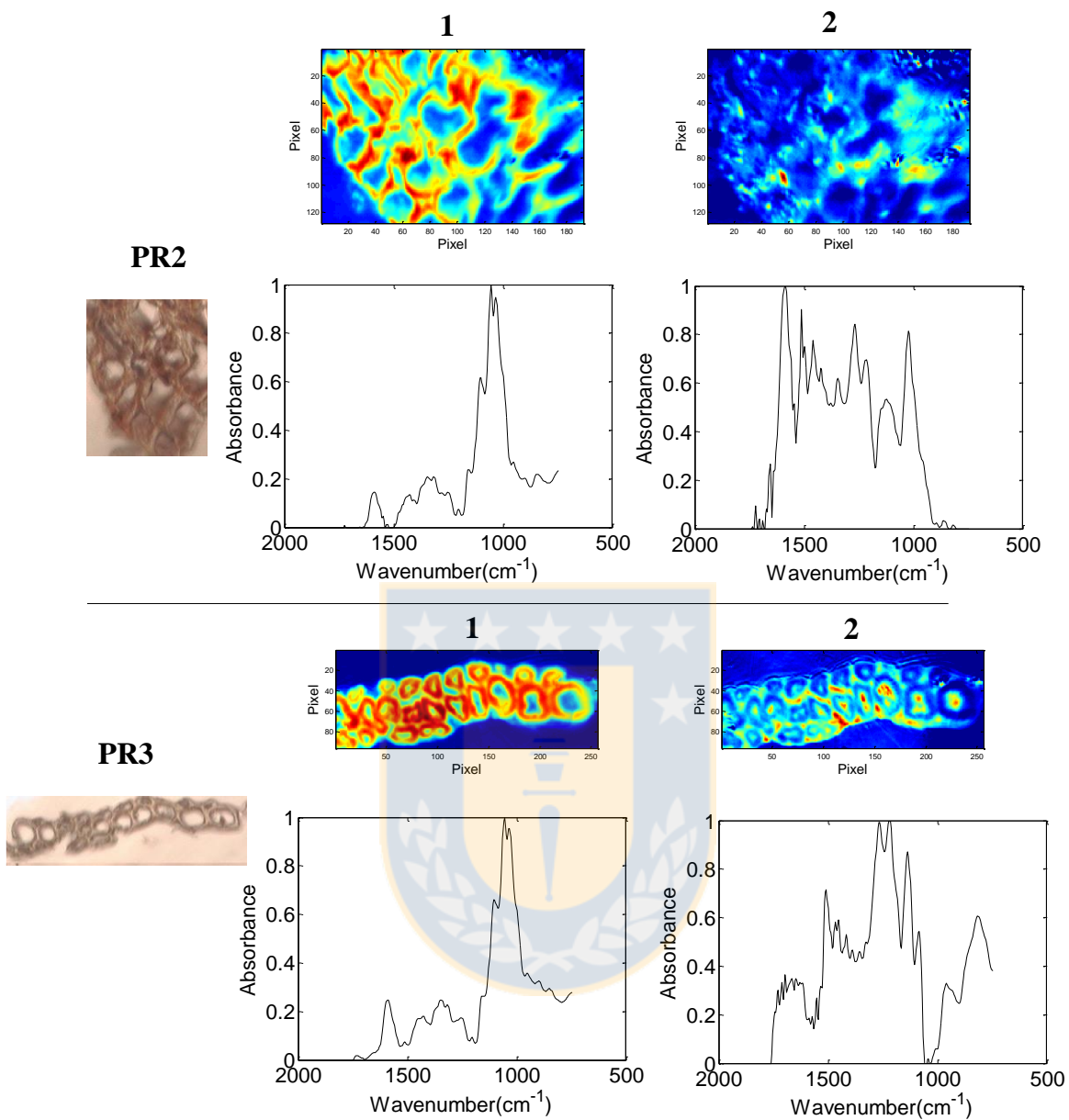
**Figura 3.22:** Imágenes obtenidas de la muestra PR2 de alto rendimiento en hidrólisis enzimática, tratada por Organosolv. (Propiedad del autor).



**Figura 3.23:** Imágenes obtenidas de la muestra PR3 de bajo rendimiento en hidrólisis enzimática, tratada por Organosolv. (Propiedad del autor).



**Figura 3.24:** Distribución y espectros de emisión puros de celulosa y lignina obtenidos mediante MCR-ALS de imágenes obtenidas por CLSM espectral (emisión de fluorescencia) de fibras pre-tratadas por Organosolv. (Propiedad del autor).



**Figura 3.25:** Análisis por MCR-ALS (mapa de concentración relativa y espectro puro) de imágenes obtenidas por microscopía FT-IR de cortes transversales de las muestras PR2 y PR3. En ambos casos el componente 1 corresponde a celulosa y el componente 2 corresponde a lignina. (Propiedad del autor).

## Referencias

1. Koo, B.-W., et al., *Structural changes in lignin during organosolv pretreatment of Liriodendron tulipifera and the effect on enzymatic hydrolysis*. Biomass and Bioenergy, 2012. **42**: p. 24-32.
2. Donohoe, B.S., et al., *Visualizing lignin coalescence and migration through maize cell walls following thermochemical pretreatment*. Biotechnology and Bioengineering, 2008. **101**(5): p. 913-925.
3. Li, H., et al., *Investigation of lignin deposition on cellulose during hydrothermal pretreatment, its effect on cellulose hydrolysis, and underlying mechanisms*. Biotechnology and bioengineering, 2014. **111**(3): p. 485-492.
4. Xu, F., et al., *Anatomy, ultrastructure and lignin distribution in cell wall of Caragana korshinskii*. Industrial crops and products, 2006. **24**(2): p. 186-193.
5. Zhang, Z., et al., *Anatomy, cell wall ultrastructure and inhomogeneity in lignin distribution of broussonetia papyrifera*. Cellulose Chemistry and Technology, 2012. **46**(3-4): p. 157-164.
6. Sannigrahi, P., et al., *Pseudo-lignin and pretreatment chemistry*. Energy & Environmental Science, 2011. **4**(4): p. 1306-1310.
7. Trajano, H.L., et al., *The fate of lignin during hydrothermal pretreatment*. Biotechnology for biofuels, 2013. **6**(1): p. 110.
8. Nicole Labbe, T.G.R. and S.S. Kelley, *FTIR imaging of wood and wood composites. Characterization of the Cellulosic Cell Wall*, 2008: p. 110.
9. Kitin, P., et al., *Analysis by confocal microscopy of the structure of cambium in the hardwood Kalopanax pictus*. Annals of Botany, 2000. **86**(6): p. 1109-1117.
10. Reme, P. and T. Helle, *Assessment of transverse dimensions of wood tracheids using SEM and image analysis*. European Journal of Wood and Wood Products, 2002. **60**(4): p. 277-282.
11. Zimmermann, T., *Spectral imaging and linear unmixing in light microscopy*. Microscopy techniques, 2005: p. 1308-1310.
12. Dickinson, M., et al., *Multi-spectral imaging and linear unmixing add a whole new dimension to laser scanning fluorescence microscopy*. Biotechniques, 2001. **31**(6): p. 1272-1279.
13. Albinsson, B., et al., *The origin of lignin fluorescence*. Journal of Molecular Structure, 1999. **508**(1): p. 19-27.
14. Wondraczek, H., et al., *Photoactive polysaccharides*. Carbohydrate polymers, 2011. **83**(3): p. 1048-1061.
15. Castellan, A., et al., *Studies on fluorescence of cellulose*. Holzforschung, 2007. **61**(5): p. 504-508.
16. Castellan, A., V. Trichet, and J. Pommier, *Studied by UV/Vis Diffuse Reflectance and Fluorescence Spectroscopy*. Journal of pulp and paper science, 1995. **21**(9).

## Capítulo 4

### Resultados: Enfoque Cuantitativo





El presente capítulo aborda los resultados obtenidos tras la aplicación de estrategias de análisis cuantitativo en imágenes espectrales obtenidas por microscopía FT-IR. De este modo se responde al objetivo específico n° 2 la presente tesis.

Los métodos de regresión requieren de patrones de concentración conocida a fin de establecer mediante alguna función matemática (generalmente una función lineal) una correlación entre una o más variables y alguna propiedad que desee ser calibrada, generalmente, la cuantificación de un analito. Sin embargo, cuando la muestra a cuantificar corresponde al pixel de una imagen híper-espectral, se presentan nuevos desafíos. Para empezar, el incremento en resolución espacial que provee la microscopía es proporcional al grado de heterogeneidad que la imagen presenta. Así, en una imagen espectral puede haber presentes compuestos (píxeles) que no hubiesen sido detectados sin el uso de microscopía. El rango de concentraciones que es probable encontrar en una imagen también se expande, puesto que algunos píxeles mostrarán casi exclusivamente un compuesto en particular, mientras que otros no tendrán presencia alguna de éste. El estudio de fibras lignocelulósicas incrementa aún más la dificultad del análisis, ya que es muy difícil contar con material de referencia adecuado para la calibración. En este capítulo se abordan estas interrogantes y se llevan a cabo algunas estrategias que permiten una estimación de la concentración de los principales componentes presentes en biomasa lignocelulósica.

El capítulo se encuentra dividido en dos secciones. La sección 1 corresponde a un manuscrito publicado en la revista *Microchemical Journal* en mayo de 2017. En él se proponen dos estrategias para la estimación de la concentración real de glucanos y lignina en la superficie de la fibra lignocelulósica, usando microscopía FT-IR y MCR-ALS. Una de dichas estrategia incluso prescinde de material de referencia para la calibración. Además, se compara las estrategias propuestas con dos métodos ya establecidos de cuantificación de imágenes por MCR-ALS, y se discuten sus ventajas y desventajas.

La sección 2 muestra un manuscrito terminado que se encuentra en proceso de envío para su publicación. En ese trabajo se vuelve a abordar la cuantificación de componentes lignocelulósicos desde imágenes híper-espectrales, esta vez de celulosa, lignina y hemicelulosa, pero desde un ángulo diferente, comparando las ventajas y desventajas de dos métodos quimiométricos diferentes: PLS y MCR-ALS. PLS es la técnica de calibración

multivariada más ampliamente usada en datos espectrales, mientras que MCR-ALS es uno de los principales métodos de resolución de mezclas. Los modelos PLS construidos tienen también una utilidad cualitativa en el análisis de la muestra, aunque a diferencia de MCR-ALS, debe ser necesariamente calibrado con muestras de referencia para conseguirlo.



## Sección 1

Manuscrito publicado en la revista *Microchemical Journal* el 29 de Mayo de 2017

DOI: <http://dx.doi.org/10.1016/j.microc.2017.05.019>

### **Single pixel quantification strategies using middle infrared hyperspectral imaging of lignocellulosic fibers and MCR-ALS analysis**

Juan A. Araya <sup>a,b,\*</sup>, Renato L. Carneiro <sup>c</sup>, Cristian Arévalo <sup>a,b</sup>, Juanita Freer <sup>b,d</sup>, Rosario del P. Castillo <sup>a</sup>

<sup>a</sup> Laboratory of Infrared Spectroscopy and Chemometrics, Faculty of Pharmacy, University of Concepción, Concepción, Chile

<sup>b</sup> Biotechnology Center, University of Concepción, Concepción, Chile

<sup>c</sup> Department of Chemistry, Federal University of São Carlos, São Carlos, SP, Brazil

<sup>d</sup> Faculty of Chemical Sciences, University of Concepción, Concepción, Chile

\* Corresponding author at: Department of Instrumental Analysis, Faculty of Pharmacy, University of Concepcion, Concepcion, Chile. Tel.: +56 412203749. E-mail address: [jarayaq@udec.cl](mailto:jarayaq@udec.cl) (Juan A. Araya Q.).

E-mail Address of other authors: [renato.lajarim@ufscar.br](mailto:renato.lajarim@ufscar.br) (Renato L. Carneiro), [cristarevalo@udec.cl](mailto:cristarevalo@udec.cl) (Cristian Arévalo), [jfreer@udec.cl](mailto:jfreer@udec.cl) (Juanita Freer), [rosariocastillo@udec.cl](mailto:rosariocastillo@udec.cl) (Rosario del P. Castillo),



# Single pixel quantification strategies using middle infrared hyperspectral imaging of lignocellulosic fibers and MCR-ALS analysis



Juan A. Araya<sup>a,b,\*</sup>, Renato L. Carneiro<sup>c</sup>, Cristian Arévalo<sup>a,b</sup>, Juanita Freer<sup>b,d</sup>, Rosario del P. Castillo<sup>a</sup>

<sup>a</sup> Laboratory of Infrared Spectroscopy and Chemometrics, Faculty of Pharmacy, University of Concepción, Concepción, Chile

<sup>b</sup> Biotechnology Center, University of Concepción, Concepción, Chile

<sup>c</sup> Department of Chemistry, Federal University of São Carlos, São Carlos, SP, Brazil

<sup>d</sup> Faculty of Chemical Sciences, University of Concepción, Concepción, Chile

## ARTICLE INFO

### Article history:

Received 15 November 2016

Received in revised form 4 April 2017

Accepted 28 May 2017

Available online 29 May 2017

### Keywords:

Infrared imaging

MCR-ALS

Lignin

Cellulose

Pixel quantification

Concentration map

## ABSTRACT

The surface of lignocellulosic pulp fibers was studied at microscopic level using Fourier transformed infrared imaging to obtain information regarding pattern distribution and concentration of the lignocellulosic components. As each pixel of the hyperspectral image contains a mixture of spectral information of all of the components present within, multivariate curve resolution-alternating least squares (MCR-ALS) analysis was performed to recover the pure spectra of the main components of lignocellulosic materials and determine their distribution on the surface of the fiber. Although MCR-ALS is not a quantitative analysis, the relative concentration profile obtained by MCR-ALS can be used to obtain real pixel-to-pixel quantitative information of the components. Some strategies have been described to transform the values obtained from MCR-ALS into real concentration units; however, they have been applied mostly to artificial mixtures, using available reference materials to create calibration models. In this work, two simple strategies are proposed to estimate single pixel concentration when no reference material is available. These strategies can be especially relevant for microscale quantitative analysis of natural, complex and heterogeneous samples.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

Lignocellulosic biomass is a promising source for ethanol production, since it is the most abundant and renewable organic matter of the biosphere, portraying 50% of the world's biomass [1]. However, wood cells have developed an orchestrated and complex structural organization, which bestow them with a natural resistance to physical and biological attacks [2]. In order to have access to fermentable sugars from lignocellulosic materials through enzymatic saccharification, it is necessary to overcome this hindrance. Different wood pretreatments have been developed to improve enzyme accessibility and cellulose digestibility [3,4]. The main premise is the removal or relocalization of lignin and the disruption of the cellulose crystalline structure [5]. Although there has been a great amount of efforts and research to understand wood recalcitrance to enzymes, the mechanisms contributing to lignocellulosic biomass recalcitrance are still not well understood [4,6]. As these mechanisms take place at a microscopic scale, it is necessary to use tools that allow microstructural, cellular or molecular analysis to

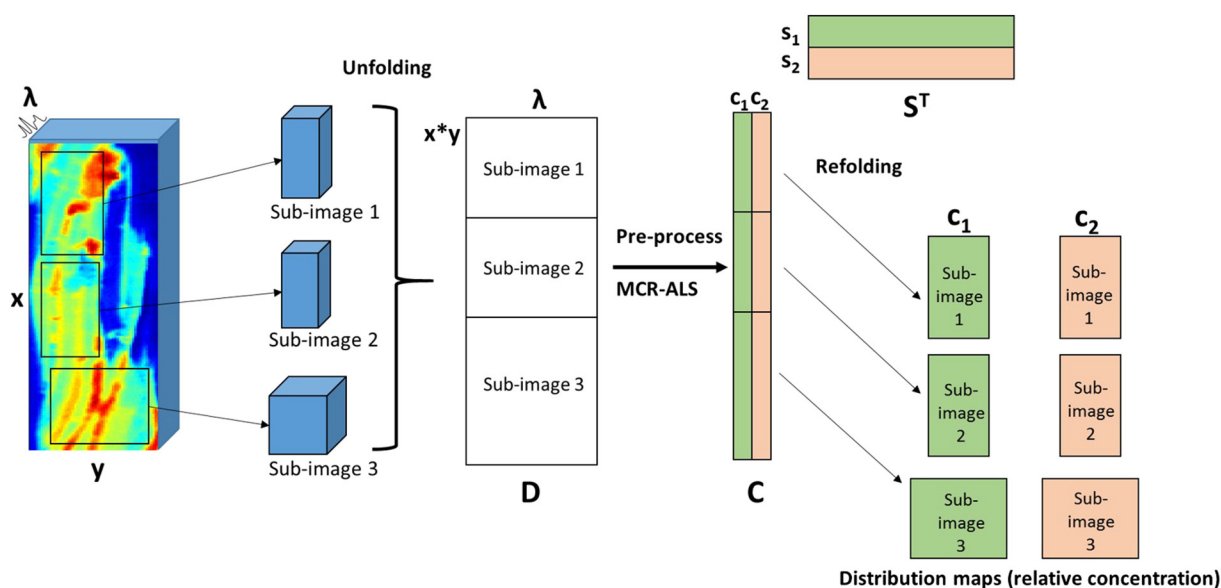
have a better understanding to achieve an efficient conversion process [3]. In this scenario, infrared imaging is an optimal technique, since it allows microstructure analysis coupled with spectral profiles, which could provide localized qualitative and quantitative chemical analysis.

Hyperspectral images can be considered as a data cube with two dimensions associated to pixels position coordinates and a third dimension containing the spectral information. As the spectra corresponds to a mixture of all the components present within a pixel, multivariate resolution methods are often performed to analyze spectroscopic images. Multivariate Curve Resolution-Alternating Least Square (MCR-ALS) method is one of the most used for signal separation of bilinear data [7]. MCR-ALS has been extensively explained in several articles [8–16]. Briefly, is an iterative resolution method that decomposes instrumental data into pure contributions according to the equation  $D = CS^T + E$ , where  $D$  is the original data matrix,  $C$  is the matrix of relative concentration profiles (or intensity of normalized spectra),  $S$  is the pure spectra matrix (normalized spectra) and  $E$  is the experimental error matrix.

The resolved concentration profiles depicted in  $C$  matrix from MCR-ALS provides only a relative concentration [17]. This information may be used to obtain real concentration units by applying appropriate strategies that set correlation using reference values. Different strategies have been proposed for real quantification using MCR-ALS, including constraints applied during the iterative optimization, as closure [18] and correlation constraints [19] as well as other strategies that take

\* Corresponding author at: Department of Instrumental Analysis, Faculty of Pharmacy, University of Concepcion, Concepcion, Chile.

E-mail addresses: [jaraya@udec.cl](mailto:jaraya@udec.cl) (J.A. Araya), [renato.lajarim@ufscar.br](mailto:renato.lajarim@ufscar.br) (R.L. Carneiro), [cristarevalo@udec.cl](mailto:cristarevalo@udec.cl) (C. Arévalo), [jfreer@udec.cl](mailto:jfreer@udec.cl) (J. Freer), [rosariocastillo@udec.cl](mailto:rosariocastillo@udec.cl) (R.P. Castillo).



**Fig. 1.** Schematic representation of bilinear matrix decomposition for MCR-ALS analysis. Three sub-images from each sample were selected, unfolded, concatenated, pretreated and analyzed altogether.

place after the MCR-ALS resolution [17,20]. However, most of them require reference materials for calibration models and are performed on prepared or artificial mixtures; therefore, their application in more complex matrices, such as lignocellulosic fibers, might be difficult. Therefore, it is necessary to explore quantification strategies that can be applied when little to no information regarding reference materials is available.

In this work, two simple strategies were assessed to obtain pixel-to-pixel quantification of lignin and glucans in lignocellulosic fibers using the MCR-ALS method. MCR-ALS constraints of closure and correlation are also used to discuss their adequacy in quantification of natural heterogeneous samples.

## 2. Material and methods

### 2.1. Wood chip pretreatments

Three different wood pre-treatment techniques were performed (Organosolv, diluted acid hydrolysis and autohydrolysis) to obtain lignocellulosic pulps from *Pinus radiata* or *Eucalyptus globulus*. The variation in conditions used for each wood pretreatment were based on three different fractional factorial designs (one for each pretreatment used), which were performed in order to obtain a high variability in lignin content and distribution patterns on the surface of the wood fiber. The specific conditions for each sample can be found in Table S1, in Supplementary material. However, as the emphasis of this article was in the quantitative analysis, only images that were representative of the samples were selected (a representativeness criteria for the images was applied, as reported in Section 2.5, strategy A), therefore, only some of those samples are included in this work.

#### 2.1.1. Organosolv pretreatment

*Pinus radiata* chips (100 g dry weight) were loaded in a 1 L Parr reactor along with 700 mL of pulping liquor, composed of water:ethanol mixtures (50:50–75:25%v/v) and 0.35 to 0.7% w/w of sulfuric acid as catalyst. The temperature varied between 180 °C to 190 °C and the reaction was held for 30 min after it reached the working temperature.

#### 2.1.2. Diluted acid hydrolysis

*Eucalyptus globulus* chips were used. Reaction conditions were applied as follow: temperature from 113° to 177 °C, 1:4 to 1:8 wood:liquor

ratio, 0.1 to 0.45% w/w sulfuric acid, and constant stirring during 20 min in a 1 L Parr reactor.

#### 2.1.3. Autohydrolysis

*E. globulus* chips were used. Time reaction varied between 4 and 81 min, in a range temperature of 160° to 190 °C in a Parr reactor.

### 2.2. Pretreated wood chemical characterization

Chemical characterization of all pretreated wood was performed by reference methods, through acid hydrolysis, followed by HPLC analysis for sugars determination, and UV spectroscopy and gravimetric analysis for lignin determination, as described by Castillo et al. [21].

### 2.3. Image acquisition

FT-IR spectra and images of pulp fibers were analyzed by micro FT-IR using an FT-IR Spectrum Frontier/Spotlight 400 Microscopy System (PerkinElmer, Inc.) with a linear array of 16 MCT detectors. Areas with side dimensions from 50 μm to 200 μm of isolated fibers previously dried for 8 h at 45 °C were scanned by attenuated total reflectance (ATR). For that, it was used a germanium crystal, a pixel size of 1.56 μm × 1.56 μm and spectral range from 4000 to 748 cm<sup>-1</sup> with spectral

**Table 1**

Chemical characterization of lignocellulosic pulps obtained through Organosolv (Or), autohydrolysis (Au) and diluted acid hydrolysis (DAH) pretreatments. The parameters for a same type pretreatment were different in order to obtain different concentration of glucans and lignin (Supplementary material).

Sample	Specie	Pretreatment	*Glucans (%)	*Lignin (%)
Or1	<i>P. radiata</i>	Organosolv	66.8	25.9
Or3	<i>P. radiata</i>	Organosolv	73.5	23.8
Or5	<i>P. radiata</i>	Organosolv	69.0	24.2
Au2	<i>E. globulus</i>	Autohydrolysis	60.7	33.0
Au3	<i>E. globulus</i>	Autohydrolysis	57.8	29.2
Au4	<i>E. globulus</i>	Autohydrolysis	60.8	32.2
Au5	<i>E. globulus</i>	Autohydrolysis	50.2	27.8
Au6	<i>E. globulus</i>	Autohydrolysis	57.1	28.4
DAH2	<i>E. globulus</i>	Diluted acid hydrolysis	55.8	29.8
DAH4	<i>E. globulus</i>	Diluted acid hydrolysis	59.5	30.6
DAH7	<i>E. globulus</i>	Diluted acid hydrolysis	45.9	26.5

\*All values are average of three replicates, with RSD < 5%.

resolution of  $8\text{ cm}^{-1}$  and 16 scans per spectra. After image acquisition, spectra were transformed to absorbance, and spectral range between  $4000$  and  $2000\text{ cm}^{-1}$  was eliminated from each image to focus the resolution analysis in the “fingerprint area”, which contains most of the qualitative spectral information.

#### 2.4. Image treatment

Unless expressed otherwise, all the calculations were performed using Matlab R2013a (Mathworks, Natick, MA). Spectral images can be displayed as a data cube, with two spatial dimensions and a third spectral dimension. As each pixel contain the spectral information of the chemical components within it, its mathematical description does not require a tri-dimensional (tri-linear) models and can be unfolded into a two dimension matrix without loss of information (pixels

× spectra, matrix D in Fig. 1). After the application of mathematical pre-processing and chemometric analysis, the two-dimension data matrix can be conveniently refolded into the original spatial dimensions to recover the original spatial description and pixel position in the image. As lignocellulosic fibers have a narrow and elongated shape, to simplify representation of resolved images, three different zones of each image were selected. Each sub-image was unfolded and concatenated as shown in Fig. 1. The augmented matrix composed of the three concatenated sub-images was considered to contain the chemical information of the original image and was submitted to smoothing (moving average, 5 points window) with first order derivative followed by normalization, prior to MCR-ALS analysis. After the MCR-ALS analysis and the quantification strategies proposed in the following paragraphs were performed, the augmented matrix was separated according the original position of the three sub-images and was refolded into the

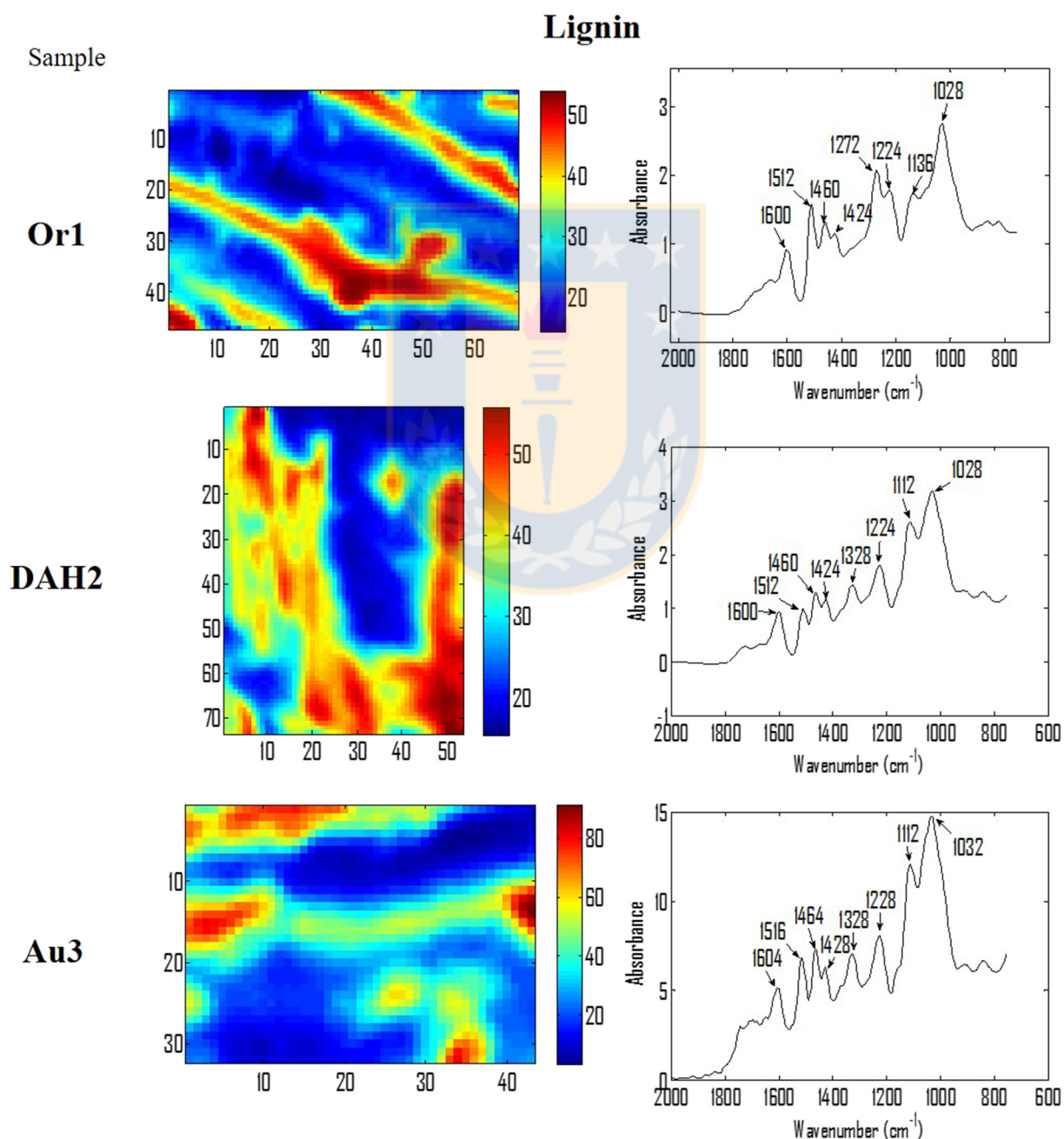


Fig. 2. MCR-ALS recovered spectra for lignin and lignin concentration maps for samples Or1, DAH2 and Au3 quantificated by strategy A. Each image corresponds to sub-images of samples from three different wood pulping pretreatment techniques.

original spatial dimensions, having as result three resolved reconstructed images from different zones of the original fiber.

### 2.5. MCR-ALS analysis and individual pixel quantification strategies

MCR-ALS analysis was performed using the updated version of the “MCR-ALS Graphic user-Friendly Interface” developed by Jaumot et al. [19] on the augmented matrix composed by three unfolded sub-images of the same sample (as explained in Section 2.4). The number of components present in samples was assessed through singular value decomposition (SVD) and two components were selected in each analysis. The interactive self-modeling mixture analysis algorithm “Pure” was used as initial estimative of the two most pure spectra in the original data. As explained previously, the sample matrix was derivative and normalized, therefore MCR-ALS analysis was performed using non-negativity constraint only in concentration profile (row mode), as

constraint of non-negativity in spectral profile would lead to erroneous results. Once the MCR-ALS analysis was done, the derivative spectra was integrated to recover the spectral features of the resolved component. After the MCR-ALS resolution, two different strategies were used to transform the resolved concentration profile into real concentration estimative of lignin and glucans in each pixel.

#### 2.5.1. Strategy A: normalization of pixels concentration using real bulk concentration

The first strategy consisted in a simple normalization step of relative concentration matrix obtained by MCR-ALS, as described by Colares et al. [20]. This approach assumes a direct relation between the global concentration of an analyte in the bulk sample, obtained by reference methods, and the average relative concentration of all the pixels for the same analyte in the MCR-ALS concentration profile, as described

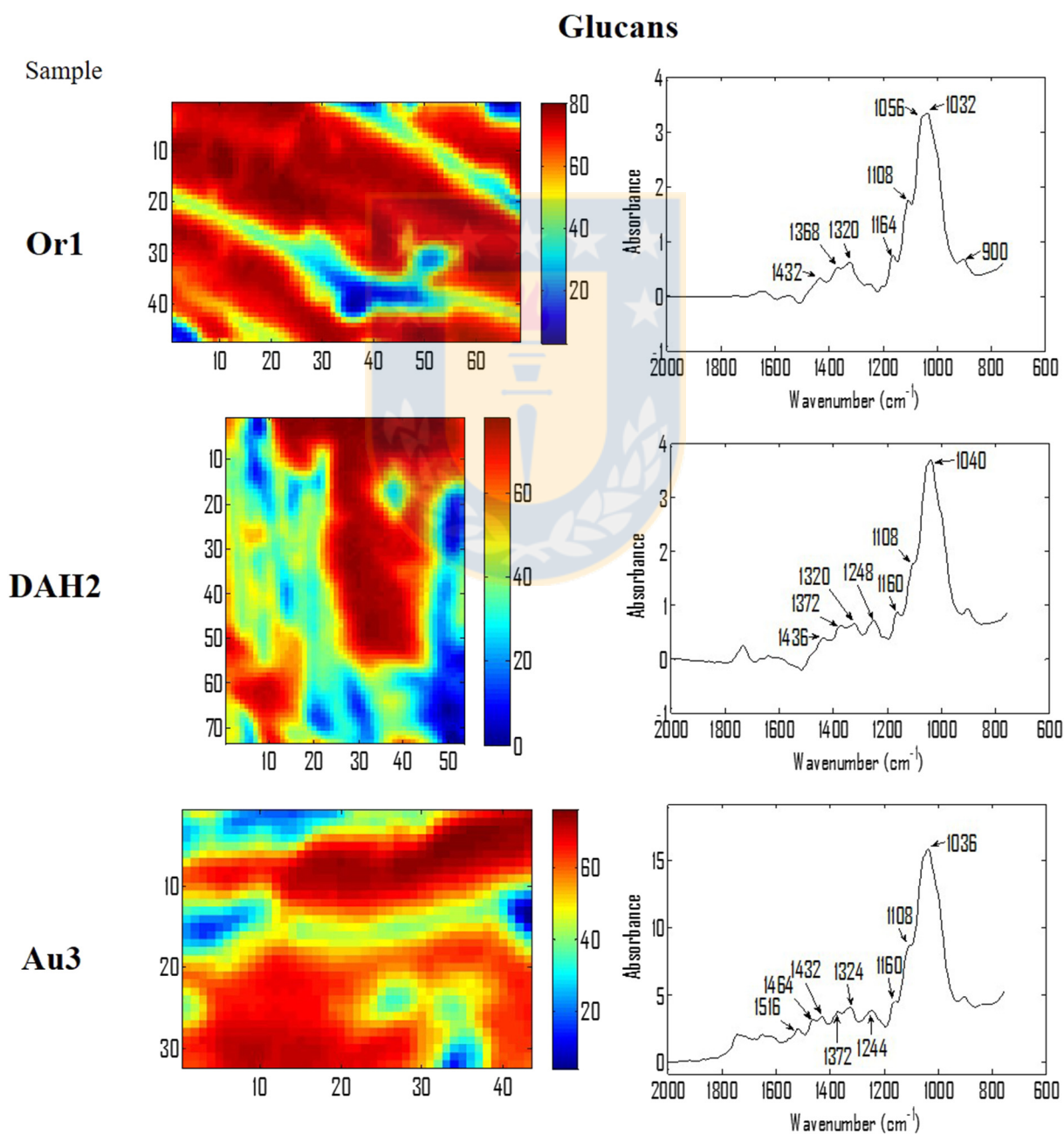


Fig. 3. MCR-ALS recovered spectra for glucans and glucans concentration maps for samples Or1, DAH2 and Au3 quantificated by strategy A. Each image corresponds to sub-images of samples from three different wood pulping pretreatment techniques.

in Eq. (1).

$$\hat{y}_i = \frac{I_i^* \bar{y}}{\bar{I}} \quad (1)$$

Thus, the concentration estimation in each pixel ( $\hat{y}_i$ ) is obtained by multiplying the relative concentration in that pixel ( $I_i$ ), obtained from the C matrix of MCR-ALS, with the reference global concentration ( $\bar{y}$ ), and divide it by the average relative concentration of all pixels ( $\bar{I}$ ). As the concentration obtained by reference methods is evaluated in the bulk sample, it was necessary to determine the representativeness of the image related to the bulk sample in order to validate this approach. Using the average spectrum of each image from different samples, and using the concentration of those samples obtained from reference methods, a partial least square regression (PLS) was performed using Pirouette® (Infometrix, Inc.) and the root mean square error of calibration and cross-validation (RMSEC and RMSECV respectively) were used as a representativeness parameter. Data for PLS analysis was transformed by first order derivative and normalization and pre-processed by mean center.

#### 2.5.2. Strategy B: normalization by pixels pureness

The second strategy consist in locating the pixel with the highest relative concentration ( $I_{\max,1}$  for component 1,  $I_{\max,2}$  for component 2) recovered by MCR-ALS, for every component in the C matrix, and assuming it contains 100% of that component. Then, the real concentration in a pixel  $i$  for component 1 ( $\hat{y}_{i,1}$ ) is estimated by converting the relative concentration of the pixel ( $I_{i,1}$  for component 1,  $I_{i,2}$  for component 2) into a percentage of  $I_{\max,1}$ , followed by a correction where the sum of concentration of all components (1 and 2) is 100%. This strategy is summarized in Eq. (2). A step-to-step description of Eq. (2) is provided in Supplementary material.

$$\hat{y}_{i,1} = \frac{1}{1 + \frac{I_{i,2}^* I_{\max,1}}{I_{i,1}^* I_{\max,2}}} \quad (2)$$

Two established quantification methods using MCR-ALS constraints were also performed for comparative purposes: closure constraint and correlation constraint, proportionated by the MCR-ALS graphical user-friendly interface [19]. For closure constraint, it is assumed that the sum of concentration of components is 100% in every pixel. For correlation constraint, an internal calibration model is built with known concentration values of calibration samples, which are grouped in the

same data matrix to be analyzed and are simultaneously resolved. The internal calibration model was built using the mean spectra of several images from different samples, and their global (bulk) concentration obtained from reference methods.

A repeated measures ANOVA was performed between the data sets of bulk concentration obtained by reference methods, the predicted bulk concentration from strategy B and the predicted bulk concentration from closure constraint in order to compare the results of those strategies (correlation constraint could not be included in this analysis since it did not produce satisfactory results). Another statistical parameter used to measure the differences between values obtained by the different methods, was the root mean square error of prediction (RMSEP) for bulk concentrations.

### 3. Results and discussion

#### 3.1. Characterization of bulk samples and samples heterogeneity

The concentration of lignin and glucans from the bulk samples obtained from reference methods are shown in Table 1.

In a previous work, FT-IR microspectroscopy and MCR-ALS was successfully used to characterize and differentiate lignin and carbohydrate rich zones on pretreated *Pinus radiata* pulps [21]. In that study we were able to observe that depending on the severity of the pretreatment process, i.e. varying temperature, acidity and reaction time, changes in the recovered spectra occurred, accounting for changes in lignin identity, and different degrees of disruption between the lignin-carbohydrates arrangement. Such heterogeneity and spectral differences between samples turn the quantification of those components at individual pixel level a very hard task, as it is unlikely to obtain standards for each sample (pixel) to create calibration sets, and in that case, we did not attempted to quantify the lignocellulosic components. In the present work, the samples proceed from two different species, *P. radiata* and *E. globulus*, which have different lignin-carbohydrates ratios, and three different pretreatments were applied even varying the conditions between samples (see Table S1 in Supplementary material). Therefore, these samples portray a higher heterogeneity in lignocellulosic components.

Among the samples in Table 1, the pulp with the minimal concentration of lignin had 23.8% while the pulp with the highest concentration presented 33.0% of this compound, portraying a range of 9.2 percentage units. For glucans, 45.9% and 73.5% were the lowest and highest concentration among samples, portraying a range of 27.6 percentage units.

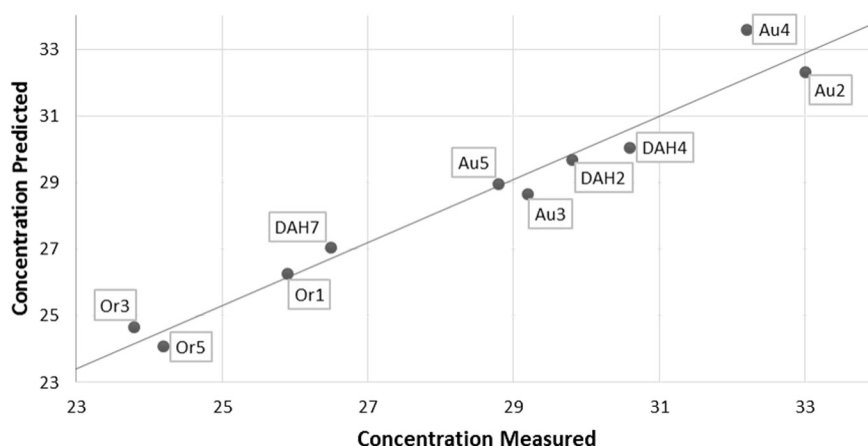


Fig. 4. Correlation between predicted versus measured global concentration values for lignin using a PLS model. Mean spectra from images was used to predict concentrations and bulk sample concentration from reference methods was used as measured concentration.



**Table 2**  
Lignin and glucans estimated concentration using strategy B, and correlation constraint.

Samples	Strategy B		Closure constraint	
	Lignin (%)	Glucans (%)	Lignin (%)	Glucans (%) <sup>a</sup>
Or1	37.5	62.5	44.7	55.3
Or3	29.6	70.4	8.9	91.1
Or5	22.1	77.9	21.0	79.0
Au2	46.3	53.7	23.9	76.1
Au3	29.5	70.5	25.9	74.1
Au4	26.0	74.0	26.6	73.4
Au5	28.8	71.2	3.0	97.0
Au6	34.3	65.7	33.7	66.3
DAH2	45.4	54.6	49.0	51.0
DAH4	22.5	77.5	16.7	83.3
RMSEP <sup>b</sup>	8.5	11.5	13.8	20.2

<sup>a</sup> Significantly different from bulk concentration obtained by reference methods ( $p < 0.01$ ).

<sup>b</sup> Root mean square errors of prediction (RMSEP) were calculated using bulk concentration obtained by reference methods.

These ranges represent the variation between the average concentrations of the bulks samples (global concentration); however, the variation within a single image (pixel to pixel) is expected to be much greater. It has been described relocation of lignin into droplets during pretreatment process, which would result in lignin rich zones and lignin depleted zones (carbohydrates rich zones) [21,22], therefore, spectral images could be expected to portray pixels containing either almost only lignin or glucans. Using bulk concentrations to create calibration models for pixel-to-pixel quantification would not cover all the concentration range expected in a single image. On the other hand to use artificial mixtures with pure materials to create a broader range calibration model is not possible, since some lignocellulosic components may suffer chemical changes during wood pretreatment [3,21], which could change spectral features compared to pure reference materials.

Therefore, in the absence of reference materials and samples in a wider composition range, it was necessary to explore quantification strategies that suits these shortcomings.

### 3.2. MCR-ALS recovered spectra

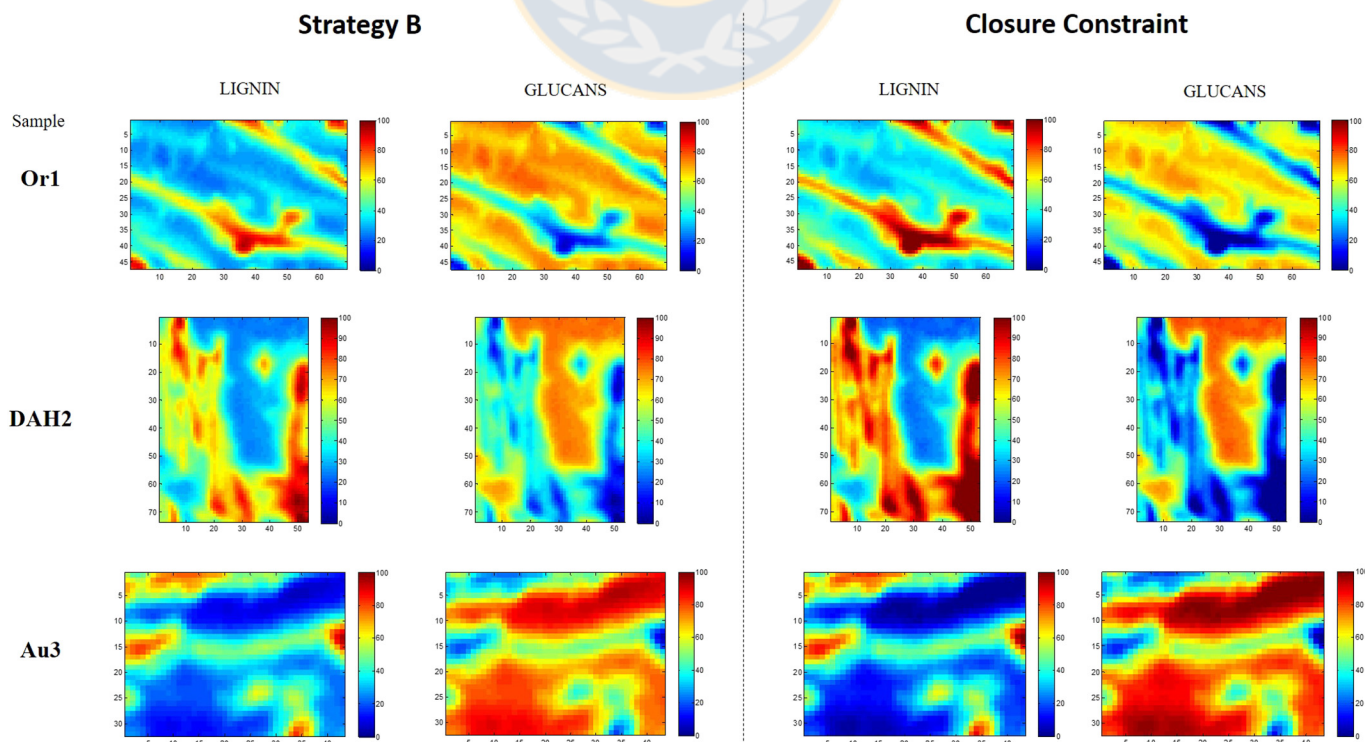
First derivative was applied as pretreatment of data in order to improve data quality, since derived spectra increases selectivity for overlapped bands increasing the performance of MCR-ALS [23]. Attempts of use MCR-ALS without derivative (using non-negativity for both concentration and spectral profile) were performed, however, the quality of recovered results was poor compared to first derivative pretreatment. The derived spectra emphasizes spectral differences enough to compensate the non-use of non-negativity constraint in spectra (column mode) to overcome rotational ambiguity.

Spectra recovered by MCR-ALS show characteristics of lignins (Fig. 2) and glucans (Fig. 3) spectra. In Fig. 2 it can be seen that spectrum from the first sample (Or1) from *P. radiata* differs from the other two samples (DAH2 and Au3) from *E. globulus* by the presence of bands in  $1270\text{ cm}^{-1}$  and  $1136\text{ cm}^{-1}$ . These bands would correspond to aromatic C-H in-plane deformation of guaiacyl units in lignin-c, which is found mainly in soft woods [24]. *E. globulus* samples, have a slightly different spectra corresponding to lignin-GS type, typical of hard wood [24].

### 3.3. MCR-ALS quantification: strategy A

Reconstructed images using strategy A are shown in left-side of Figs. 2 and 3. Only one sub-image from each wood species and pretreatment was selected as example. The other reconstructed sub-images from the same samples can be found in Supplementary data (Fig. 15).

As explained in Section 2.5 this method forces the value of global concentration from a sample obtained by reference methods, into the average relative concentration of all the pixels obtained from MCR-



**Fig. 5.** Lignin and glucans concentration maps for samples Or1, DAH2 and Au3 quantified using strategy B and using closure constraint from MCR-ALS.

ALS to estimate individual pixel concentration. It is, therefore, highly dependent on the representativeness of the image related to the bulk sample. Microscopic analysis has an inherent representativeness issue, since the analyzed region is limited and may not be representative of a macroscopic sample. This is especially relevant for highly heterogeneous materials such as a lignocellulosic fiber, whose chemical content may vary depending on the anatomical origin of that fiber in the tree [25]. If the image corresponds to an individual fiber whose lignin and cellulose content differs much from the bulk sample, then the concentration estimated for individual pixels through this method would be biased proportionally.

In order to assess representativeness of the image with the bulk sample, we used a PLS regression model for lignin prediction, using the average spectra of each image and the concentrations obtained by reference methods (see Fig. 4). The PLS model showed a good prediction of lignin for the bulk, with low estimation errors of calibration (RMSEC 0.2477%,  $r_{\text{val}}^2$  0.998) and cross validation (RMSECV 0.6434%,  $r_{\text{val}}^2$  0.978), using 3 latent variables (85% of explained variance). If the fiber images could not predict the bulk concentration adequately, then we could not state that the fibers were representative samples of the bulk. As lignin concentration from all the bulk samples were satisfactorily predicted using the average spectra of the images, we can conclude that the

images are representative samples of the bulk. This verification serves as validation strategy for this method.

#### 3.4. MCR-ALS quantification: strategy B versus closure constraint

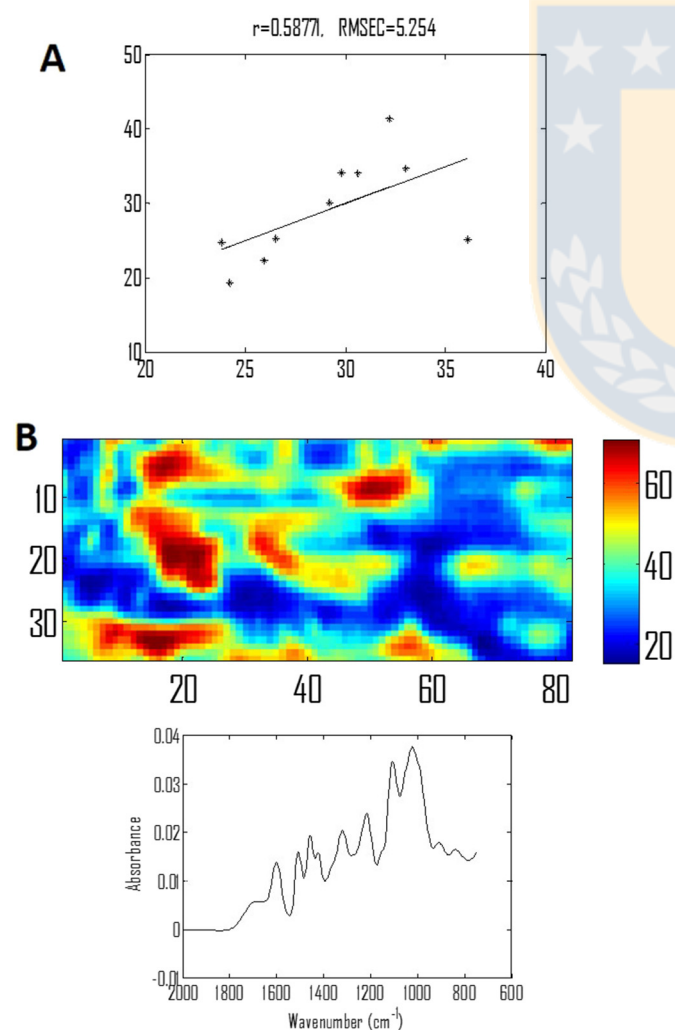
An interesting characteristic of this strategy is that it does not need information from samples concentrations obtained from reference methods. It assumes that the pixel with the highest concentration for each component recovered by MCR-ALS will contain nearly 100% of that component. Depending on the sample nature and the image acquisition technique, this assumption could be more or less valid. In attenuated total reflectance (ATR) acquisition mode, the evanescent waves penetrate approximately 900 nm into the sample when germanium is used as the ATR crystal [26]. The pixel size selected for this study was  $1.56 \mu\text{m} \times 1.56 \mu\text{m}$  (the minimum spatial resolution possible with the instrument). On the other hand, lignin droplets that relocate on the wood fiber surface after hydrothermal pretreatment have sizes around  $3 \mu\text{m}$  [21], therefore it might be expected that a single pixel would contain either lignin or glucans in approximately 100% of concentration in these samples, as they are the two major components in wood fibers. Hemicelluloses, also present in wood, were not able to resolve using MCR-ALS, possible due to the high spectral similarity with glucans and the low concentration expected, since hydrothermal pretreatment removes these component from the wood fiber.

Estimation of pixel real concentrations was performed using Eq. (2). The pixels with the highest concentrations for each component are used as reference to estimate concentrations in each pixel, under the condition that the sum of the two components does not exceed 100% of value. This condition is similar to the application of the closure constraint in the MCR-ALS analysis, at which each pixel is supposed to accomplish a constant mass balance of 1, representing 100% of concentration [11,18]. However, in closure constraint this assumption is imposed during the iterative optimization, while in strategy B the concentration data is normalized to 100% once the MCR-ALS analysis was already done. Table 2 shows the average concentration values for images from different samples, estimated by strategy B and closure constraint. Root mean square errors of prediction (RMSEP) were calculated using the bulk concentration obtained from reference methods in Table 1. The RMSEP for bulk concentration prediction using strategy B was smaller than RMSEP of the closure constraint approach (see Table 2). A comparison with strategy A was not performed, since mean concentration values in strategy A are normalized by bulk concentration, therefore they will be the same values as those obtained in the bulk sample by reference methods.

A repeated measures ANOVA between the data sets of bulk concentration obtained by reference methods, and the bulk concentration predicted by strategy B and by closure constraint shows that, for glucan quantification, variation between column means is significantly greater than expected by chance ( $p$  value = 0.0123). A Tukey-Kramer multiple comparisons post test was applied, which revealed that the differences of means for glucans between the bulk reference and the prediction of the closure constraint approach is significant ( $p$  value < 0.01). The post test also showed that for glucans there was not significant differences of means between the bulk concentration and the values predicted by strategy B. For lignin quantification, the ANOVA test showed that variation between column means was not significant ( $p$  value > 0.05).

Reconstructed sub-images using strategy B and using closure constraint are shown in Fig. 5. Other reconstructed images from the same samples using strategy B are available in the Supplementary material, in Fig. 2S. Pixel concentrations estimated through this strategy are reasonable close to those obtained with strategy A, even though no information regarding reference methods were used in this case.

This strategy depends on the MCR-ALS performance and would be affected by any limitation the algorithm may have. The sum of concentration of the components resolved by MCR-ALS are forced to 100% in each pixel and any less abundant component that MCR-ALS fails to



**Fig. 6.** Image quantification of lignin from sample Eg-Au2 using MCR-ALS correlation constraint. (A) Scatter plot and correlation between predicted and known concentration values for the internal univariate calibration model. (B) Lignin concentration map and recovered spectra.

resolve would induce to error in the quantification. This is a limitation of the method, although it could be argued that if MCR-ALS fails to recover a component present in the mixture, then it is expected that its concentration is too low to be resolved and it would have a small impact in the quantification. On the other hand, strategy B is proposed as a quantification alternative when there is little to no prior knowledge about the mixture and the components within it, as is the case in the blind source separation problem. Therefore, we believe this method is suitable and reasonably reliable when working with complex natural samples, when no prior information about its components identity or concentration is available.

### 3.5. MCR-ALS quantification using correlation constraint

Correlation as an MCR-ALS constraint was first introduced by Antunes et al. [27] in 2002. Since then it has been used as a MCR-ALS based quantification method [27–31]. In this method, univariate calibration models are established for selected compounds which rescales the concentration values found by the ALS optimization procedure into real concentration units during the iterative optimization [19]. Although this strategy seems to work very well for artificial mixtures, it seems to be less applicable to more heterogeneous or natural samples, when there is less information available regarding reference materials. To implement correlation constraint optimally it requires access to reference materials and previous knowledge of the exact identity of the components present within the mixture to build an appropriate calibration curve. If such materials are available, correlation constraint is the most suitable method for quantification with MCR-ALS. In our particular case, it was very difficult to obtain reference materials. Besides lignin chemical variability in different wood species, each pretreatment process used (Organosolv, autohydrolysis and diluted acid hydrolysis) incorporates chemical changes to lignin and other components due to the application of temperature, solvents and pressure. Therefore, lignin or cellulose from commercial origin were not suitable as reference material for the construction of calibration curves. To use correlation constraint we built the internal calibration model using the mean spectra of several images from different samples, and their global (bulk) concentration obtained from reference methods. This approach is far from optimal. The mayor issue we found using this approach was divergence during the iterative optimization. Glucans could not be resolved in any of our samples, and lignin was resolved only in few samples (see Fig. 6). The problem seems to be that the range between the highest and lowest bulk concentration in samples used for internal calibration is less than the concentration range within a single image, thus, the model fails to resolve or predict out-range data.

## 4. Conclusions

Different strategies for single pixel quantification in infrared images of natural/heterogeneous samples were applied, using the MCR-ALS concentration profile information. Some assumptions must be made in each case; therefore, the estimated concentration accuracy will depend on the legitimacy of each assumption. Strategy A requires normalizing the average relative concentration of an image with the global bulk concentration for those samples obtained by reference methods. This strategy is very simple if reference concentration values from the bulk sample are available, however, it is needed to check whether the image is representative of the sample to validate this approach. Strategy B re-escalates single pixel concentration using the pixel with the maximum relative concentration for every component as reference, assuming it has 100% of concentration. Although this assumption could be debatable, results are reasonably closed to those obtained using strategy A, considering that no external reference information is used regarding concentration values. Furthermore, strategy B could be promising for complex samples where little to no information about its components is provided and no reference materials are available. Both methods

performed well and could be a more suitable alternative than MCR-ALS constraint based strategies for pixel-to-pixel quantification of heterogeneous samples.

## Funding

This work was supported by FONDECYT, Chile [Initiation Project 11130388]. Juan Araya thanks to CONICYT-PCHA/Doctorado Nacional scholarship, Chile [Folio 2013-21130641].

## Acknowledgements

Authors would like to thank “Centro de Microscopia Avanzada CMA Bío-Bío” of the University of Concepción, Chile.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.microc.2017.05.019>.

## References

- [1] O.P. Ward, A. Singh, Bioethanol technology: developments and perspectives, *Adv. Appl. Microbiol.* 51 (2002) 53–80.
- [2] F. Xu, et al., Anatomy, ultrastructure and lignin distribution in cell wall of *Caragana korshinskii*, *Ind. Crop. Prod.* 24 (2) (2006) 186–193.
- [3] P. Alvira, et al., Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review, *Bioresour. Technol.* 101 (13) (2010) 4851–4861.
- [4] C. Wyman, Key features of pretreated lignocelluloses biomass solids and their impact on hydrolysis, *Bioalcohol Production: Biochemical Conversion of Lignocellulosic Biomass 2010*, p. 73.
- [5] N. Mosier, et al., Features of promising technologies for pretreatment of lignocellulosic biomass, *Bioresour. Technol.* 96 (6) (2005) 673–686.
- [6] X. Meng, A.J. Ragauskas, Recent advances in understanding the role of cellulose accessibility in enzymatic hydrolysis of lignocellulosic substrates, *Curr. Opin. Biotechnol.* 27 (2014) 150–158.
- [7] M.B. Lopes, et al., Near-infrared hyperspectral unmixing based on a minimum volume criterion for fast and accurate chemometric characterization of counterfeit tablets, *Anal. Chem.* 82 (4) (2010) 1462–1469.
- [8] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution, *J. Chemom.* 9 (1) (1995) 31–58.
- [9] J. Felten, et al., Vibrational spectroscopic image analysis of biological material using multivariate curve resolution–alternating least squares (MCR-ALS), *Nat. Protoc.* 10 (2) (2015) 217–240.
- [10] A. de Juan, R. Tauler, Multivariate curve resolution (MCR) from 2000: progress in concepts and applications, *Crit. Rev. Anal. Chem.* 36 (3–4) (2006) 163–176.
- [11] J. Jaumot, et al., A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, *Chemom. Intell. Lab. Syst. 76* (1) (2005) 101–110.
- [12] R. Tauler, *Chemometric Tools for Image Analysis. Infrared and Raman Spectroscopic Imaging*, 2009.
- [13] A. De Juan, R. Tauler, Chemometrics applied to unravel multicomponent processes and mixtures: revisiting latest trends in multivariate resolution, *Anal. Chim. Acta* 500 (1) (2003) 195–210.
- [14] B.O. Budevska, S.T. Sum, T.J. Jones, Application of multivariate curve resolution for analysis of FT-IR microspectroscopic images of in situ plant tissue, *Appl. Spectrosc.* 57 (2) (2003) 124–131.
- [15] A. de Juan, J. Jaumot, R. Tauler, Multivariate curve resolution (MCR). Solving the mixture analysis problem, *Anal. Methods* 6 (14) (2014) 4964–4976.
- [16] C. Ruckebusch, L. Blanchet, Multivariate curve resolution: a review of advanced and tailored applications and challenges, *Anal. Chim. Acta* 765 (2013) 28–36.
- [17] S. Piqueras, et al., Relevant aspects of quantification and sample heterogeneity in hyperspectral image resolution, *Chemom. Intell. Lab. Syst.* 117 (2012) 169–182.
- [18] J.M. Amigo, et al., Study of pharmaceutical samples by NIR chemical-image and multivariate analysis, *TrAC Trends Anal. Chem.* 27 (8) (2008) 696–713.
- [19] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: new features and applications, *Chemom. Intell. Lab. Syst.* 140 (2015) 1–12.
- [20] C.J. Colares, et al., Near infrared hyperspectral imaging and MCR-ALS applied for mapping chemical composition of the wood specie *Swietenia macrophylla* King (Mahogany) at microscopic level, *Microchem. J.* 124 (2016) 356–363.
- [21] R.d.P. Castillo, et al., Fourier transform infrared imaging and microscopy studies of *Pinus radiata* pulps regarding the simultaneous saccharification and fermentation process, *Anal. Chim. Acta* 866 (2015) 10–20.
- [22] B.S. Donohoe, et al., Visualizing lignin coalescence and migration through maize cell walls following thermochemical pretreatment, *Biotechnol. Bioeng.* 101 (5) (2008) 913–925.
- [23] R.L. Carneiro, R.J. Poppi, Homogeneity study of ointment dosage forms by infrared imaging spectroscopy, *J. Pharm. Biomed. Anal.* 58 (2012) 42–48.

- [24] O. Faix, Classification of lignins from different botanical origins by FT-IR spectroscopy, *Holzforschung Int. J. Biol. Chem. Phys. Technol. Wood* 45 (s1) (1991) 21–28.
- [25] Z. Zhang, et al., Anatomy, cell wall ultrastructure and inhomogeneity in lignin distribution of *Broussonetia papyrifera*, *Cellul. Chem. Technol.* 46 (3–4) (2012) 157–164.
- [26] B.C. Smith, *Fundamentals of Fourier transform infrared spectroscopy*, CRC Press, 2011.
- [27] M. Antunes, et al., Multivariate curve resolution of overlapping voltammetric peaks: quantitative analysis of binary and quaternary metal mixtures, *Analyst* 127 (6) (2002) 809–817.
- [28] T. Azzouz, R. Tauler, Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples, *Talanta* 74 (5) (2008) 1201–1210.
- [29] H.C. Goicoechea, A.C. Olivieri, R. Tauler, Application of the correlation constrained multivariate curve resolution alternating least-squares method for analyte quantitation in the presence of unexpected interferences using first-order instrumental data, *Analyst* 135 (3) (2010) 636–642.
- [30] L.B. Lyndgaard, F. van den Berg, A. de Juan, Quantification of paracetamol through tablet blister packages by Raman spectroscopy and multivariate curve resolution-alternating least squares, *Chemom. Intell. Lab. Syst.* 125 (2013) 58–66.
- [31] J. Jaumot, et al., Blending process modeling and control by multivariate curve resolution, *Talanta* 117 (2013) 492–504.



## SUPPLEMENTARY MATERIAL

Table 1S: Specific conditions used for wood pretreatment for each sample, based on 3 different DoE.

Sample	Specie	Pretreatment	*Conditions
Au1	<i>E. globulus</i>	Autohydrolysis	190° C/ 70 min
Au2	<i>E. globulus</i>	Autohydrolysis	175° C/ 81 min
Au3	<i>E. globulus</i>	Autohydrolysis	160° C/ 70 min
Au4	<i>E. globulus</i>	Autohydrolysis	190° C/ 15 min
Au5	<i>E. globulus</i>	Autohydrolysis	160° C/ 15 min
Au6	<i>E. globulus</i>	Autohydrolysis	175° C/ 4 min
Or1	<i>P. radiata</i>	Organosolv	180 / 0,7% / 50:50 / 30 min
Or2	<i>P. radiata</i>	Organosolv	190 / 0,7% / 50:50 / 30 min
Or3	<i>P. radiata</i>	Organosolv	190 / 0,35% / 50:50 / 30 min
Or4	<i>P. radiata</i>	Organosolv	190 / 0,35% / 25:75 / 30 min
Or5	<i>P. radiata</i>	Organosolv	190 / 0,35% / 50:50 / 30 min
Or6	<i>P. radiata</i>	Organosolv	190 / 0,7% / 25:50 / 30 min
Or7	<i>P. radiata</i>	Organosolv	180 / 0,7% / 25:75 / 30 min
Or8	<i>P. radiata</i>	Organosolv	180 / 0,35% / 25:75 / 30 min
Or9	<i>P. radiata</i>	Organosolv	180 / 0,35% / 50:50 / 30 min
DAH1	<i>E. globulus</i>	Diluted Acid Hydrolysis	120° C / 0,10% / 1:4 / 20 min
DAH2	<i>E. globulus</i>	Diluted Acid Hydrolysis	170° C / 0,10% / 1:4 / 20 min
DAH3	<i>E. globulus</i>	Diluted Acid Hydrolysis	120° C / 0,75% / 1:4 / 20 min
DAH4	<i>E. globulus</i>	Diluted Acid Hydrolysis	170° C / 0,75% / 1:4 / 20 min
DAH5	<i>E. globulus</i>	Diluted Acid Hydrolysis	120° C / 0,10% / 1:8 / 20 min
DAH6	<i>E. globulus</i>	Diluted Acid Hydrolysis	170° C / 0,10% / 1:8 / 20 min
DAH7	<i>E. globulus</i>	Diluted Acid Hydrolysis	120° C / 0,75% / 1:8 / 20 min
DAH8	<i>E. globulus</i>	Diluted Acid Hydrolysis	170° C / 0,75% / 1:8 / 20 min
DAH9	<i>E. globulus</i>	Diluted Acid Hydrolysis	145° C / 0,43% / 1:6 / 20 min
DAH10	<i>E. globulus</i>	Diluted Acid Hydrolysis	112,7° C / 0,43% / 1:6 / 20 min
DAH11	<i>E. globulus</i>	Diluted Acid Hydrolysis	177,3° C / 0,43% / 1:6 / 20 min
* Autohydrolysis (temperature / reaction time)			
Organosolv (temperature / percentage of sulfuric acid / ethanol:water / min)			
Diluted Acid Hydrolysis (temperature, percentage of sulfuric acid/wood:liquor / min)			

Step to step description of equation 2:

$$(i) \quad \hat{y}_{i,j} = \frac{I_{i,j} * 100}{I_{max,j}}$$

$$(ii) \quad \hat{y}_{i CORRECTED, component j} = 100 * \frac{\hat{y}_{i, component j}}{\sum_{j=1}^n (\hat{y}_{i, n})}$$

Where:

$\hat{y}_{i,j}$  = concentration in pixel i for component j

$I_{i,j}$  = relative intensity in pixel i for component j

$I_{max,j}$  = pixel with the highest relative concentration of component j

By merging (i) and (ii) for two components:

$$\hat{y}_{i,1} = \frac{1}{1 + \frac{I_{i,2} * I_{max,1}}{I_{i,1} * I_{max,2}}}$$

(Equation2)



Figure 1S: Lignin and glucans concentration maps of two different zones of samples Or1, DAH2 and Au3 reconstructed using strategy A.

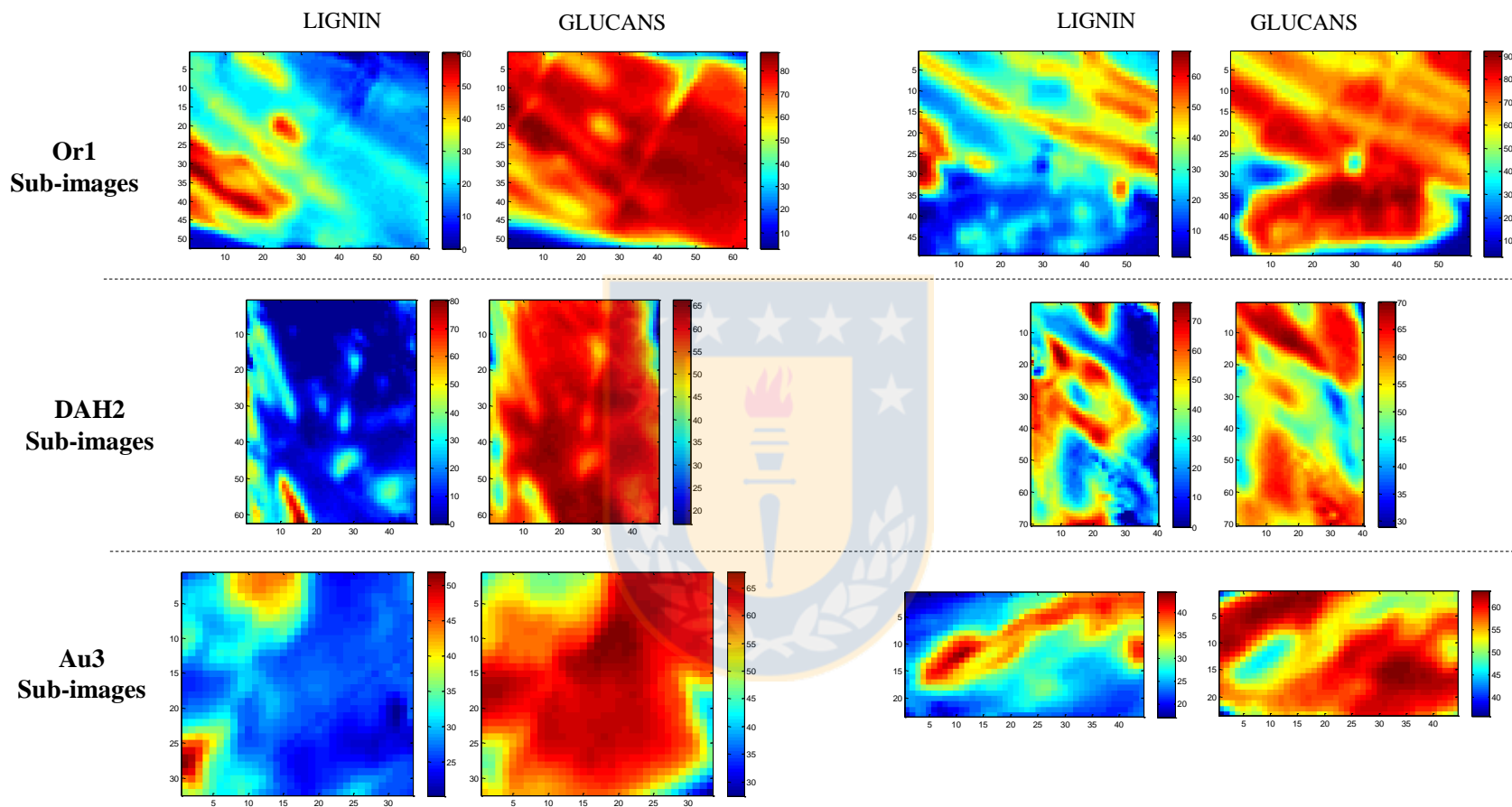
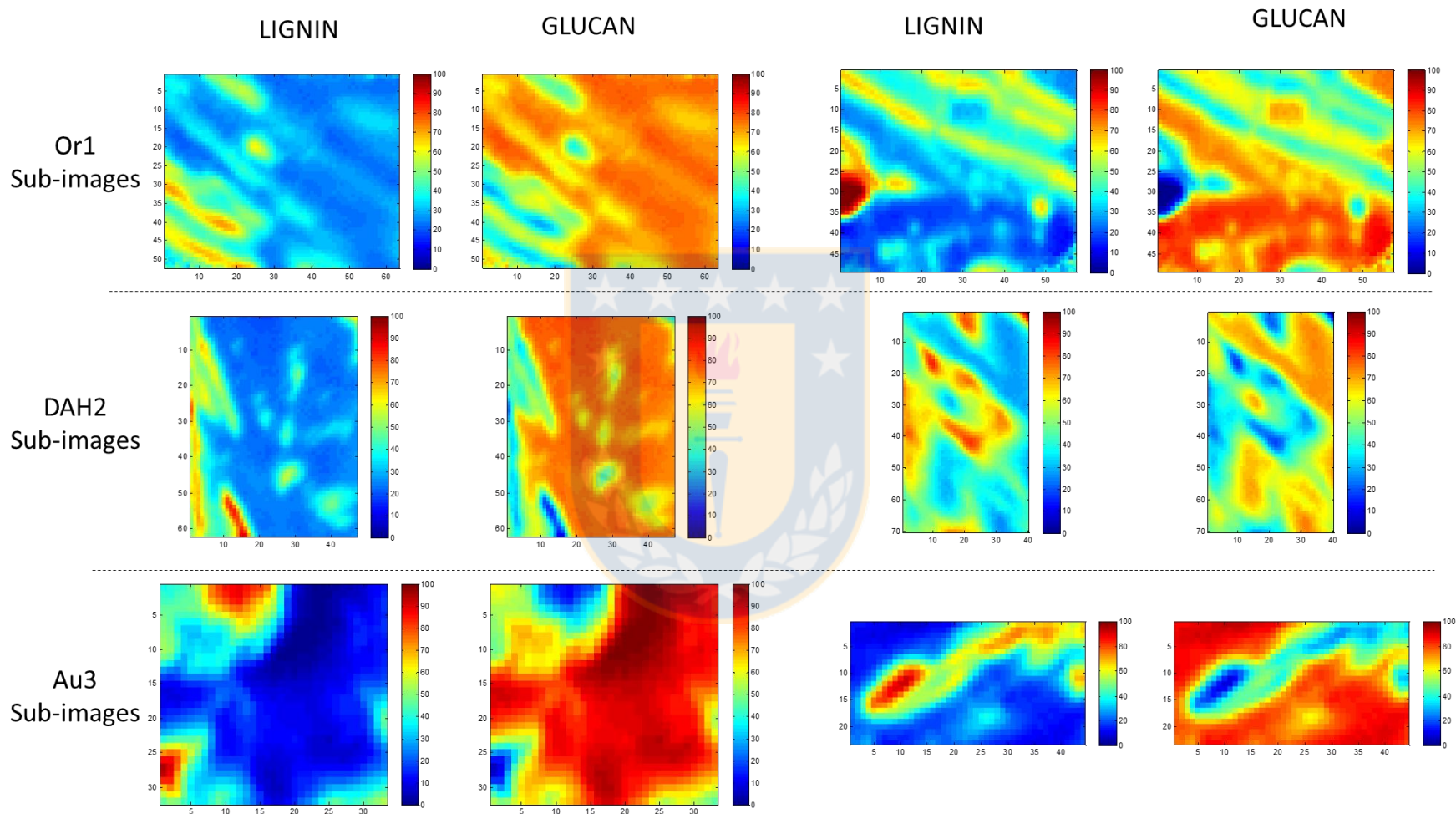


Figure 2S: Lignin and glucans concentration maps of two different zones of samples Or1, DAH2 and Au3 reconstructed using strategy B corrected.





## Sección 2

Manuscrito enviado a la revista *Analytica Chimica Acta* el 28 de Septiembre de 2017.  
Número asignado al manuscrito: ACA-17-2246

### **Fourier transformed infrared imaging and quantitative analysis of pre-treated wood fibres: a comparison between partial least squares (PLS) and multivariate curve resolution with alternating least squares (MCR-ALS) methods**

Juan A. Araya<sup>a,b</sup>, Renato L. Carneiro<sup>c</sup>, José Neira<sup>a</sup>, Juanita Freer<sup>b,d</sup>, Rosario del P. Castillo<sup>a,b\*</sup>

<sup>a</sup> Laboratory of Infrared Spectroscopy and Chemometrics, Faculty of Pharmacy, University of Concepción, Concepción, Chile

<sup>b</sup> Biotechnology Center, University of Concepción, Concepción, Chile

<sup>c</sup> Department of Chemistry, Federal University of São Carlos, São Carlos, SP, Brazil

<sup>d</sup> Faculty of Chemical Sciences, University of Concepción, Concepción, Chile

\* Corresponding author at: Department of Instrumental Analysis, Faculty of Pharmacy, University of Concepcion, Concepcion, Chile. Tel.: +56 412203749. E-mail address: [rosariocastillo@udec.cl](mailto:rosariocastillo@udec.cl) (Rosario del P. Castillo).

E-mail Address of other authors: [jarayaq@udec.cl](mailto:jarayaq@udec.cl) (Juan Araya Q.), [renato.lajarim@ufscar.br](mailto:renato.lajarim@ufscar.br) (Renato L. Carneiro), [ynaiera@udec.cl](mailto:ynaiera@udec.cl) (José Yamil Neira), [jfreer@udec.cl](mailto:jfreer@udec.cl) (Juanita Freer).

Manuscript Number:

Title: Fourier Transform Infrared Imaging and Quantitative Analysis of pre-treated wood fibers: a comparison between Partial Least Squares (PLS) and Multivariate Curve Resolution with Alternating Least Squares (MCR-ALS) methods

Article Type: Full Length Article

Section/Category: CHEMOMETRICS

Keywords: Fourier transform infrared imaging; Partial Least Squares; Multivariate curve resolution with alternating least squares; multivariate image regression; Lignocellulosic fibers.

Corresponding Author: Ms. ROSARIO CASTILLO,

Corresponding Author's Institution: University of Concepcion

First Author: Juan A. Araya

Order of Authors: Juan A. Araya; Renato L. Carneiro; Juanita Freer; José Y. Neira; ROSARIO CASTILLO

Manuscript Region of Origin: CHILE

Suggested Reviewers: Ronei Poppi  
UNICAMP

ronei@iqm.unicamp.br

Recognized researcher in the field of chemical imaging and applied chemometrics

Reinaldo Teófilo

Universidade Federal de Vicosa

rteofilo@gmail.com

Expertise in chemometric analysis of infrared data from lignocellulosic materials, wood chemistry and multivariate calibration

Carlos Esquerre

University College Dublin

carlos.esquerre@ucd.ie

Expertise in chemical imaging, included quantitative applications by chemometric methods.

Sergei Kazarian

Imperial College London

s.kazarian@imperial.ac.uk

Recognized researcher in the field of biospectroscopy, chemical imaging and vibrational spectroscopy. He published diverse articles using infrared chemical imaging at microscopic level.

Concepcion - Chile, November 25<sup>th</sup>, 2017


To the assigned Editor of *Analytica Chimica Acta*

Dear Editor,

Enclosed find our manuscript entitled “Fourier Transform Infrared Imaging and Quantitative Analysis of pre-treated wood fibers: a comparison between Partial Least Squares (PLS) and Multivariate Curve Resolution with Alternating Least Squares (MCR-ALS) methods”. This work proposes the quantification of the main chemical components of lignocellulosic fibers using hyperspectral imaging and two different chemometric methods: MCR-ALS and PLS. We also provide a discussion of the suitability of both multivariate methods in this scenario. Quantitative analysis of hyperspectral images is relatively new, and current literature provides methods applicable mostly to simple artificial samples, using reference material available to create calibration models. Lignocellulosic fibers provide a good study case for complex and heterogeneous samples, when proper reference materials for calibration are not available. We consider our work provides a new and different approach since we address the problems found when working with real matrices from natural origin.

Our work includes hyperspectral imaging and chemometric techniques, which are included in the scope of *Analytica Chimica Acta*. We hope our manuscript meets the requirements to be accepted in this renowned journal.

Sincerely yours,



Rosario del P. Castillo, PhD.  
Associate Professor  
Faculty of Pharmacy  
University of Concepcion, Chile

## \*Highlights

- Quantitative analysis is performed from hyperspectral images of wood fibers
- Discussion of suitability of MCR-ALS and PLS methods in this scenario is provided
- Glucans, lignin and hemicelluloses were quantified in non-destructive way by both methods
- Concentration and distribution of chemical components is analyzed simultaneously



1  
2  
3  
4 **Fourier Transform Infrared Imaging and Quantitative Analysis of pre-**  
5 **treated wood fibers: a comparison between Partial Least Squares (PLS)**  
6  
7 **and Multivariate Curve Resolution with Alternating Least Squares**  
8 **(MCR-ALS) methods**  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18

19 Juan A. Araya<sup>a, b</sup>, Renato L. Carneiro<sup>c</sup>, Juanita Freer<sup>b</sup>, José Y. Neira<sup>a, d</sup>, Rosario del P.  
20 Castillo<sup>a, b\*</sup>  
21  
22  
23  
24  
25

26  
27  
28 <sup>a</sup> Laboratory of Infrared Spectroscopy and Chemometrics, Faculty of Pharmacy, University  
29 of Concepcion, Concepcion, Chile  
30  
31  
32

33  
34 <sup>b</sup> Biotechnology Center, University of Concepción, Concepción, Chile  
35  
36

37 <sup>c</sup> Department of Chemistry, Federal University of São Carlos, São Carlos, SP, Brazil  
38  
39

40 <sup>d</sup> Faculty of Chemical Sciences, University of Concepción, Concepción, Chile  
41  
42  
43  
44  
45

46 \* Corresponding author at: Department of Instrumental Analysis, Faculty of Pharmacy,  
47 University of Concepcion, Concepcion, Chile. Tel.: +56 412203749. E-mail address:  
48 [rosariocastillo@udec.cl](mailto:rosariocastillo@udec.cl) (Rosario del P. Castillo).  
49  
50  
51  
52  
53

54 E-mail Address of other authors: [jarayaq@udec.cl](mailto:jarayaq@udec.cl) (Juan Araya Q.),  
55 [renato.lajarim@ufscar.br](mailto:renato.lajarim@ufscar.br) (Renato L. Carneiro), [yneidera@udec.cl](mailto:yneidera@udec.cl) (José Yamil Neira),  
56  
57 [jfreer@udec.cl](mailto:jfreer@udec.cl) (Juanita Freer).  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Abstract

Pre-treated lignocellulosic fibers were used as a study case to compare two chemometric methods for quantification of chemical components by Fourier-transformed infrared (FT-IR) imaging. Partial least squares (PLS) and multivariate curve resolution with alternating least squares (MCR-ALS) methods were applied to the images for quantification of glucan, lignin and hemicellulose content. The main problem for calibration in samples from natural origin is the unavailability of proper reference material for pixel-to-pixel quantification. Furthermore, chemical components in wood experience changes after different pre-treatment conditions (such as temperature, pressure and acid content); therefore commercially available reference material may not have the same identity of the components present in the sample. Global concentration information of samples, along with their median spectrum was used as an alternative for calibration in this scenario. Results show that both methods provided similar distribution for lignin and hemicellulose concentration maps, but image reconstruction of glucans shows differences between the two methods. PLS models used to quantify pixels in an image was previously validated through the prediction of global concentration of samples, using the median spectrum of different images (RMSEP=1.3% for glucans, 1.0% for lignin and 0.9% for hemicelluloses) . PLS provided a faster method for estimation of components distribution; however, the

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

range of pixel concentration predicted was too narrow due to the lack of a calibration set with a wider concentration pool. MCR-ALS reconstructed images have a more realistic range of concentration for pixels, although in the circumstances present in this study it required an additional step to transform resolved relative concentration profiles into real concentration units.

**Keywords:** Fourier transform infrared imaging; Partial Least Squares; Multivariate curve resolution with alternating least squares; multivariate image regression; Lignocellulosic fibers.



## 1. Introduction

Hyperspectral imaging is a technique that provides chemical qualitative and quantitative information in a spatial context. It produces a three-dimensional array named “hypercube” which consists of two spatial dimensions containing the position coordinates, and a third dimension containing the spectral information of each pixel within the image. It allows identification and localization of the chemical constituents in a sample, depending on the spectral technique used (e.g. mid infrared, Raman spectroscopy). Due to the high amount of information that spectral images contain, it is necessary to use chemometric tools to deplete and extract relevant data [1].

Although the main advantage of hyperspectral imaging is to provide information regarding distribution of the chemical components in a sample, it is also possible to perform quantitative analysis if appropriate reference information is available. Partial least squares regression (PLS) has been successfully used by decades as the chemometric tool of preference to solve a wide variety of analytical problems [2]. Its application in the analysis of hyperspectral images for the quantification of chemical components has been applied in several studies [3-6]. PLS method require calibration, usually using artificial mixtures prepared with pure materials distributed in a convenient concentration range. Most research in PLS analysis of hyperspectral images has been performed on synthetic samples, using reference materials for calibration. When samples from natural origin are analyzed, the calibration step is a challenge, since the concentration range of analytes in the bulk is not representative compared to the concentration range in the pixels. This occurs because the increase in spatial resolution that microscopic tools provide is correlative to the degree of



1  
2  
3  
4 chemical heterogeneity the resultant image may have, especially in samples from natural  
5  
6 sources, which is known as micro-heterogeneity.  
7  
8

9  
10 Multivariate Curve Resolution - Alternating Least Squares (MCR-ALS) is a bilinear  
11  
12 method developed to resolve mixtures when no prior information about the composition of  
13  
14 the sample is available [7-10]. It iterates the alternating least square algorithm, using  
15  
16 coherent chemical constraints, to converge on estimates of the “true” pure response  
17  
18 profiles, which in spectroscopic measurements correspond to the relative intensity  
19  
20 (concentration) and spectra[11]. Calibration and acquisition of concentration maps (pixel to  
21  
22 pixel quantification) is also possible, through the application of a correlation constraint  
23  
24 which uses an internal calibration model, that rescales intensity values into real  
25  
26 concentration units during the iterative optimization[12]. Hence, appropriate calibration  
27  
28 samples are also needed. MCR-ALS have been used along with hyperspectral images to  
29  
30 obtain chemical distribution maps, mainly from matrices with known composition, such as  
31  
32 samples of pharmaceutical origin[13-16], however, quantitative applications in  
33  
34 hyperspectral images of natural matrices have only recently being described[17, 18].  
35  
36  
37  
38  
39  
40  
41

42 Lignocellulosic fibers obtained from pre-treated wood is a good example of samples from  
43  
44 natural origin, with complex chemical composition. Wood pre-treatment is consider a  
45  
46 crucial step for the biotechnological conversion of lignocellulosic material into ethanol and  
47  
48 other industrial products of interest [19-21]. Pre-treatment is needed to overcome the  
49  
50 natural plant cell wall recalcitrance to enzymatic hydrolysis of cellulose, which is inherent  
51  
52 to its rigid and compact structure [22]. Analytical tools, such as microscopic hyperspectral  
53  
54 imaging, deliver chemical and visual information that improves the understanding of the  
55  
56 physical consequences of pre-treatments in lignocellulosic fibers. The spectral  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 characteristics, quantity and distribution of their main components (i.e. cellulose, lignin and  
5  
6 hemicelluloses) may change depending on the conditions applied in each pre-treatment;  
7  
8 therefore, it is difficult to obtain reliable reference materials for calibration. In this  
9  
10 scenario, an alternative to reference values for pixel calibration in hyperspectral images is  
11  
12 the use of the global concentration of a sample if it is properly matched with a mean or  
13  
14 median spectrum from a representative single image[23]. Median spectrum is preferred  
15  
16 over mean spectrum since it is less influenced by extreme anomalous pixels (outliers)  
17  
18 within the image[23].  
19  
20  
21  
22  
23

24 In this work, PLS and MCR-ALS methods are used to create chemical maps and pixel to  
25  
26 pixel quantification of the three principal constituents of pre-treated wood fibers, namely,  
27  
28 glucans, lignin and hemicellulose from microscopic FT-IR images. Limitations and  
29  
30 advantages of each method in this study case are also discussed.  
31  
32  
33  
34  
35  
36  
37

## 38 **2. Material and Methods**

### 39 *2.1 Samples*

40  
41  
42  
43  
44 Lignocellulosic fibers from three different wood pre-treatment techniques were used.  
45  
46 Diluted acid hydrolysis (DAH) and auto-hydrolysis (AH) pre-treatments were performed on  
47  
48 *Eucalyptus globulus* chips, as described by Arévalo *et al.*[24]. Organosolv pre-treatment  
49  
50 was performed on *Pinus radiata* chips, as described by Castillo *et al.*[25]. Conditions in  
51  
52 each pre-treatment were selected using a factorial fractional experimental design to obtain  
53  
54 different concentration and distribution patterns of glucans, lignin and hemicelluloses.  
55  
56  
57  
58  
59 More information about samples are given in section 3.5.  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8 *2.2 Pre-treated Wood Chemical Characterization*  
9

10  
11 Chemical characterization of all pre-treated wood was performed by reference methods,  
12 through acid hydrolysis, followed by HPLC analysis for sugars determination, and UV  
13 spectroscopy and gravimetric analysis for lignin determination, as performed by Castillo *et*  
14 *al.*[25]. As samples are destroyed during chemical characterization, concentrations obtained  
15 correspond to the global concentration value of the bulk sample.  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

26  
27 *2.3 Image acquisition*  
28  
29

30  
31 Fourier transformed infrared (FT-IR) images of pulp fibers were acquired using an FT-IR  
32 Spectrum Frontier/Spotlight 400 Microscopy System (PerkinElmer, Inc.) with a linear array  
33 of 16 mercuric cadmium telluride (MCT) detectors. Isolated fibers previously dried for 8 h  
34 at 45°C were selected for imaging and scanned by attenuated total reflectance (ATR) using  
35 a germanium crystal. A pixel size of 1.56  $\mu\text{m}$  x 1.56  $\mu\text{m}$  and spectral range from 4000 to  
36 748  $\text{cm}^{-1}$  with spectral resolution of 8  $\text{cm}^{-1}$  and 16 scans per spectra were the parameters  
37 selected.  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50

51 *2.4 Data processing*  
52  
53

54 Lignocellulosic fibers have a sinuous and elongated shape; therefore, a region of interest  
55 (ROI) of each image was selected to avoid pixels with no chemical information, such as  
56 background or zones with poor contact with the ATR crystal. Hypercubes of the ROI were  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 later unfolded into two dimensions, pixel and spectra, using Matlab R2013a (Mathworks,  
5  
6 Natick, MA). Spectra were transformed to absorbance, and spectral range between 4000 to  
7  
8 2000 cm<sup>-1</sup> was eliminated from each image to focus the resolution analysis in the  
9  
10 fingerprint region of mid infrared, which contains most of the qualitative spectral  
11  
12 information. After chemometric analysis (prediction of pixel-to-pixel concentration), data  
13  
14 was refolded into its original spatial dimensions in order to create concentration maps for  
15  
16 each component.  
17  
18  
19  
20  
21  
22  
23  
24

### 25 *2.5 PLS models*

26  
27  
28 PLS models were built using the median spectrum of a single image (ROI) of each sample,  
29  
30 and the global concentration of those samples obtained from reference methods. Three  
31  
32 different partial least square regression (PLS) models were performed using Pirouette®  
33  
34 (Infometrix, inc) to predict glucans, lignin and hemicellulose concentrations, respectively.  
35  
36 Samples that did not fit a linear regression in each model were detected by the outlier  
37  
38 diagnostics available in the software and were eliminated. Modelling and predictive ability  
39  
40 of the models was tested using cross validation (leaving one out) and external validation.  
41  
42 External validation set was constructed for each model, using the median spectrum of  
43  
44 different images not included the modelling set. Root mean square error of calibration  
45  
46 (RMSEC), root mean square error of cross validation (RMSECV) and root mean square  
47  
48 error of external prediction (RMSEP) were used as parameters to evaluate the models.  
49  
50  
51  
52  
53  
54  
55

56 For the glucans PLS model, twelve samples of *E. globulus*, from two different pre-  
57  
58 treatments (six of diluted acid hydrolysis and six of auto hydrolysis), were used as the  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 calibration set with a concentration range between 47.1% and 61.1% w/w of glucans  
5 content. Spectral range between 1680  $\text{cm}^{-1}$  to 748  $\text{cm}^{-1}$  was selected, no transformations  
6 were applied and the data was pre-processed by mean center. Six images were used in the  
7 external validation set.  
8  
9

10  
11  
12  
13  
14 For the lignin PLS model nine samples were used for calibration, which includes samples  
15 from two different species (*E. globulus* and *P. radiata*) and three different pre-treatments  
16 (three of Organosolv, three of diluted acid hydrolysis and three of auto hydrolysis), with a  
17 concentration range between 23.8% and 33.0% w/w of lignin content. Spectral range  
18 selected was 1676  $\text{cm}^{-1}$  to 748  $\text{cm}^{-1}$ , data was transformed by smoothing (Savitzky-golay  
19 filter, 11 points), first derivative (Savitzky-golay, 13 points) and normalization, and pre-  
20 processed by mean center. External validation set contained 10 samples.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

32 PLS model for hemicelluloses was built using nine samples (*E. globulus*) from two  
33 different pre-treatments (three of diluted acid hydrolysis and six of auto hydrolysis) with a  
34 concentration range from 1.5% to 13.7% w/w. Spectral range used was 2000  $\text{cm}^{-1}$  to 748  
35  $\text{cm}^{-1}$ . Data was transformed by smoothing (Savitzky-golay filter, 11 points) and first  
36 derivative (Savitzky-golay, 13 points), and pre-processed by mean center. Five images were  
37 used in the external validation set.  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50

## 51 2.6 MCR-ALS

52  
53  
54 The MCR-ALS toolbox[26] for Matlab was used for the analysis. Three components were  
55 selected for the analysis and “PURE” routine (from MCR-ALS toolbox) was used to find  
56 the purest spectra in the dataset, which were used as initial estimative of spectral profile for  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 MCR-ALS. For ALS procedure non- negativity in concentration and spectra were used as  
5  
6 constraints, with 50 iterations and convergence criteria of 0.1%. After MCR-ALS analysis,  
7  
8 values obtained in the concentration profile for a pixel ( $I_i$ ) were converted into real pixel  
9  
10 concentration ( $\hat{y}_i$ ) by assuming a direct relation between the global concentration value of  
11  
12 the sample ( $\bar{y}$ ) and the average spectrum of the image ( $\bar{I}$ ), and recalculating single pixel  
13  
14 concentration afterwards, as described by different authors[17, 18]. This mathematical  
15  
16 transformation is shown in equation (1):  
17  
18  
19  
20

$$\hat{y}_i = \frac{I_i * \bar{y}}{\bar{I}} \quad \dots(1)$$

### 3. Results and Discussion

21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37 Table 1 shows global PLS parameters for the three models constructed. The number of  
38  
39 latent variables used in each model (three in lignin and in hemicelluloses, four in glucans)  
40  
41 were selected based on the decrease of RMSECV, and was consistent with the number of  
42  
43 chemical components expected in lignocellulosic fibers. RMSECV for glucans was 1.8%,  
44  
45 for lignin was 0.3% and for hemicellulose was 1.7%. Global concentrations of external  
46  
47 validation samples were satisfactory predicted by all models with RMSEP of 1.3% for  
48  
49 glucans prediction (n = 6), 1.0% for lignin prediction (n = 10) and 0.9% for hemicelluloses  
50  
51 prediction (n = 5). As the global prediction of concentration was satisfactory, the models  
52  
53 were considered adequate to predict pixel-to-pixel concentrations and refold data into map  
54  
55 concentration of each component.  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 The Figure 1 shows regression plot and regression vectors of PLS models for glucan, lignin  
5 and hemicelluloses. For glucans (Fig. 1A), the most important bands in the regression  
6 vector are consistent with the cellulose spectrum [27-29], except for the presences of  
7 positive bands at  $1516\text{ cm}^{-1}$  and  $1612\text{ cm}^{-1}$ , which would correspond to lignin aromatic  
8 skeletal vibration and C=O stretching [30]. Most prominent bands with positive coefficient  
9 in the vector are  $1312\text{ cm}^{-1}$ ,  $1108\text{ cm}^{-1}$ ,  $1060\text{ cm}^{-1}$ ,  $1166\text{ cm}^{-1}$  that correspond to cellulose  
10 vibrations as shown in table 2. Negative bands at  $1232\text{ cm}^{-1}$  and  $1128\text{ cm}^{-1}$  are related to  
11 lignin[30], while band at  $1040\text{ cm}^{-1}$  is associated with several carbohydrates[31]. The  
12 negative direction of those bands are consistent with what is expected for the distribution of  
13 glucans in the sample, since in natural conditions lignin is found associated with  
14 hemicellulose, and glucans rich zones in the fiber would contain less of these other  
15 components. When the region between  $1480\text{ cm}^{-1}$  to  $1620\text{ cm}^{-1}$ , which is associated to  
16 lignin, was excluded from the modelling set to avoid bias in glucans prediction, position  
17 and intensity of the other variables in a new regression vector remained unchanged, and  
18 reconstruction of images were almost identical using both models. Furthermore, RMSECV  
19 increased compared with the model that included those bands, suggesting that this region  
20 contributes to the prediction of glucans. Therefore, the final PLS model for glucans  
21 included those bands.

22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49 Regression vector of lignin model (see Fig. 1B) show the most important variables with  
50 positive coefficient at  $1124\text{ cm}^{-1}$ ,  $1220\text{ cm}^{-1}$ ,  $1020\text{ cm}^{-1}$ ,  $932\text{ cm}^{-1}$ ,  $1340\text{ cm}^{-1}$ ,  $856\text{ cm}^{-1}$  and  
51  $1480\text{ cm}^{-1}$ , which are reasonably close to the position of lignin typical bands (see table 2).  
52  
53  
54  
55  
56 Important negative bands are at  $1288\text{ cm}^{-1}$ ,  $1088\text{ cm}^{-1}$  (close to lignin C-O deformation in  
57 secondary alcohols and O-H vibration in hemicelluloses), and  $980\text{ cm}^{-1}$ , which is related to  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 carbohydrates. This suggest that some hemicellulose's spectral features are indirectly  
5  
6 increasing the model quality.  
7

8  
9 For hemicellulose's regression vector (Fig. 1C), variables at 1740  $\text{cm}^{-1}$ , 980  $\text{cm}^{-1}$ , 1265  $\text{cm}^{-1}$ ,  
10  
11  $^1$ , 880  $\text{cm}^{-1}$  and 1040  $\text{cm}^{-1}$  are consistent with hemicellulose and carbohydrates bands.  
12  
13 Another positive band at 1420  $\text{cm}^{-1}$  might be related to lignin and cellulose, however, most  
14  
15 of bands with negative coefficient are also close to lignin typical bands (1216  $\text{cm}^{-1}$ , 1508  
16  
17  $\text{cm}^{-1}$ , and 1330  $\text{cm}^{-1}$ ). Other negative bands at 1116  $\text{cm}^{-1}$  and 1059  $\text{cm}^{-1}$  are related to  
18  
19 cellulose.  
20  
21  
22  
23  
24

25 In MCR-ALS analysis, a preliminary principal component analysis was performed to  
26  
27 determine the number of components present in the data matrix. The percentage of  
28  
29 explained variance suggested three principal components. Two typical MCR-ALS resolved  
30  
31 images, from samples AH 2 and sample DAH 4, are depicted in Fig. 2 (D, E and F) and  
32  
33 Fig. 3 (D, E and F), respectively. Parameters of the optimization procedure are summarized  
34  
35 in table 3 and show that experimental data was well fitted, presenting a high percentage of  
36  
37 explained variance. The three components resolved were identified as glucans, lignin and  
38  
39 hemicelluloses, depending on their spectral characteristics. Glucans resolved spectra (see  
40  
41 Fig. 2D and 3D) show characteristic bands and intensities of this component at 1030  $\text{cm}^{-1}$ ,  
42  
43 1052  $\text{cm}^{-1}$ , 1104  $\text{cm}^{-1}$ , 1156  $\text{cm}^{-1}$ , 1316  $\text{cm}^{-1}$ , 1368  $\text{cm}^{-1}$  and 1428  $\text{cm}^{-1}$ . MCR-ALS resolved  
44  
45 spectra for lignin (see Fig. 2E and 3E) also show mostly typical bands for lignin (e.g. 1460  
46  
47  $\text{cm}^{-1}$ , 1512  $\text{cm}^{-1}$ , 1604  $\text{cm}^{-1}$ ), although some carbohydrates bands are also present. The third  
48  
49 resolved component frequently displayed features of hemicellulose spectrum, with the  
50  
51 presence of typical bands of other carbohydrates and lignin (see Fig. 2F and 3F). This may  
52  
53 evidence the presence of lignin-carbohydrate complexes on the surface of the fiber.  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



1  
2  
3  
4 Covalent bonds between lignin and hemicelluloses are naturally formed as the plant cell  
5 wall is lignified[32], then, the presence of this complex suggest that severity of the pre-  
6 treatment did not entirely disrupt lignin-carbohydrates chemical association[25, 33]. During  
7 this work, this component would be appointed as hemicellulose to facilitate comparison  
8 with PLS predicted concentration map for hemicellulose.  
9  
10  
11  
12  
13  
14  
15  
16

17 Fig. 2 and 3 also show concentration maps for glucans (A), lignin (B) and hemicelluloses  
18 (C) predicted by PLS models. Distribution of lignin and hemicellulose is similar in maps  
19 predicted by PLS (Fig. 2B and 2C; and Fig. 3B and 3C) and maps obtained by MCR-ALS  
20 (Fig. 2E and 2F; and Fig. 3E and 3F).  
21  
22  
23  
24  
25  
26

27 It is worth noting that lignin and hemicellulose concentration maps predicted by PLS  
28 models are very similar in distribution (as expected), despite the fact that lignin  
29 concentration of samples used to build the model increases while hemicellulose  
30 concentration decreases, and vice versa (see Fig 1). This is coherent with the distribution  
31 of these components in the cell plant, and provides confidence to the performance of the  
32 predictions. Concentration maps of lignin and hemicellulose reconstructed by MCR-ALS  
33 also have a similar distribution.  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44

45 Glucans distribution maps show evident differences between PLS predicted images (Fig.  
46 2A and 3A) and MCR-ALS reconstructed image (Fig. 2D and 3D). Hot and cold zones in  
47 both images do not seem to match, even when glucans regression vector in the PLS and  
48 MCR-ALS pure spectra assigned to glucans show prevalent glucans typical bands. The  
49 variability in spatial distribution of pixel concentration in an image can be defined as the  
50 *distributional heterogeneity* of an image[13]. Thus, concentration maps of lignin and  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 hemicellulose obtained by PLS and MCR-ALS methods have a similar *distributional*  
5  
6 *heterogeneity*, while glucan concentration maps are more heterogeneous.  
7  
8

9  
10 Besides glucan distribution in concentration maps, the most substantial difference between  
11  
12 PLS and MCR-ALS models is the range of concentration pixels are depicted. PLS  
13  
14 predicted maps for glucans and lignin have a very narrow range of concentration compared  
15  
16 to maps predicted by MCR-ALS, whose ranges are considerably broader. This difference is  
17  
18 notorious in the concentration bars of images in Fig. 2 and 3; however, to highlight this  
19  
20 difference, Fig. 4 and 5 show the histograms of pixel concentration of the same images  
21  
22 shown in Fig. 2 and 3, respectively. Concentration axis of the histogram had the same range  
23  
24 for PLS and MCR-ALS images of the same component, to help comparison. The  
25  
26 distribution of concentration values around the mean value of an image (in other words,  
27  
28 how broad is the concentration range) can be defined as the *constitutional*  
29  
30 *heterogeneity*[13] of a sample. Hence, glucans and lignin concentration maps obtained by  
31  
32 MCR-ALS have a higher *constitutional heterogeneity* than the ones obtained by PLS. A  
33  
34 broad concentration range is more realistic, since the spatial resolution of microscopic  
35  
36 images would allow to find at least a few pixels with low or high content of any  
37  
38 component.  
39  
40  
41  
42  
43  
44  
45  
46

47 The narrow concentration range of images predicted by PLS could be explained by the little  
48  
49 difference of concentration bulk samples had. As no reference material was available, PLS  
50  
51 models were constructed using the information of the global concentration of the  
52  
53 lignocellulosic fibers. As can be seen in Figure 1, the range of concentration those samples  
54  
55 had was not wide. Therefore, concentration predicted in pixels using a short range of  
56  
57 concentration will be restricted to the dimensions in which the model was calibrated, that in  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 this case is far from optimal. If samples in a wide concentration range are available, PLS  
5  
6 prediction model could produce a more realistic concentration map. MCR-ALS approach  
7  
8 does not have this problem. On the other hand, in the histogram of lignin concentration  
9  
10 obtained by MCR-ALS shown in Fig. 5, some values are above 100%. This is only seen in  
11  
12 a small number of pixels, and could be explained by an error in the reference methods used  
13  
14 to quantify lignin in the original bulk sample.  
15  
16  
17  
18

19 Another factor to take into account is the complexity of the methods. In this work, when  
20  
21 MCR-ALS concentration profiles were resolved, the relative concentration values obtained  
22  
23 were escalated into real concentration units afterwards, in a subsequent step. This process  
24  
25 require more time of analysis, compared to the fast PLS prediction of pixel concentration,  
26  
27 and each image has to be resolved individually. Although several images can be  
28  
29 concatenated and resolved altogether, this would restrict the size of the image (amount of  
30  
31 pixels) or the operating capacity of the computer used, and still re escalation of the  
32  
33 concentration values should be applied to each image individually afterwards. To avoid this  
34  
35 subsequent calibration step, other strategies could be applied, such as correlation constraint.  
36  
37  
38  
39  
40  
41 Correlation constraint is a MCR-ALS constraint that allows to calibrate real concentration  
42  
43 values at the same time the algorithm resolves the data. Is the fundamental quantification  
44  
45 method when working with MCR-ALS[12, 34-37], however, it requires the construction of  
46  
47 an internal calibration model using reference information, therefore, in this case, we would  
48  
49 have the same limitations regarding concentration range we had with the PLS models.  
50  
51  
52 Furthermore, when an attempt to build the internal calibration model for correlation  
53  
54 constrained using the reference information available was made (as done with the PLS  
55  
56 model), results were not satisfactory (data not shown).  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7 **4. Conclusions**  
8  
9

10  
11  
12  
13 PLS and MCR-ALS models were performed on ATR FT-IR hyperspectral images of  
14 lignocellulosic fibers to obtain concentration maps for glucans, lignin and hemicellulose,  
15 which are the main components of wood cells. For lignin and hemicelluloses, relative  
16 distribution of concentration maps was similar using both methods, however relative  
17 distribution for glucans showed differences. PLS models are easier and faster to apply,  
18 regression vectors were consistent with spectral characteristics of the each component  
19 calibrated and global concentration prediction using the median spectrum of several images  
20 was satisfactory. However, concentration range used to calibrate the model was very short,  
21 as they were limited to the global concentration of the lignocellulosic fibers for each  
22 component. Therefore, concentration range of maps predicted was too narrow, discordant  
23 with what is expected in natural lignocellulosic fibers. If a more adequate calibration set  
24 was available for calibration, PLS could potentially be a more suitable method for  
25 prediction of concentration maps in lignocellulosic fibers. On the contrary, MCR-ALS  
26 method requires more time of analysis and provides concentration maps that produced  
27 pixels in all the range of concentration possible, which is more consistent with reality.  
28 Consequently, MCR-ALS could be more reliable when appropriate calibration samples are  
29 not available.  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58

59 **5. Acknowledgements**  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 Authors would like to thank “Centro de Microscopía Avanzada CMA Bío-Bío” of the  
5  
6 University of Concepción, Chile.  
7  
8

9  
10 Funding: This work was supported by ENLACE UDEC [217.071.065.1-0] Project. Juan  
11  
12 Araya thanks to CONICYT-PCHA/Doctorado Nacional scholarship [Folio 2013-  
13  
14 21130641], Chile.  
15  
16  
17  
18  
19  
20

## 21 6. References

22  
23

- 24 1. Fackler, K. and L.G. Thygesen, *Microspectroscopy as applied to the study of wood*  
25 *molecular structure*. Wood science and technology, 2013. **47**(1): p. 203-222.
- 26 2. Brereton, R.G., *Applied chemometrics for scientists*. 2007: John Wiley & Sons.
- 27 3. Reich, G., *Near-infrared spectroscopy and imaging: basic principles and pharmaceutical*  
28 *applications*. Advanced drug delivery reviews, 2005. **57**(8): p. 1109-1143.
- 29 4. Gowen, A.A., et al., *Hyperspectral imaging – an emerging process analytical tool for food*  
30 *quality and safety control*. Trends in Food Science & Technology, 2007. **18**(12): p. 590-  
31 598.
- 32 5. Schönbichler, S., et al., *Comparison of NIR chemical imaging with conventional NIR,*  
33 *Raman and ATR-IR spectroscopy for quantification of furosemide crystal polymorphs in*  
34 *ternary powder mixtures*. European Journal of Pharmaceutics and Biopharmaceutics, 2013.  
35 **84**(3): p. 616-625.
- 36 6. Burger, J. and P. Geladi, *Hyperspectral NIR imaging for calibration and prediction: a*  
37 *comparison between image and spectrometer data for studying organic and biological*  
38 *samples*. Analyst, 2006. **131**(10): p. 1152-1160.
- 39 7. Tauler, R., *Multivariate curve resolution applied to second order data*. Chemometrics and  
40 intelligent laboratory systems, 1995. **30**(1): p. 133-146.
- 41 8. de Juan, A. and R. Tauler, *Multivariate curve resolution (MCR) from 2000: progress in*  
42 *concepts and applications*. Critical Reviews in Analytical Chemistry, 2006. **36**(3-4): p. 163-  
43 176.
- 44 9. Tauler, R., A. Smilde, and B. Kowalski, *Selectivity, local rank, three- way data analysis*  
45 *and ambiguity in multivariate curve resolution*. Journal of Chemometrics, 1995. **9**(1): p. 31-  
46 58.
- 47 10. Budevska, B.O., S.T. Sum, and T.J. Jones, *Application of multivariate curve resolution for*  
48 *analysis of FT-IR microspectroscopic images of in situ plant tissue*. Applied spectroscopy,  
49 2003. **57**(2): p. 124-131.
- 50 11. Jaumot, J., et al., *A graphical user-friendly interface for MCR-ALS: a new tool for*  
51 *multivariate curve resolution in MATLAB*. Chemometrics and Intelligent Laboratory  
52 Systems, 2005. **76**(1): p. 101-110.
- 53 12. Jaumot, J., A. de Juan, and R. Tauler, *MCR-ALS GUI 2.0: new features and applications*.  
54 Chemometrics and Intelligent Laboratory Systems, 2015. **140**: p. 1-12.  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

13. Piqueras, S., et al., *Relevant aspects of quantification and sample heterogeneity in hyperspectral image resolution*. Chemometrics and Intelligent Laboratory Systems, 2012. **117**: p. 169-182.
14. Carneiro, R.L. and R.J. Poppi, *Homogeneity study of ointment dosage forms by infrared imaging spectroscopy*. Journal of Pharmaceutical and Biomedical Analysis, 2012. **58**: p. 42-48.
15. Cruz, J., et al., *Nir-chemical imaging study of acetylsalicylic acid in commercial tablets*. Talanta, 2009. **80**(2): p. 473-478.
16. Vajna, B., et al., *Comparison of chemometric methods in the analysis of pharmaceuticals with hyperspectral Raman imaging*. Journal of Raman Spectroscopy, 2011. **42**(11): p. 1977-1986.
17. Araya, J.A., et al., *Single pixel quantification strategies using middle infrared hyperspectral imaging of lignocellulosic fibers and MCR-ALS analysis*. Microchemical Journal, 2017.
18. Colares, C.J., et al., *Near infrared hyperspectral imaging and MCR-ALS applied for mapping chemical composition of the wood specie Swietenia Macrophylla King (Mahogany) at microscopic level*. Microchemical Journal, 2016. **124**: p. 356-363.
19. Alvira, P., et al., *Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review*. Bioresource technology, 2010. **101**(13): p. 4851-4861.
20. Wyman, C., *Key features of pretreated lignocelluloses biomass solids and their impact on hydrolysis*. Bioalcohol production: biochemical conversion of lignocellulosic biomass, 2010: p. 73.
21. Wyman, C.E., *What is (and is not) vital to advancing cellulosic ethanol*. TRENDS in Biotechnology, 2007. **25**(4): p. 153-157.
22. Zhao, X., L. Zhang, and D. Liu, *Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose*. Biofuels, Bioproducts and Biorefining, 2012. **6**(4): p. 465-482.
23. Burger, J. and P. Geladi, *Hyperspectral NIR image regression part II: dataset preprocessing diagnostics*. Journal of Chemometrics, 2006. **20**(3- 4): p. 106-119.
24. Arévalo, C., et al., *Study of the Ultrastructure of Eucalyptus globulus Wood Substrates Subjected to Auto-Hydrolysis and Diluted Acid Hydrolysis Pre-treatments and Its Influence on Enzymatic Hydrolysis*. BioEnergy Research, 2017: p. 1-14.
25. Castillo, R.d.P., et al., *Fourier transform infrared imaging and microscopy studies of Pinus radiata pulps regarding the simultaneous saccharification and fermentation process*. Analytica chimica acta, 2015. **866**: p. 10-20.
26. De Juan, A. and R. Tauler, *Chemometrics applied to unravel multicomponent processes and mixtures: Revisiting latest trends in multivariate resolution*. Analytica Chimica Acta, 2003. **500**(1): p. 195-210.
27. Fackler, K., et al., *Localisation and characterisation of incipient brown-rot decay within spruce wood cell walls using FT-IR imaging microscopy*. Enzyme and Microbial Technology, 2010. **47**(6): p. 257-267.
28. Pandey, K. and K. Theagarajan, *Analysis of wood surfaces and ground wood by diffuse reflectance (DRIFT) and photoacoustic (PAS) Fourier transform infrared spectroscopic techniques*. European Journal of Wood and Wood Products, 1997. **55**(6): p. 383-390.
29. Pandey, K., *A study of chemical structure of soft and hardwood and wood polymers by FTIR spectroscopy*. Journal of Applied Polymer Science, 1999. **71**(12): p. 1969-1975.
30. Faix, O., *Classification of lignins from different botanical origins by FT-IR spectroscopy*. Holzforschung-International Journal of the Biology, Chemistry, Physics and Technology of Wood, 1991. **45**(s1): p. 21-28.
31. Kacurakova, M., et al., *FT-IR study of plant cell wall model compounds: pectic polysaccharides and hemicelluloses*. Carbohydrate polymers, 2000. **43**(2): p. 195-203.

- 1  
2  
3  
4 32. Trajano, H.L., et al., *The fate of lignin during hydrothermal pretreatment*. Biotechnology for biofuels, 2013. **6**(1): p. 110.  
5  
6 33. Arévalo, C., et al., *Study of the Ultrastructure of Eucalyptus globulus Wood Substrates Subjected to Auto-Hydrolysis and Diluted Acid Hydrolysis Pre-treatments and Its Influence on Enzymatic Hydrolysis*. BioEnergy Research, 2017: p. 1-14.  
7  
8  
9 34. Lyndgaard, L.B., F. van den Berg, and A. de Juan, *Quantification of paracetamol through tablet blister packages by Raman spectroscopy and multivariate curve resolution-alternating least squares*. Chemometrics and Intelligent Laboratory Systems, 2013. **125**: p. 58-66.  
10  
11  
12  
13 35. Goicoechea, H.C., A.C. Olivieri, and R. Tauler, *Application of the correlation constrained multivariate curve resolution alternating least-squares method for analyte quantitation in the presence of unexpected interferences using first-order instrumental data*. Analyst, 2010. **135**(3): p. 636-642.  
14  
15  
16 36. Antunes, M., et al., *Multivariate curve resolution of overlapping voltammetric peaks: quantitative analysis of binary and quaternary metal mixtures*. Analyst, 2002. **127**(6): p. 809-817.  
17  
18  
19  
20  
21 37. Azzouz, T. and R. Tauler, *Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples*. Talanta, 2008. **74**(5): p. 1201-1210.  
22  
23  
24  
25



**Table 1:** PLS parameters for prediction of glucans, lignin and hemicellulose content in pretreated wood. Models were validated by cross validation and by an external validation set, through prediction of global concentration of the images. n= number of images; L.V.= number of latent variables; RMSEC= root mean square error of calibration; RMSECV= root mean square error of cross validation; rCal= correlation coefficient of calibration; rCV= correlation coefficient of cross validation; RMSEP= root mean square error of prediction of “n” external validation images.

Y variable	Calibration						Prediction			
	n	Range %	L.V.	RMSEC	RMSEC V	rCal	rCV	n	RMSE P (median)	r
Glucans	12	47.1 – 61.1	4	1.1	1.8	0.9827	0.9131	6	1.3	0.9622
Lignin	9	23.8 – 33.0	3	1.0	0.3	0.9997	0.9968	10	1.0	0.9555
Hemicellulose	9	1.5 – 13.7	3	1.0	1.7	0.9823	0.8837	5	0.9	0.9854



**Table 2:** Assignment of IR bands of different components in lignocellulosic material observed in this study, according to literature<sup>27-30</sup>

Wavenumber (cm <sup>-1</sup> )	Assignment
856	C-H out of plane in position 2,5, and 6 of guaiacyl units
880	Glucose ring stretch, C <sub>1</sub> -H deformation
932	C-H out of plane; aromatic
980	C-O stretching in cellulose and hemicellulose
1020-1030	C-O of primary alcohol. Guaiacyl C-H
1040	C-O stretching
1052-1060	C-O stretching
1078	O-H
1088	C-O deformation in secondary alcohols and aliphatic ethers
1104-1108	glucose ring stretch (asymmetric)
1116-1128	C-H (guaiacyl and syringyl)
1156-1166	C-O-C asymmetric vibration
1216-1220	C-O of guaiacyl ring
1232	C-C plus C-O plus C-O stretching
1265	C-O stretching in lignin, acetyl and carboxylic vibration in xylans
1312-1316	O-H in plane bending of alcohol groups
1330-1340	S ring plus G ring substituted in position 5 in lignin, O-H in plane deformation in cellulose
1368	C-H symmetric deformation
1420-1430	Aromatic skeletal vibration combined with C-H in plane deformation in lignin, C-H asymmetric deformation in cellulose
1460-1480	C-H deformation; asym. in-CH <sub>3</sub> and -CH <sub>2</sub> -
1508-1516	Aromatic skeletal vibration
1604-1612	Aromatic skeletal vibration plus C=O stretching
1640	water
1740	C=O stretching vibration

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

**Table 3:** MCR-ALS parameters of the optimization procedure for hyperspectral images of samples AH 2 and DAH 4

Samples	AH 2	DAH 4
n° iterations for optimal fitting	23	26
SD of the residuals respect experimental data	0.00034107	0.0016228
Lack of fit (%) respect to PCA data	0.89389	1.3569
Lack of fit (%) respect to experimental data	2.5113	2.8757
Percent of variance explained	99.9369	99.9173



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## List of Figures

**Figure 1:** PLS Regression Vectors (right) and regression lines of the calibration data set (left) for glucans (A), lignin (B) and hemicellulose (C).

**Figure 2:** Image reconstruction of sample AH 2. Concentration maps predicted by PLS models of glucans (A), lignin (B) and hemicellulose (C); and concentration maps and resolved spectra of glucans (D), lignin (E) and hemicellulose (F) reconstructed by MCR-ALS.

**Figure 3:** Image reconstruction of sample DAH 4. Concentration maps predicted by PLS models of glucans (A), lignin (B) and hemicellulose (C); and concentration maps and resolved spectra of glucans (D), lignin (E) and hemicellulose (F) reconstructed by MCR-ALS.

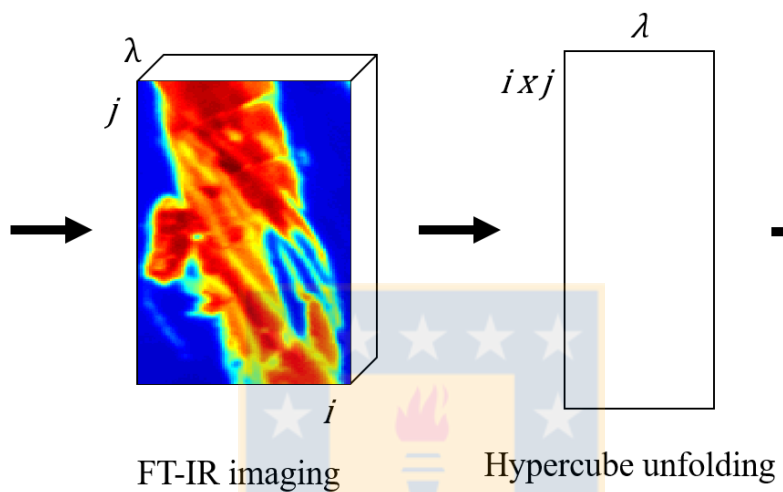
**Figure 4:** Histogram of concentration maps of glucans, lignin and hemicellulose obtained by PLS and MCR-ALS methods in sample AH 2.

**Figure 5:** Histogram of concentration maps of glucans, lignin and hemicellulose obtained by PLS and MCR-ALS methods in sample DAH 4.

Graphical Abstract



Pre-treated wood



**PLS vs MCR-ALS  
strategies for  
quantification of:**

- Glucans
- Lignin
- Hemicellulose

Figure 1

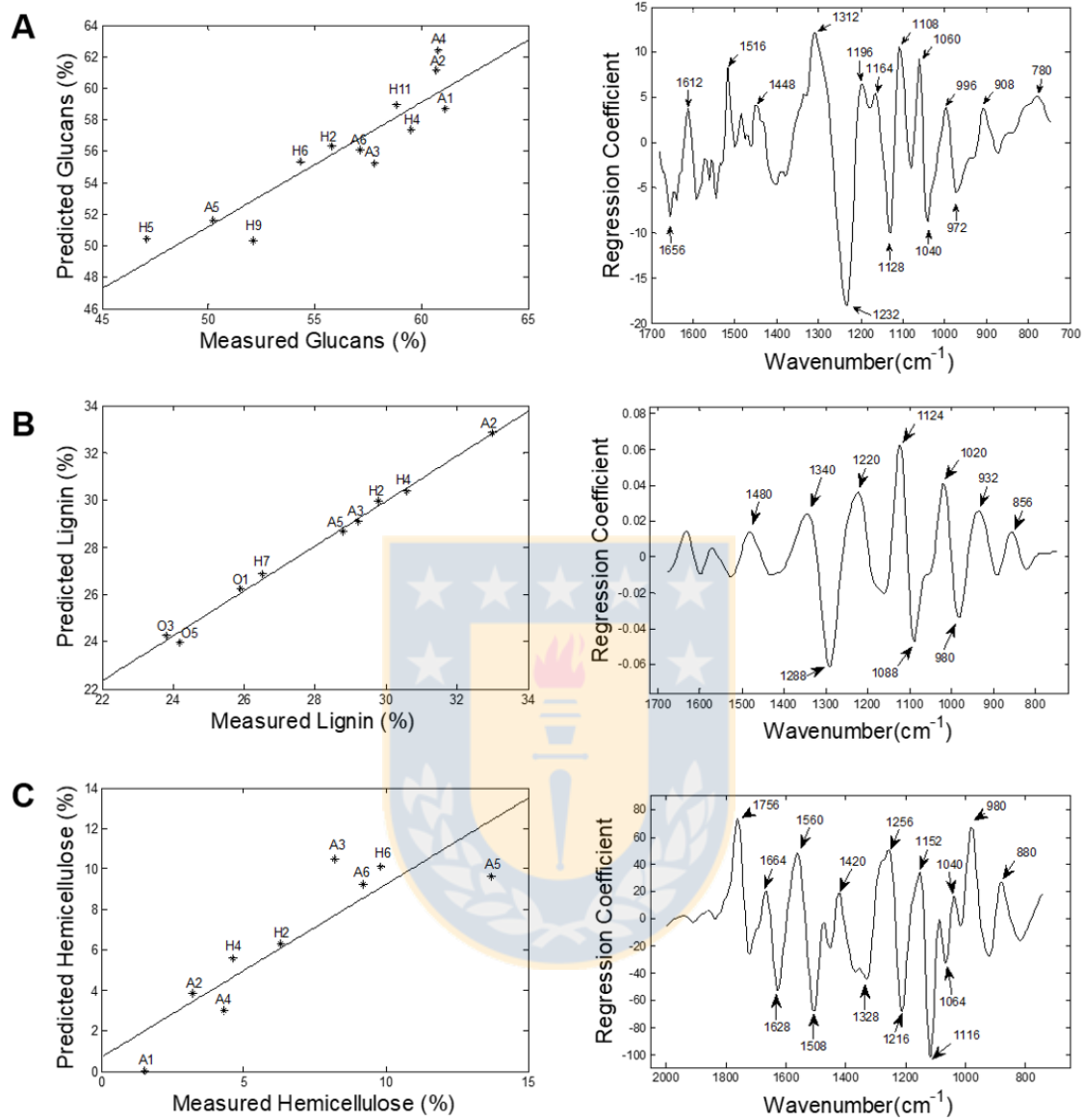


Figure 2

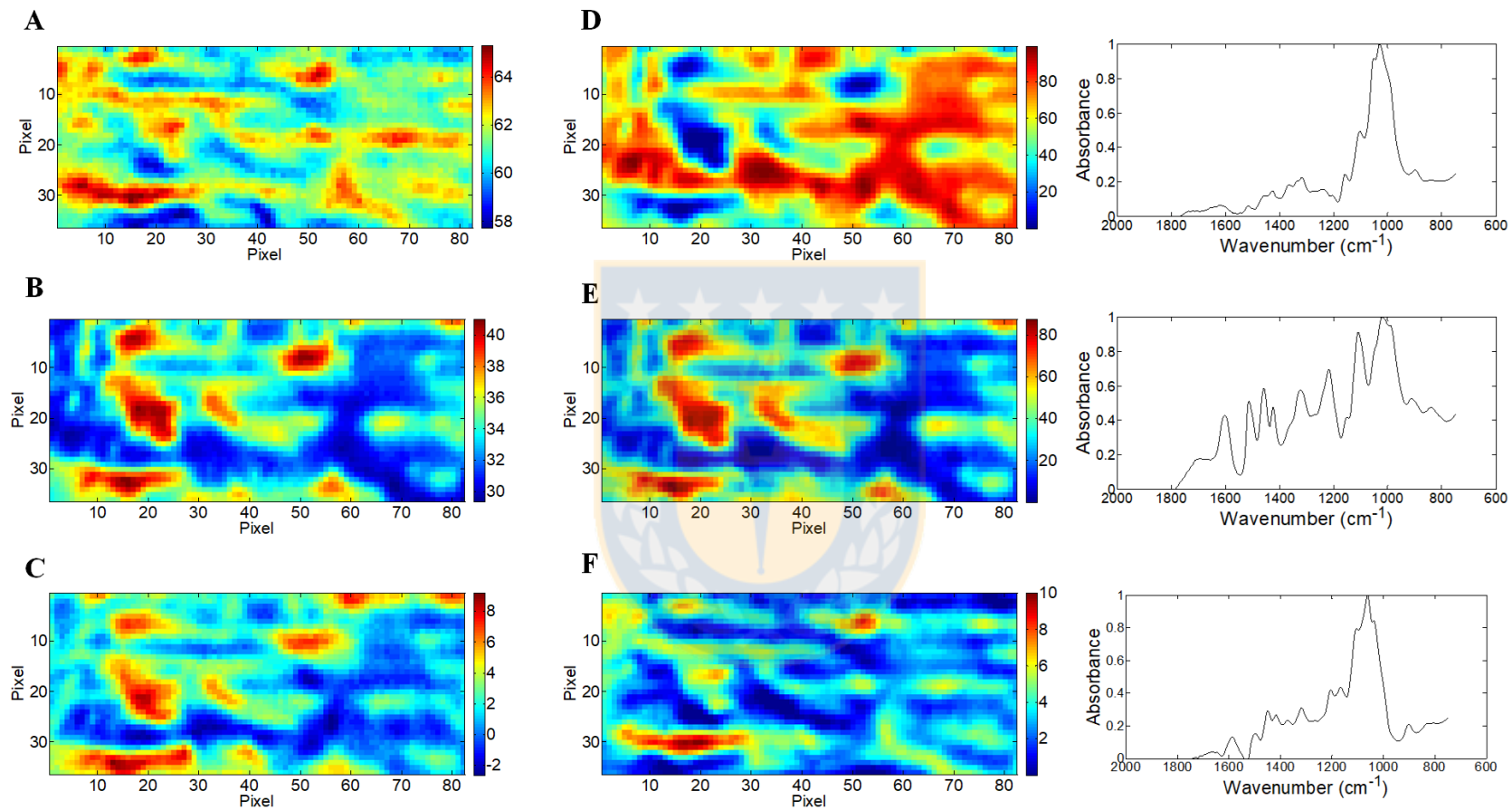


Figure 3

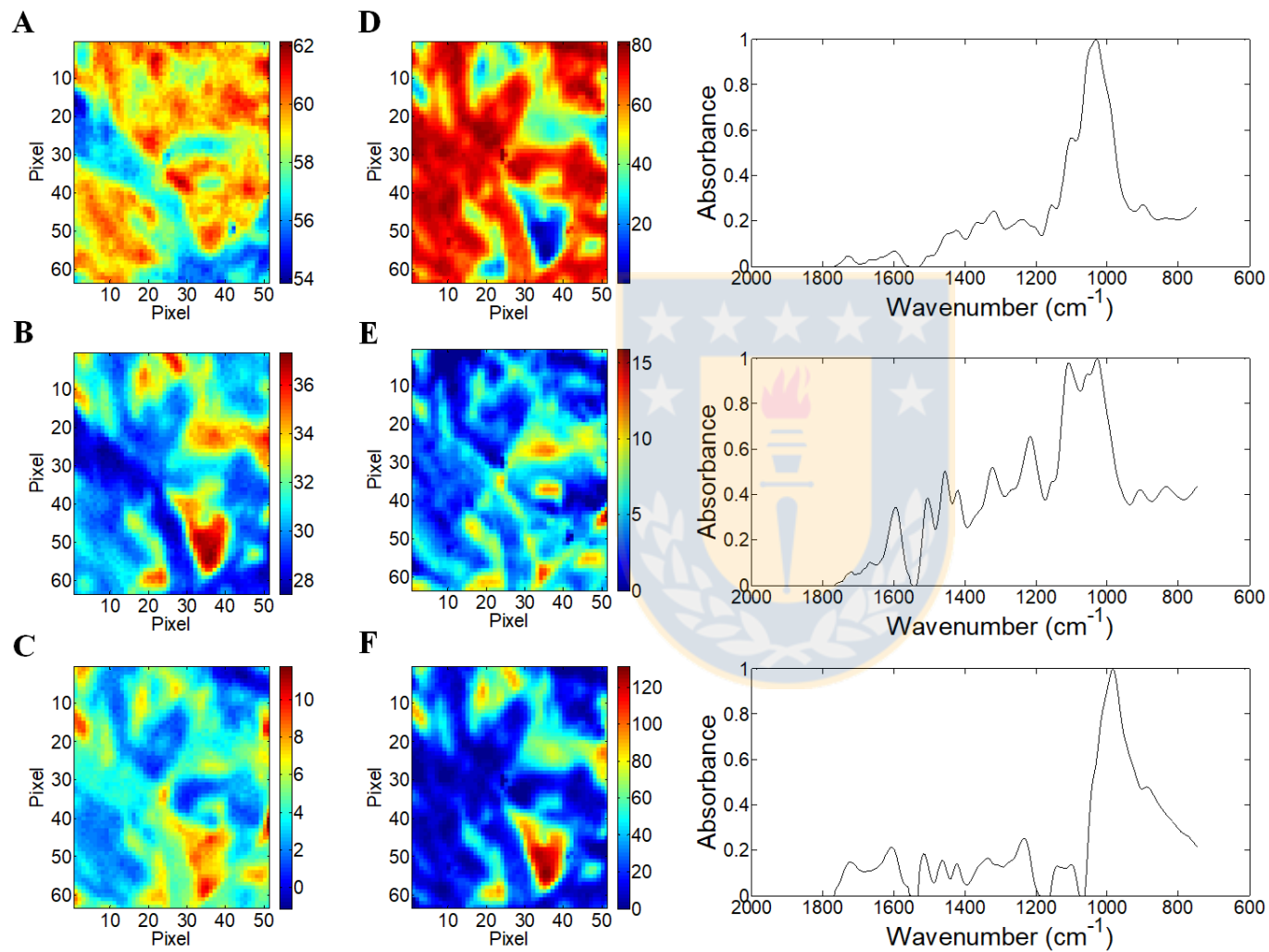


Figure 4

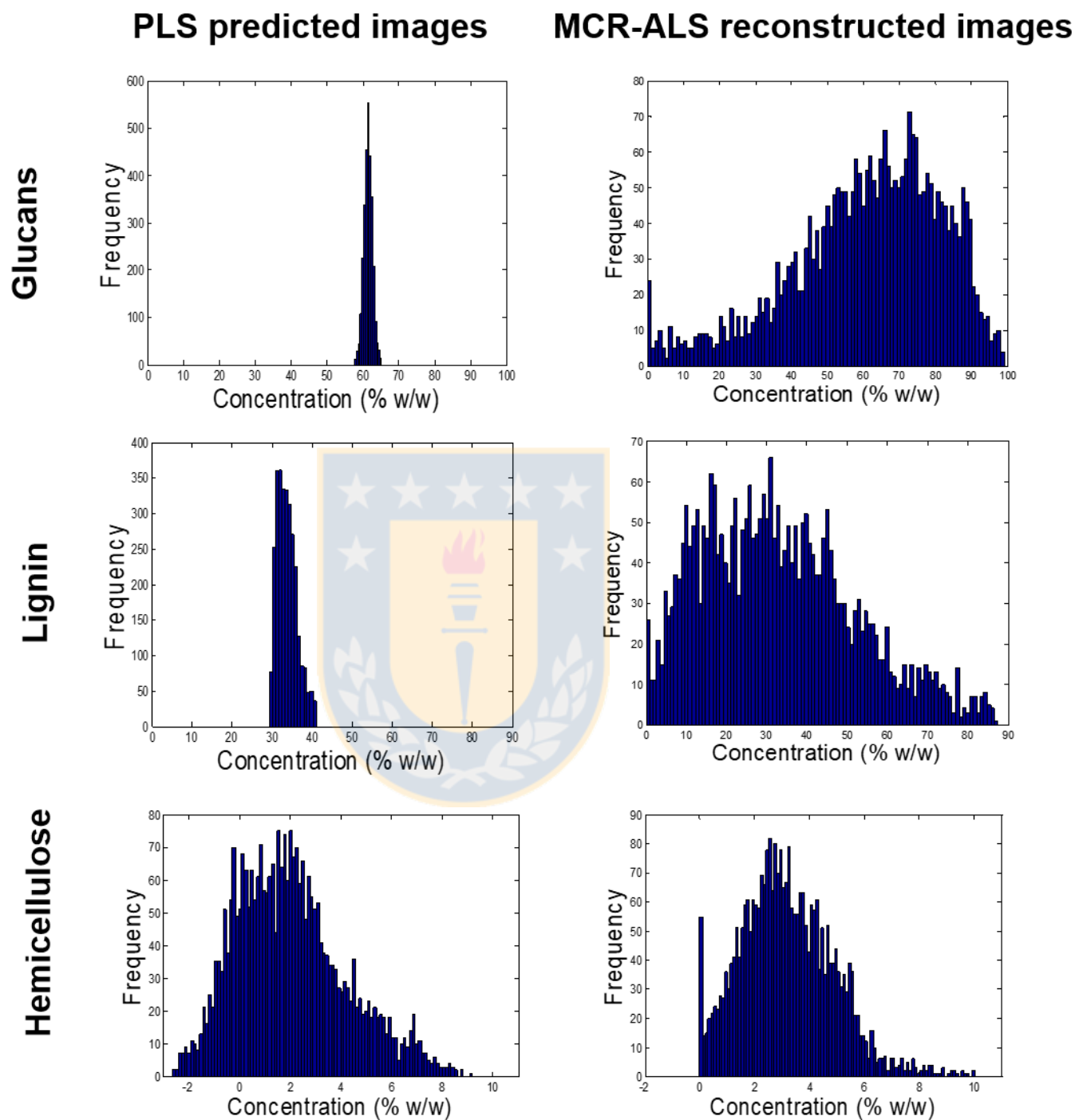
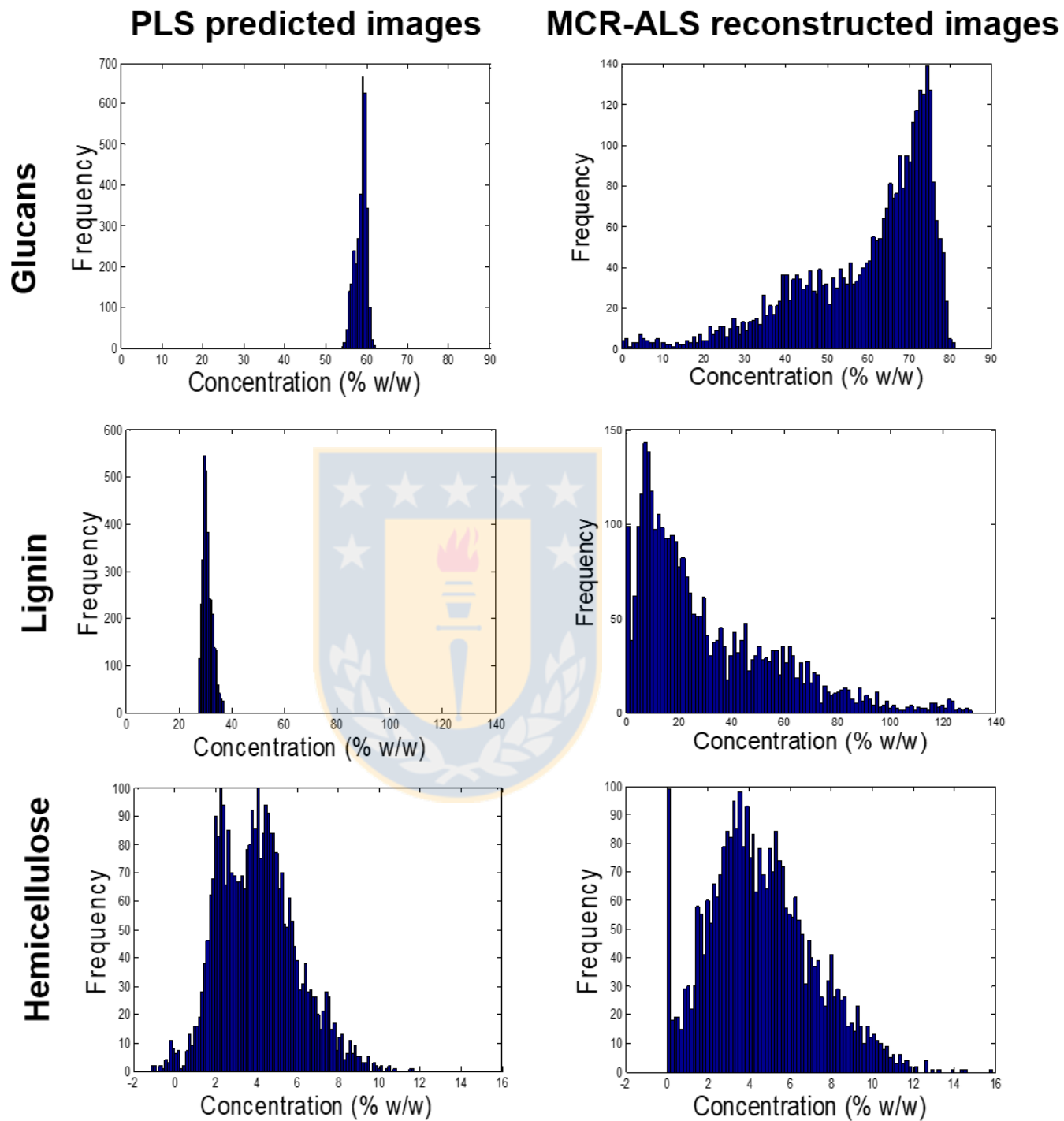




Figure 5



# Capítulo 5

## Conclusiones Finales



1. Las diferentes técnicas microscópicas utilizadas en este trabajo proporcionan información diferente y complementaria. SEM tiene la mayor resolución espacial (lateral) y permite visualizar la superficie de la fibra con alto nivel de detalle, pero no entrega ningún tipo de información química. CLSM permite visualizar a través de la fluorescencia total la fibra a diferentes planos focales (resolución espacial lateral y confocal) y permite reconstrucciones 3-D de ésta. CLSM espectral (usando espectros de emisión de fluorescencia) mejora la selectividad de la técnica, permitiendo la identificación de celulosa y lignina, aunque de ningún otro componente. Tampoco es capaz de detectar diferentes tipos de lignina. La microscopía FT-IR proporciona la mayor cantidad de información cualitativa, permitiendo identificar diferencias químicas entre un mismo macro-componente, sin embargo tiene la menor resolución espacial de todas las técnicas usadas y está limitada por el contenido de humedad de la muestra.
2. CLSM espectral de emisión de fluorescencia permite detectar (además de lignina) la auto-fluorescencia en zonas de celulosa, un fenómeno que, aunque controversial, fue demostrado en el presente trabajo. Tradicionalmente, los estudios realizados en material lignocelulósico mediante CLSM utilizan sondas o tinciones para la detección de celulosa, por lo que la aplicación de la estrategia de análisis aquí desarrollada supone una mejora en esta disciplina respecto de las metodologías existentes en la actualidad.
3. La microscopía FT-IR acoplado análisis por MCR-ALS permite construir verdaderos mapas químicos de compuestos puros, sin siquiera contar con material de referencia para ello.
4. De los métodos quimiométricos usados en esta tesis en conjunto con imágenes espectrales se puede concluir que la complejidad química de la matriz estudiada hacen que sea necesario herramientas adicionales al PCA para su descripción. MCR-ALS fue el método que mostró mayor versatilidad. Permite la descripción química (identificación) de los componentes químicos principales encontrados en madera con un alto grado de detalle (selectividad). Además, mediante MCR-ALS se desarrollaron diferentes estrategias para la cuantificación de imágenes espectrales, las que incluso se ajustan a situaciones en que se cuenta con muy poca información de referencia

para la calibración. PLS también permite la cuantificación de los analitos en una imagen espectral, permitiendo de paso describir su localización, sin embargo es fundamental tener materiales de referencia adecuados (en todo el rango de concentración posible) para la calibración.

5. Mediante el análisis quimiométrico de imágenes espectrales realizado en este trabajo es efectivamente posible determinar la identidad, cantidad y localización de los principales componentes químicos presentes en madera (celulosa, lignina y hemicelulosas) a nivel microscópico. La implementación de esta estrategia analítica proporciona además una herramienta nueva para el estudio de la accesibilidad enzimática a la superficie de la fibra lignocelulósica.
6. De la recalcitrancia de la fibra lignocelulósica se puede concluir que no es necesariamente el contenido total de lignina lo que afecta la accesibilidad enzimática, sino que la distribución en que ésta se encuentra en la fibra. Por ejemplo, la relocalización de lignina producto del pre-tratamiento en esferas (coalescencia) o cúmulos amorfos, a diferentes niveles de la pared celular, deja libres zonas ricas en celulosa, lo que aumenta el rendimiento de hidrólisis enzimática.
7. Los distintos pre-tratamientos usados en este trabajo producen diferentes patrones de distribución de lignina en la ultraestructura de la fibra. La autohidrólisis remueve lignina principalmente desde la lamela media la que es desplazada hacia el exterior de la pared celular en la forma de esferas y conglomerados amorfos. La hidrólisis con ácido diluido remueve la lignina desde la pared celular secundaria, acumulándose en la lamela media (se observó un ensanchamiento de esa zona) y dejando la celulosa de la pared más expuesta. En Organosolv, la lignina es relocalizada formando numerosas esferas (principalmente) y conglomerados amorfos de diferentes tamaños, los que se distribuyen tanto al interior como al exterior de la pared celular.
8. A mayor severidad del pre-tratamiento los patrones de relocalización de lignina mencionados se ven acentuados, produciendo finalmente el colapso de la pared celular, lo que se traduce en un mayor rendimiento de hidrólisis enzimática.