On the impact of drought-induced abiotic stress on the composition of Douglas-fir lignin for valorization

by

Daniel Barker-Rothschild

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Chemical Engineering

Department of Chemical and Materials Engineering

University of Alberta

Abstract

Rising interest in the transition to renewable energy, materials, and chemicals ushers biorefining strategies that effectively utilize all components of lignocellulosic biomass feedstocks. Concurrently, more frequent and severe drought brought on by accelerating climate change pose questions regarding the reliability of biomass feedstocks. Lignin is a phenolic branched copolymer deposited mainly in the secondary plant cell wall of some plant cells providing strength, rigidity, and pathogen resistance to most terrestrial plants. Lignin also accounts for ~15-30% of lignocellulosic biomass composition and up to 40% of its energy content. Lignin has been correlated with drought stress response mechanisms of plants due to the various properties it confers to the plant cell wall. Lignin has also become a center-point of contemporary biorefining schemes due to its impact on biomass processing chemistry and inherent potential for various high-value products as the most abundant renewable aromatic carbon source.

One of the most ecologically and economically important conifers in North America known for its large size, high lignin content, and drought resistance; Douglas-fir, is a promising feedstock for contemporary biorefineries. This thesis has investigated the impact of drought induced stress on the composition of Douglas-fir lignin to assess to what degree lignin may play a role in its drought stress tolerance mechanisms. In particular, this work has looked at lignin content and structural composition which may be useful indicators for assessing biomass feedstock quality.

Wood from the outer rings of stems of Douglas-fir [*Pseudotsuga menziesii* var. *menziesii* (Mirb.) Franco] seedlings grown in either severe drought or under normal conditions for one year have been harvested for analysis. Lignin has been extracted from the wood samples using a mild organic solvent extraction approach for more detailed analysis of its structural features. Wood and

lignin samples were analysed with wet chemical approaches, attenuated total reflectance Fourier transform infrared spectroscopy, pyrolysis-gas chromatography – mass spectrometry, quantitative phosphorus-31 nuclear magnetic resonance spectroscopy and quantitative solid-state multiple-cross polarization carbon-13 nuclear magnetic resonance spectroscopy. While wet chemical analysis did not discern differences in lignin content, quantitative solid-state carbon-13 nuclear magnetic resonance spectroscopy of wood revealed a 5% higher lignin content in the drought stressed wood compared to the control. This result provides some evidence to suggest lignin biosynthesis may be associated with the drought stress response mechanism in Douglas-fir. Furthermore, quantitative phosphorus-31 and solid-state carbon-13 nuclear magnetic resonance spectroscopy of the isolated lignin indicated slight variations in concentration of distinct hydroxyl group functionalities and other structural features between the samples. Altogether, while this thesis has provided encouraging evidence for associating lignin content and composition with the drought stress response of Douglas-fir, further investigation will be necessary to better understand this mechanism. The findings of this thesis have aided in identifying promising pathways for more comprehensive and sophisticated investigations. The limitations of the current work and new directions for future studies have been discussed in detail.

The first five chapters of this document provide a detailed background of the key areas related to the investigation carried out for this the thesis while chapter 6 describes the experimental work. Chapter 1 provides an introduction and background to the subject along with characterizing the status of the field, defining challenges to be addressed, and presenting a hypothesis for this thesis. Chapter 2 reviews the literature regarding the association between plant drought stress tolerance and lignin, discussing the approaches, current understandings, challenges and limitations. Chapter 3 provides an overview of lignocellulosic biomass fractionation strategies,

covering both the industrial approaches as well as laboratory methods. Chapter 4 reviews lignin characterization and the analytical tools for elucidating structural features. In chapter 5, a review of chemometric approaches applied for lignocellulosic biomass and lignin characterization is presented including a detailed background on chemometrics and an organized assessment of approaches applied to different analytical challenges in the field. Chapter 6 describes and discusses the experimental work performed for this investigation, including an analysis of the results and the conclusions regarding the effects of drought stress on the composition of Douglas-fir wood observed in this work. Finally, chapter 7 summarizes the work done for this thesis, presents key findings, limitations, and prospects for future work.

Preface

This thesis is an original work by Daniel Barker-Rothschild under the supervision of Professor Phillip Choi. The investigation conducted for this thesis is the product of an interdisciplinary research collaboration between the Department of Chemical and Materials Engineering of the University of Alberta and two labs under the Department of Natural Resources Canada: The Canadian Wood Fibre Centre, Victoria, British Colombia, of the forest sector and CanmetENERGY, Devon, Alberta, of the energy sector. The project carried out during the work of this thesis was funded through the Canadian Wood Fibre Centre. Supervision and mentorship in addition to analysis and writing support was provided by Stanislav Stoyanov, Cosmin Filipescu, Mike Cruickshank and Rafal Gieleciak of Natural Resources Canada. Douglas-fir seed [*Pseudotsuga menziesii* var. *menziesii* (Mirb.) Franco] originated from full-sibling families from the BC Ministry of Forests Douglas-fir breeding program. Sample harvesting, preparation, and experiments were performed by Daniel Barker-Rothschild and other members of the Natural Resources Canada project team. Solid-state ^{13}C NMR spectroscopy was performed by the University of Alberta Chemistry Department. All the data and experimental results were analysed by Daniel Barker-Rothschild under the supervision of the project team. Versions of Chapters 2, 3, 4, 5 and 6 are being prepared for or have been submitted for publication.

Acknowledgements

I would like to express my sincere gratitude to Phillip Choi for taking me under his supervision, for his inspirational approach to teaching, for his eagerness and enthusiasm to share his time with me and for the always entertaining and thought-provoking conversations. Stanislav Stoyanov for his greatly appreciated and valuable supervision. His guidance, experience, and mentorship have been an invaluable support for my growth over the last two years. Cosmin Filipescu for his kind nature, for all his help in organizing and supporting my project, for his valuable perspective, and for his positive enthusiasm in all our interactions. Rafal Gieleciak for his valuable time, feedback, and efforts which supported the work on my project. Mike Cruickshank for all his work to support my project, for always providing his honest opinion, his detailed and insightful feedback, and for the knowledge that I gained while working with him. Felysia Green and Dominique Lejour for their welcoming nature and their valuable work to support the project. David Dunn and Rebecca Dixon for welcoming me into their laboratory and for their work on the project. Rebecca Mertens for her friendship and hard work in the lab. Daniel Mazerolle and the rest of the wood fibre centre for their role in the very positive experience I had during my masters. Jenny Hristova for welcoming me into her lab and providing many useful and inspiring life lessons. Trevor McHarg, Derek Chao, Craig McMullen, Matthew Kusinski, Hosnay Mobarok, Jonathan Mane and the rest of CanmetENERGY for their support with my project. Brittney Klein, Vladimir Michaelis, and the University of Alberta Department of Chemistry for their work on the solid-state NMR experiments and consultation. Seth Beck for his friendship, advice and collaboration, and motivation to challenge myself. To the rest of Professor Choi's research group, graduate peers and department members for fostering a healthy and productive environment. My parents Catherine Barker and Wayne Rothschild for teaching me to value critical thinking, encouraging me to be the best I can be, and for providing me with the opportunity to pursue my interests. To my sister Miriam and the rest of my family that love and support me. My partner Ashley Turner for her love and support through my undergraduate and graduate degrees. To all my family, friends, teammates, teachers, and mentors.

Thank you sincerely for your support.

Table of Contents

List of Tables

List of Figures

List of Symbols

R^2	*Coefficient of Determination*
$\mathbf{X}$	*Independent Variables Matrix Representation*
$\mathbf{Y}$	*Dependent Variables Matrix Representation*
$\boldsymbol{x}$	*Independent Variables Vector Representation*
$\boldsymbol{y}$	*Dependent Variables Vector Representation*
x	*Independent Variables Scalar Representation*
y	*Dependent Variables Scalar Representation*
V_h	*Hydrodynamic Volume*
e_i	*Residuals*
y_i	*True Value*
$\hat{y}_i$	*Predicted Value*
M_w	*Weight Average Molecular Weight*
st	*Stretching vibration*
δ_{ip}	*In-plane deformation vibration*
δ_{op}	*Out-of-plane deformation vibration*
I_i	*Integrated area for the "i" region where i is a text subscript which specifies the region of interest*
I'_i	*Normalized integrated area for the "i" region where i is a text subscript which specifies the region of interest*
^{13}C	*Carbon-13 Isotope*
^{31}P	*Phosphorus-31 Isotope*

Glossary of Terms

ABA	*Abscisic acid*
ADF	*Acid-detergent fibre*
ADL	*Acid-detergent lignin*
AH	*Acid-hydrolysis*
ALS	*Alternating Least Squares*
AFEX	*Ammonia fibre explosion/expansion*
ANN	*Artificial Neural Network(s)*
APX	*Ascorbate peroxidase*
ATR	*Attenuated Total Reflectance*
AP2/ERF	*APETALA2/Ethylene responsive factor*
BTX	*Mixtures of benzene, toluene, and xylenes*
FRCBOSS	*Bootstrapping Soft Shrinkage Combined with Frequency and Regression Coefficient of Variables*
COMT	*Caffeic acid-O-methyltransferase*
CCoAOMT	*Caffeoyl-CoA O-methyltransferase*
CEL	*Cellulolytic enzyme lignin*
CAD	*Cinnamyl alcohol dehydrogenase*
CCR	*Cinnamoyl-CoA Reductase*
CP-MAS	*Cross Polarization Magic Angle Spinning*
CRISPR	*Clustered regularly interspaced short palindromic repeats*
CV	*Cross-validation*
DES	*Deep euctectic solvents*
DCM	*Dichloromethane*
DE	*Direct Exposure-Mass Spectrometry*
DOSY	*Diffusion-Ordered Spectroscopy*
DFRC	*Derivatization Followed by Reductive Cleavage*
DTG	*Derivative Thermogravimetry*
EHL	*Enzymatic hydrolysis lignin*

EMAL	*Enzymatic/mild acidolysis lignin*
EMSC	*Extended Multiplicative Signal Correction*
FT-IR	*Fourier Transform-Infrared Spectroscopy*
G	*Guaiacyl lignin monomer unit*
GA	*Genetic Algorithms*
GC	*Gas Chromatography*
GC-MS	*Gas Chromatography-Mass Spectrometry*
HCA	*Hierarchical Cluster Analysis*
H/G ratio	*Ratio of p-hydroxylphenol (H) to Guaiacyl (G) lignin monomer units*
HMBC	*Heteronuclear Multiple Bond Correlation*
HPSEC	*High-performance Size Exclusion Chromatography*
HSQC	*Heteronuclear Single Quantum Coherence*
ICA	*Independent Component Analysis*
IR	*Infrared*
ISO	*International Organization for Standardization*
IL	*Ionic liquid*
k-NN	*k-Nearest Neighbor*
LASSO	*Least Absolute Shrinkage and Selection Operator*
LCC	*Lignin-carbohydrate Complexes*
LDA	*Linear Discriminant Analysis*
LHW	*Liquid hot water*
MAS	*Magic Angle Spinning*
MDA	*Malondialdehyde*
MALDI	*Matrix Assisted Laser Desorption/Ionization*
MS	*Mass Spectrometry*
MIR	*Mid-infrared Spectroscopy*
MCR	*Multivariate Curve Resolution*
MC-UVE	*Monte-Carlo Uninformative Variable Elimination*
MLR	*Multiple Linear Regression*

MultiCP	*Multiple Cross-polarization*
MSE	*Mean Squared Error*
MWD	*Molecular Weight Distribution*
MSC	*Multiplicative Signal Correction*
MWL	*Milled Wood Lignin*
m/z	*Mass-to-Charge Ratio*
NDF	*Neutral-detergent fibre*
NIR	*Near-infrared Spectroscopy*
NBO	*Nitrobenzene Oxidation*
NIPALS	*Nonlinear Iterative Partial Least Squares*
NMF	*Non-negative Matrix Factorization*
NMR	*Nuclear Magnetic Resonance Spectroscopy*
O-PLS/OPLS	*Orthogonal Projections to Latent Structures*
OPLS-DA	*Orthogonal Projections to Latent Structures–Discriminant Analysis*
OSC	*Orthogonal Signal Correction*
H	*p-hydroxyphenyl lignin monomer unit*
PAL	*Phenylalanine ammonia-lyase*
PARAFAC	*Parallel Factor Analysis*
PARADISe	*PARAFAC2 based Deconvolution and Identification System*
PO	*Permanganate Oxidation*
POD	*Peroxidase*
PCA	*Principal Component Analysis*
PCR	*Principal Component Regression*
pGrAdd	*Python Group Additivity*
pMuTT	*Python Multiscale Thermochemistry Toolbox*
PLS/PLSR	*Partial Least Squares Regression*
PLS-DA	*Partial Least Squares–Discriminant Analysis*
qRT-PCR	*Real-time quantitative reverse transcription polymerase chain reaction*
PRESS	*Predicted Residual Error Sum of Squares*

Py-MBMS	*Pyrolysis Molecular Beam–Mass Spectrometry*
Py-GC/MS	*Pyrolysis Gas Chromatography–Mass Spectrometry*
RAV	*Related to ABI/VP1*
RING	*Rule Input Network Generator*
RMSE	*Root Mean Squared Error*
RNA-seq	*Ribonucleic acid sequencing*
ROS	*Reactive oxygen species*
RSS	*Residual Sum of Squares*
RT-PCR	*Reverse transcription polymerase chain reaction*
SEC	*Size Exclusion Chromatography*
SEP_{Test}	*Standard Error of Prediction using an External Test Set*
SEP_{Cal}	*Standard Error of Prediction using an Internal Calibration Set*
SE	*Steam explosion*
S/G Ratio	*Ratio of Syringyl (S) to Guaiacyl (G) lignin monomer units*
SIMCA	*Soft Independent Modeling of Class Analogy*
SMARTS	*SMILES Arbitrary Target Specification*
SMILES	*Simplified Molecular Input Line Entry System*
SNV	*Standard Normal Variate*
SOD	*Superoxide dismutase*
SOM	*Kohonen Self-organizing Map*
SVD	*Singular Value Decomposition*
SVM	*Support Vector Machines*
S	*Syringyl lignin monomer unit*
TGA	*Thermogravimetry Analysis*
TG-MS	*Thermogravimetry Mass Spectrometry*
TA	*Thioacidolysis*
TPPO	*Triphenyl phosphorus oxide*
UV	*Ultraviolet*
UVRR	*Ultraviolet Resonance Raman*

VCA	*Vertex Component Analysis*
^{1}H NMR	*Proton Nuclear Magnetic Resonance Spectroscopy*
^{13}C NMR	*Carbon-13 Nuclear Magnetic Resonance Spectroscopy*
^{31}P NMR	*Phosphorus-31 Nuclear Magnetic Resonance Spectroscopy*
1D	*1 Dimensional*
2D	*2 Dimensional*
TMDP	*2-chloro-4,4,5,5-tetramethyl-1,3,2-dioxaphospholane*
4CL	*4-coumarate-CoA ligase*

Chapter 1: Introduction

1.1 Background

Growing concerns of anthropogenic contributions to climate change have prompted increased research into energy, materials, and chemicals that would reduce emissions and sustain the global demand for resources. To this end, lignocellulosic biomass has emerged as a promising candidate as a renewable and sustainable feedstock. Derived from organic plant matter, lignocellulosic biomass is primarily composed of three biopolymers: cellulose, hemicellulose, and lignin. Plants consume carbon dioxide and water and convert light energy from the sun into chemical bonds that build up its structure. The energy stored within these chemical bonds and the naturally synthesized biopolymers within plants can be accessed and exploited and therefore present a unique opportunity as widely available renewable feedstocks.

There are a variety of renewable resources that have the potential to meet the growing demand, all with unique challenges. Renewable energy technologies, such as solar and wind, have become well known and increasingly advanced. However, they face challenges with integration into existing electrical grid distribution and consumption schemes based on their intermittent nature [1]. Biomass has the potential to address some concerns faced by other common renewables such as intermittency but has its own challenges as well. The challenges to using biomass as a renewable energy source include carbon emissions, environmental concerns during the refining process and cost-effective utilization at scale. Of the various renewable resources available, plant biomass is the only renewable carbon source in nature, providing it with a unique capability for production of value-added products [2]. Around 40% of a plant's dry matter is carbon, making lignocellulosic biomass not only a potential renewable source of energy but also a renewable source of carbon [3].

Due to the renewable nature of lignocellulosic biomass along with its wide accessibility and diverse utility, some have proposed that our conventional fossil-fuel dominated society could be replaced with a bio-based economy established upon the principals of green chemistry. There are several environmental, social, technical, logistical, and economic considerations that need to be addressed before there is a large-scale transition from conventional resources to lignocellulosic

biomass. Such considerations include variations in biomass availability, harvesting and consolidation, high production costs (pre-treatment being among the most costly steps [4]), existing technical limitations (conversion), and complex and costly fuel and supply chain logistics [5]. There are a number of techno-economic and life cycle analyses of lignocellulosic biomass as a renewable resource available in literature that touch on many of these considerations. These studies are highly dependent on geographical location, feedstock used, and the end products, among other considerations, and therefore pose a challenge for accurate assessments [6]. Generally, there is agreement that with continued research and development efforts, lignocellulosic biomass is a promising resource.

The Department of Natural Resources Canada (NRCan; French: Ministère des Ressources Naturelles Canada; RNCan), responsible for natural resources for the federal government of Canada, defines renewable energy as the energy derived from natural processes that are replenished at a rate equal to or faster than the rate at which they are consumed [7]. Canada, a large landmass with diverse geography, has an abundant supply of renewable resources which include moving water, wind, biomass, solar, geothermal, and ocean energy. Behind hydro and wind, biomass is the third largest in renewable electricity generation. Canada has a unique potential to utilize biomass as a renewable energy source because of its large landmass and active forest and agricultural industries [7]. Despite the fact that bioenergy is still in its development stages, Canada produces more than 400 petajoules of bioenergy every year in the industrial sector, over half of which is derived from the pulp and paper industry [7]. Within this context, since 2014 the Government of Canada has several measures and mandates (see Table 1) in place to support the production and use of renewable fuels. These measures include renewable fuels regulations, $200 million over four years to support farmer participation in the industry, $1.5 billion over nine years to support domestic production through an operating incentive program, and a $500 million NextGen Biofuels Fund™ to support the next-generation technologies for biofuels from non-conventional feedstocks [7].

Table 1. Provincial Renewable Fuel Mandates, 2014 [7].

		Renewable Alternatives to Gasoline	Renewable Alternatives to Diesel
Federal		5%	2%
Provincial			
	British Columbia	5%	4%
	Alberta	5%	2%
	Saskatchewan	7.5%	2%
	Manitoba	8.5%	2%
	Ontario	5%	2%
	Quebec	5% (target only)	--

In 2009, the Forest Products Association of Canada and its partners, including FPInnovations and the Canadian Forest Service of Natural Resources Canada, have initiated the Bio-pathways Project [8]. The project has involved over 65 experts from a variety of sectors which include finance, biotechnology, and energy development to assess Canada's forest industries position for entering the future of biomass production for energy and value-added projects [8]. They have identified six lines of inquiry to pursue and assess opportunities based on financial, socio-economic, and environmental perspectives [8]. So far, the project has shown promising findings for the integration of the Canadian forest industry and the next generation bioeconomy. It must be noted that the findings indicated the valorization of biomass to bioproducts in addition to bioenergy would be essential for sustaining the industry in the long term [8].

Traditionally, it has been proposed that separating lignocellulosic biomass to its individual components would be ideal for further processing into their respective products and there have been significant efforts to this objective. In particular, the separation of lignin from the polysaccharides is an important element of this approach. This is because lignin is considered the main contributor to the recalcitrant nature of the plant cell wall. In this context, we consider recalcitrance as the term often used to describe the state of the plant cell wall being resistant to fractionation/processing into its isolated components. The primary challenge of utilizing lignocellulosic biomass has long been considered the recalcitrant nature of the plant cell walls,

preventing conversion into valuable products at a cost that is competitive with that of the equivalent petroleum-based alternatives. Since lignin is ascribed most of the fault for this inconvenient property of plants, it has been the subject of extensive research in the past few decades. In-fact, one could perhaps depict it as its own entire field of research, encompassing a variety of approaches and extending over a range of disciplines. A vast volume of literature has been developed in this field of research surrounding lignin covering a variety of approaches and perspectives. These efforts, along with growing emphasis on adhering to the principles of green chemistry have transformed the perceptions regarding biomass utilization. More complete and effective utilization of lignocellulosic biomass is now considered a necessity for contemporary refineries. This shift in approach and advancements in applications has stimulated a great deal of interest in what are termed contemporary biorefinery strategies. There is no strict definition for the term biorefinery, and perceptions also vary geographically. Here, we consider biorefineries to be any integrated industrial process that utilizes biomass feedstock(s) to produce bio-based fuels, chemicals, or materials [9]. The vision of a contemporary biorefinery is that which efficiently utilizes all the main components of lignocellulosic biomass, including lignin, for value-added applications. Based on our definition, this could include pulp and paper industries or bioethanol refineries that that effectively utilize their lignin streams for value-added applications.

While both effective isolation methods and value-added applications of lignin were once limited, there are now many developing approaches and opportunities. There are a wide range of treatments and well studied pathways to valuable products. Lignin currently has a wealth of proposed applications including fuels, chemicals, and advanced materials. For example, as a renewable aromatic, lignin offers potential for producing biomass-derived jet turbine fuels, for transformation into conventionally petroleum derived benzene, toluene, and xylenes (BTX) which are widely used as industrial solvents and feedstocks in many industries, and even for advanced materials application such as high-value carbon fibres [10] [11].

Since lignin is a major component of lignocellulosic biomass, the effective utilization of the resource is a prerequisite of any modern biorefinery. Currently, lignin is underutilized, with industries such as pulp and paper producing large quantities (50-70 million US tonnes world-wide annually [11]) and burning as much as 98% of the extracted lignin as low value fuel [12]. The most abundant aromatic polymer in nature, accounting for 10-35% by weight, 40% by energy of

biomass, will likely prove to be a key piece in the development of modern biorefineries [2]. Zhang et al. [4] have provided an economic analysis based on lignocellulose intermediates and the potential revenues from the final products. These authors calculate the value of the lignocellulose intermediates based on three categories of utilization: simple, partial, and complete. The simple utilization is defined as all sugars being glucose and lignin is burned as fuel. The partial utilization is defined as all sugars being glucose, half of the lignin is burned as fuel and half is sold as materials, with acetic acid as a commodity. The complete utilization is defined as each sugar sold at its high price, all lignin utilized as polymeric materials and acetic acid as a commodity. The authors estimate the final revenues of the simple, partial, and complete utilizations of lignocellulosic biomass components to be $121, $254, and $628 per ton respectively [4]. These projections clearly communicate the economic importance of the valorization of all lignocellulosic components, particularly lignin. Key to the acceleration of lignin valorization is understanding its chemical structure and interactions with other components within the plant cell wall.

1.2 The plant cell wall

Despite lignocellulosic biomass having a great potential as a renewable resource, there are challenges to the efficient conversion of raw material to refined products. To understand these challenges, scientists are focusing on understanding the biochemistry of the plant cell wall. Plants are composed of around 40 cell types with distinct and dynamic cell walls that play a major role in determining the growth and overall configuration of the plants [13]. These cell walls determine the shape of the cells and provide many functions ranging from structural support and protection to transport and communication [14]. Plant cell walls can be further divided into two main layers, the primary and secondary plant cell walls, which are distinct in composition, developmental timeline, and position.

Primary cell walls which are ubiquitous to plant cells are composed primarily of cellulose, hemicellulose, and pectin. The cellulose polymer chains stack and aggregate into stable microfibrils with high axial stiffness [15] [16]. Hemicelluloses are a class of polysaccharides which coat cellulose microfibrils and link these together forming a network [17]. Pectins are another class of polysaccharides that mainly include homogalacturonan and rhamnogalacturonan-I and -II [18]. Pectin is also found in the middle lamella where it plays a key role in binding

adjacent cells together. Primary cell walls are thin, flexible, and well hydrated [13]. After the primary cell wall is fully developed, some plant cells develop secondary cell walls between the primary cell wall and the plasma membrane. The secondary cell wall has some structural similarities to the primary cell wall with some important differences. The secondary plant cell wall is mainly composed of cellulose, hemicellulose, and lignin, with minor components being proteins, lipids, soluble sugars, and minerals [19]. The key distinguishing feature that sets the secondary plant cell wall apart from the primary plant cell wall is the presence of lignin [17]. Other distinctions include differences in hemicellulose composition and the absence of pectin present in the secondary plant cell walls [14]. Lignins are complex, amorphous, aromatic heteropolymers consisting of phenylpropanoid monomer units. The distinct compositional differences of the secondary plant cell wall provide the plant with increased strength and rigidity. Additionally, the properties of lignin such as hydrophobicity and resistance to degradation provide improved water transport and pathogen protection. Thus, plants with highly lignified plant cell walls have distinct properties as compared to those without. In-fact, it is believed that the development of lignin biosynthesis was a key factor in plant adaptations to terrestrial ecosystems around 400 million years ago [20].

While there have been increasing developments in our understanding of the plant cell walls – largely as a result of advancements in technology – the plant cell wall is still the least understood cellular structure in plants [21]. While there are still many unanswered questions, the current understanding of the secondary plant cell wall architecture proposes that the cellulose microfibrils are organized in a matrix with hemicellulose hydrogen bonding to the microfibril surfaces as well as covalently bonding to other hemicellulose and lignin [19]. The covalent bonds between hemicellulose and lignin are known as lignin-carbohydrate complexes (LCC) and are proposed to be another important contributor to the difficulties in the isolation and extraction of lignin from lignocellulosic biomass because of their enzyme impenetrable cross-links [22] [19].

1.3 Plant cell wall biopolymers

Cellulose is a linear homopolymer composed of D-glucopyranose units linked via the $\beta(1 \rightarrow 4)$ glycosidic bonds. These glucan chains assemble into sheets via edge-to-edge hydrogen bonds and stack into crystalline structures known as cellulose microfibrils [23]. Cellulose

microfibrils have regions of highly ordered crystalline structures as well as disordered amorphous regions [16]. These long and rigid microfibrils form a structural network acting as a load-bearing scaffold in the cell wall [13]. Branched polysaccharides (hemicellulose/pectin) cross-link the cellulose microfibrils. Hemicellulose is an amorphous branched copolymer composed of pentoses (D-xylose, D-arabinose), hexoses (D-mannose, D-glucose, D-galactose) and occasionally uronic acids and acetyl moieties as side chain groups [24] [25]. Lignin is a branched and cross-linked aromatic copolymer composed primarily of three canonical 4-hydroxyphenylpropanoids which differ by the number of methoxy units on the 3rd and 5th position on the phenyl ring. These are known as *p*-coumaryl (4-hydroxycinnamyl), coniferyl (3-methoxy-4-hydroxycinnamyl), and sinapyl (3,5-dimethoxy-4-hydroxycinnamyl) alcohol, or as *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units, respectively, in the context of the lignin polymer (See Figure 1). Lignin fills the gaps in the cell wall polysaccharide structures and can also covalently link with hemicelluloses or cellulose.

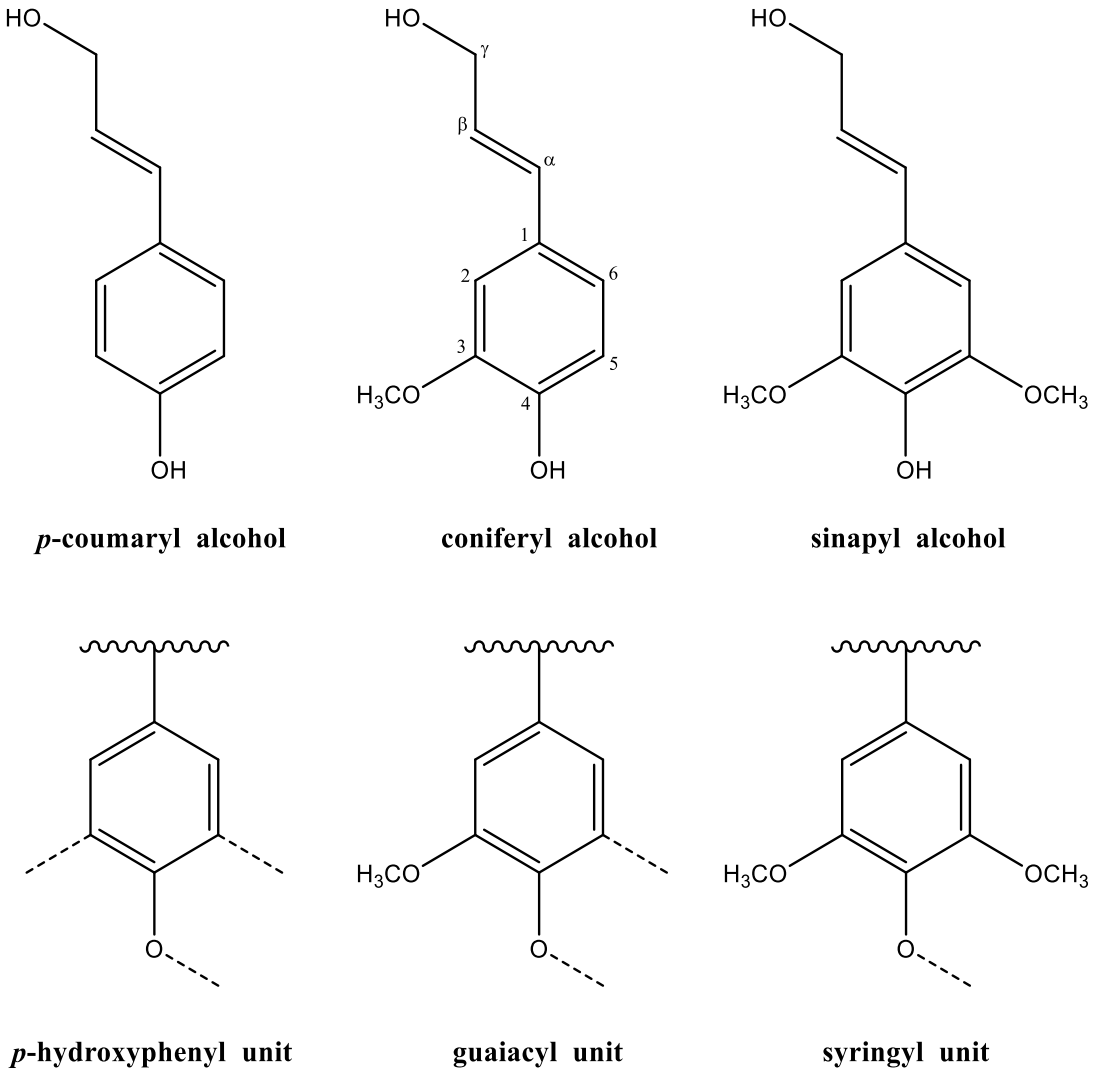


Figure 1. The three main lignin monomer unit precursors (top) and their respective units in the context of the lignin polymer (bottom).

Years of rigorous research efforts to elucidate lignins structure has provided a good basis for understanding its composition. Yet, the complexity of the heterogenous lignin polymer, the limitations of studying it in situ, and its high variability have proven significant challenges and there are still many open areas of debate. Currently the most detailed analytical investigations can provide information on lignin monomer composition, interunit linkages, functional groups, molecular weight, interactions with carbohydrates and more. These features can be used to generate models of lignins structures. Despite this, the lignin structure is still subject to debate. Historically, lignin has been widely accepted to be a branched polymer; however, this has been

called into question. Crestini et al. [26] have found that milled wood lignin, which is a product of a common pre-treatment method and considered one of the closer representations of that of native lignin, is a (mostly) linear oligomer. This has been a controversial finding because it is not consistent with the classical understanding of lignins structure. Additionally, in a 2019 review of the lignin structure Ralph et al. [27] have noted that there is currently no structural evidence for branching. They state that the commonly assumed notion of a highly condensed 3D lignin structure is now in question. More recently, the findings of Crestini et al. [26] have been reviewed by Balakshin et al. [28], who have found strong evidence that spruce milled wood lignin has a high degree of branching/crosslinking. The conflicting opinions on such a crucial and basic structural feature of lignin underline the fact that the true structure of lignin is still unknown. As research methods become more advanced, it is likely that new developments will be made in the near future. Certainly, more research on the lignin structure will be critical for the work of scientists and engineers developing advanced lignin valorization strategies.

The lignin structure is extrapolated from the relative proportions of the main monomer units, inter-unit linkages, structural moieties, functional groups, and degree of branching among other features. Common structures and inter-unit linkages are illustrated in Figure 2. Among the various linkages, the $\beta - O - 4$ ether linkage is the most frequent and the weakest of the covalent linkages, making it the easiest to cleave during extraction/isolation and thus a target of many fractionation approaches [22].

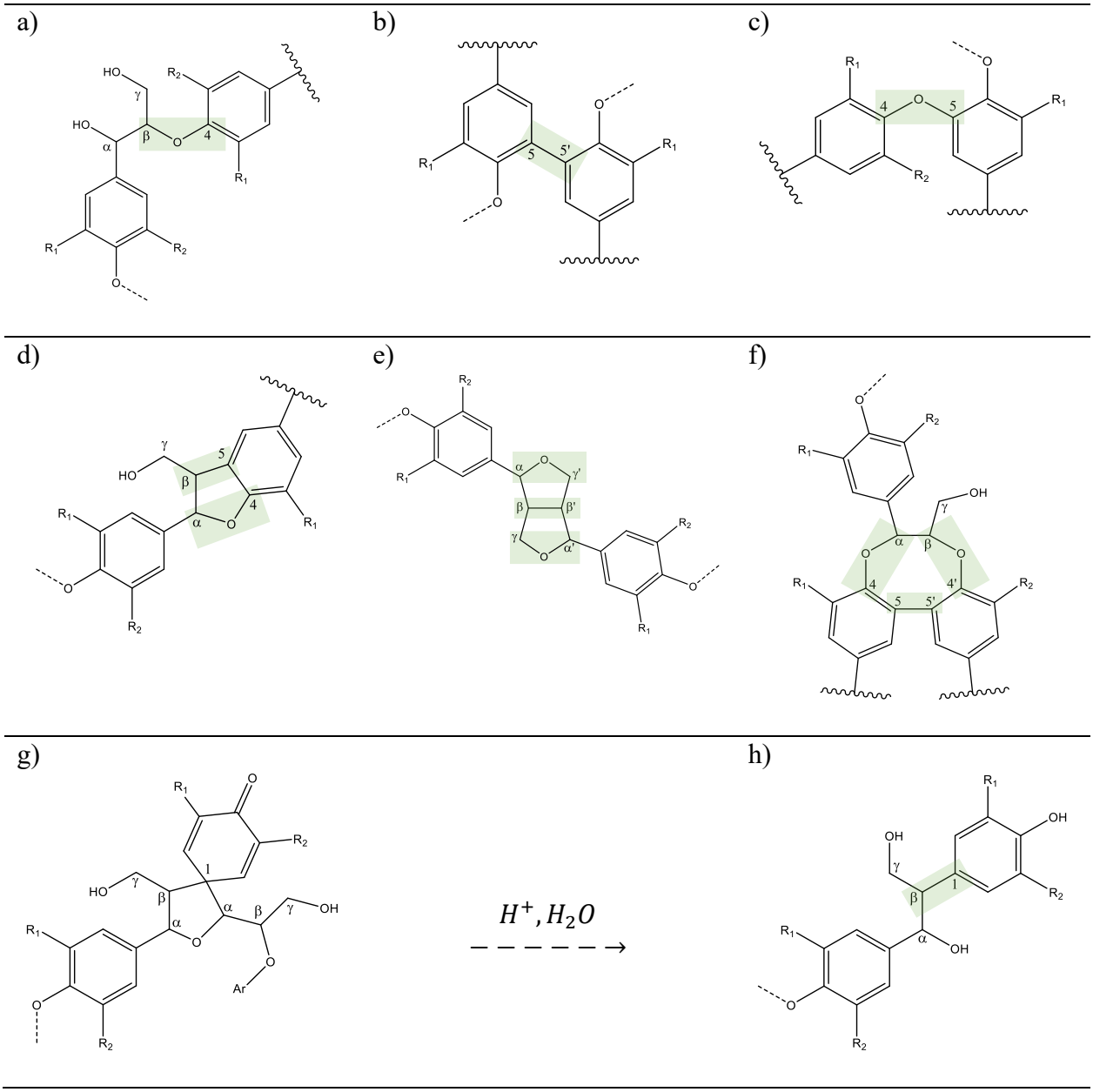


Figure 2. Illustrations of the main lignin bonding patterns including structures and inter-unit linkages a) β −arylether ($\beta - O - 4$) b) Biphenyl ($5 - 5'$) c) Biphenyl ether ($4 - O - 5$) d) Phenylcoumaran ($\beta - 5, \alpha - O - 5$) e) Resinol ($\beta - \beta', \alpha - O - \gamma$) f) Dibenzodioxocin ($5 - 5', \beta - O - 4, \alpha - O - 4$) g) Spirodienone h) Diarylpropane ($\beta - 1$). Where $R_1 = R_2 = H$ in p-hydroxyphenyl units; $R_1 = OCH_3, R_2 = H$ in guaiacyl units, $R_1 = R_2 = OCH_3$ in syringyl units. Adapted from Lapierre [29].

1.4 Sub-topic introduction and content guide

The main objective of this thesis was to assess the impact of drought-induced stress on the content and composition of Douglas-fir lignin to better understand if and how plant lignin biosynthesis mechanisms play a role in drought stress tolerance. Furthermore, a major focal point was to concentrate on understanding to what degree these potential changes affect the value of Douglas-fir for wood as a feedstock for contemporary biorefining approaches. I hypothesised that drought-induced stress would result in altered composition of Douglas-fir wood, specifically involving an increase in lignin content as a defense mechanism providing increased water retention and conferring extra strength and support to the tree in its vulnerable state. In addition to altered lignin content, it could be readily anticipated the spatio-temporal regulation of lignin biosynthesis and deposition could be altered in response to metabolic adjustments as the plant responds to drought stress conditions. The consequences could include altered proportions of monomer units, including non-conventional units, or bonding patterns which would result in modified frequency of certain inter-unit linkages. Variations in lignin content and composition resulting from drought would have important bi-directional implications for both selective breeding for feedstock improvement and for refining strategies to value-added products.

This thesis comprises seven chapters and three appendices covering a variety of sub-topics related to the field of lignocellulosic biomass with a focus on lignin valorization. Many of the chapters in this thesis provide a detailed supporting background of the multifaceted investigation that has been undertaken. Chapter 1 provides an introduction and overview of the chemistry of lignocellulosic biomass concentrating on the three main components of the plant cell wall and in particular the chemical structure of lignin. Once the basics of the plant cell wall chemistry have been discussed, Chapter 2 will provide a literature review focusing on the relationship between drought stress tolerance and plant cell wall lignin biosynthesis. Chapter 3 will then review lignocellulosic biomass fractionation strategies, including both commercial pretreatment and laboratory isolation approaches. Chapter 4 will review analytical chemistry approaches for lignin characterization. Chapter 5 presents a review paper titled “Chemometrics and Lignocellulosic Biomass – Advanced Strategies for the Accelerated Analysis and Valorization of Lignin” which provides a detailed introductory background of chemometrics and subsequently systematically reviews their applications to the field of lignocellulosic biomass, highlighting accomplishments

and opportunities. Chapter 6 presents a research paper "Assessing the impact of drought stress on the composition of Douglas-fir wood and the structure of softwood lignin for valorization" which describes and discusses the experimental work done for this thesis project. Chapter 7 provides a summary of the work performed for this thesis, key findings and limitations, and discussed prospects and strategies for future work.

Chapter 2: Drought response and plant cell wall lignification

2.1 Introduction

Drought is responsible for more annual loss in crop yield than all other pathogens combined and hence is one of the most important factors concerning feedstock reliability for biorefineries and many other agricultural and food industries [30]. Drought conditions are expected to become more frequent and severe due to climate change, disproportionately affecting northern latitudes [31]. Drought conditions have a major impact on plant development both in terms of growth and composition. Plants have developed a variety of adaptations that allow them to survive in different environments and phenotypic plasticity can allow them to adjust to changing environments [32]. These adaptation mechanisms are sophisticated and can include physiological and biochemical responses. Strategies for coping with drought stress are termed broadly as drought resistance, which can be further categorized into three mechanisms: escape, avoidance, and tolerance [33]. Drought escape refers to the acceleration of the life cycle to completion prior to the effects of drought. Drought avoidance refers to mechanisms that help plants to maintain water content. Drought tolerance is the term for mechanisms that support plants in enduring the condition of reduced tissue water content.

It is known that lignin biosynthesis is the product of a complex genetic and transcriptional network in which both biotic and abiotic stressors can influence and regulate the production and deposition of lignin [34]. While the lignin biosynthesis pathway is well studied, it is not so well understood how environmental stressors influence this process. It is suspected lignin biosynthesis is involved in drought stress response mechanisms. For example, it is proposed that plants may increase lignin content in response to drought stress, reducing the permeability of its tissues to restrict moisture loss and strengthening the cell walls preventing cavitation. However, there is no firm consensus on the effects of drought on lignin biosynthesis and efforts to understand these mechanisms have accelerated in recent years. This is due to the increasing interest in lignocellulosic feedstocks as important feedstocks for contemporary biorefineries. With the growing interest in lignins role in biomass processing in parallel with rising concerns of climate

change and the resulting impacts such as drought, it is clear that connecting lignin biosynthesis with drought stress response mechanisms is an important objective. It is interesting to note that many selective breeding and bioengineering approaches have focused on producing trees with reduced lignin content to increase processing efficiency and gain access to the polysaccharide components such as cellulose. However, this type of approach could produce plants that are vulnerable to drought stress, thus reducing the reliability of the feedstock. Future approaches may look to tailor lignocellulosic feedstocks for drought resistance and optimized composition for industrial processing.

Efforts to understand the relationship between lignin biosynthesis and drought stress response mechanisms have increased in recent years. In addition, the approaches to probe these mechanisms have become progressively sophisticated. While early studies were limited to evaluating just lignin content, composition, or distribution, more recent approaches have focused on connecting drought stress response at the transcriptional level with the phenotypic traits such as lignin content and more recently to lignin composition. Such examples would include performing drought trials on genetically modified plants with selectively over/under expressed transcription factors which regulate the expression of specific genes. In this section, the literature from the last twenty or so years related to understanding the mechanisms of plant drought stress response and its relationship with lignin biosynthesis is reviewed. Studies were summarized to provide a full picture of the objectives, approach, methods, and findings. The section concludes with a discussion on the overall findings from the evaluation of the literature. A summary table of the studies identified in this review along with their key findings related to lignin and drought stress can be found in Appendix A.

2.2 Literature Review

Donaldson [35] investigated *Pinus radiata* logs harvested from a forest located on coastal sanddunes with shallow sand over a hard iron pan subsoil – an environment likely to subject the trees to periodic severe drought conditions. Confocal fluorescence microscopy using acriflavine (stains the lignified cell walls green and non-lignified red) indicated a severely abnormal lignin distribution in terms of quantity and distribution in the tracheids in the outer part of the stem compared to normal wood. Transmission electron microscopy provided additional support to the

results of the confocal fluorescence microscopy with more detail and higher resolution. The results of the study showed that under drought stress, *Pinus radiata* logs have tracheids with concentric layers of abnormal lignification within the secondary cell wall and reduced lignification in the middle lamella.

Interested in the prospect of water deficit causing locally altered cell wall composition in the basal region of the root elongation zone, Fan et al. [36] investigated the changes in metabolism and accumulation of cell wall phenolic substances in maize (Zea *mays*) seedling primary roots. A candidate gene approach with semiquantitative reverse transcription polymerase chain reaction (RT-PCR) and northern blots was used to look at the potential expression of transcripts of lignin synthesis genes as well as the time scale effects of water deficit in the maize root elongation zone. FT-IR spectroscopy was also used to evaluate cell wall compositional changes and UV-fluorescence microscopy for visualizing lignin in the cross sections from equivalent root regions. They found pieces of interrelated evidence supporting the hypothesis that wall extensibility and root growth are inhibited by water deficit through alterations of phenolics in the cell wall in the root elongation zone. Transcripts of two cinnamoyl-CoA reductase (CCR) genes involved in lignin biosynthesis were detected in the elongation zone of both well-watered and stressed roots. CCR transcript levels were found to have increased after just 1 hour of water deficit, and prior to reductions in cell wall extensibility, indicating a potential causal association. Progressive increases in IR absorbance and UV fluorescence were indicative of increased cell wall phenolics from 3 to 9 mm behind the tops of water deficit-treated roots and colocalized with the progressive inhibition of wall extensibility and growth. Water deficit was observed to induce increases in UV fluorescence and lignin staining in the cell walls of inner tissues of the stele, which seemed to specifically limit root growth rates.

Lee et al. [37] tested the hypothesis that enzymes related to lignification responsible for stress tolerance mechanisms would activate at different rates in relation to the degree of drought stress, by investigating the activity of enzymes, growth, and stress associated parameters in drought stressed white clover leaves. They found that physiological parameters and lignification related enzymes responded to leaf water potential under water-deficit. They also identified two distinct phases of drought stress which begins with a “mild endurance period” in which ascorbate peroxidase (APX) and phenylalanine ammonia-lyase (PAL) are activated providing antioxidant

protection, and a “severe injury period” in which lignifying peroxidases guaiacol peroxidase, CPOX, and syringaldazine peroxidase are enhanced, and lignin and lipid peroxidation is increased, restricting growth.

Hu et al. [38] focused on identifying expression changes in maize leaf proteins under drought stress and investigated their physiological functions in drought tolerance. They identified the expression of many drought-induced proteins including three involved in lignin biosynthesis. Leaf lignin content was determined and found to increase significantly under severe and moderate drought treatments. They also observed significant differences in leaf lignification between drought-tolerant and drought-sensitive inbred lines. The authors conclude that the results indicate gene selection based on molecular markers for breeding drought tolerant maize could be a promising approach.

Moura-Sobczak et al. [39] sought to determine if drought stress caused changes in the quantity and composition of lignin in the basal and apical regions of the stems of *Eucalyptus globulus* Labill and in the hybrids E. *urograndis* (E. *urophylla* x E. *grandis*) and E. *uroglobulus* (E. *globulus x* E. *urograndis*). Total lignin was determined using thioglycolic acid and GC-MS were used to determine lignin monomeric composition. There was not a consistent theme in regard to lignin content and the plants displayed varying responses to drought stress in terms of lignin content. However, it was found that the adjustment of total lignin resulted in an increased S/G ratio, either by increasing the amount of lignin S units or by the reduction of lignin G units.

Li et al. [40] sought to study drought tolerance and post-drought recovery for two different genotypes of white clover (*Trifolium repens* L., a drought-tolerant small-leafed white clover and a drought-sensitive large-leafed white clover) associated with antioxidative enzyme and lignin metabolism. The activity of lignin related enzymes polyphenol oxidase and cinnamyl alcohol dehydrogenase (CAD) were found to vary between small and large leafed clovers but the enzyme activities were observed to increase under drought stress and decrease after re-watering for both clovers. The lignin content of the small-leafed clover was found to significantly increase under drought stress; however, the large-leafed clover did not display the same trend. Additionally, root lignin content of both materials increased significantly under drought stress and returned to a uniform increase of lignin content after rewatering. Their results indicated that lignin metabolism

and/or the cell wall lignification response is more sensitive in roots than leaves. They suggest that based on their findings that lignification decreases cell wall plasticity, inhibits cell wall growth, and assists with water retention. Hence, lignin metabolism likely played an important role in drought stress response and post-drought recovery for white clovers.

Leaf rolling is likely an evolve trait of plants in response to drought stress as a mechanism that reduces the surface area of its leaves reached by light which as a result reduces transpiration and leaf dehydration. Terzi et al. [41] investigated the role of lignification in the leaf rolling mechanism of *Ctenanthe setosa* plants in response to long-term drought stress. They studied the enzymes associated with lignification in the unrolled leaves as control and at two different leaf rolling indices at days 35 and 47 of the drought period. They found increases in activity of various lignification-related enzymes in response to drought stress along with a positive correlation between leaf rolling and lignin content. Overall based on the results of their study, the authors proposed that drought stress could stimulate lignification in leaves of *Ctenanthe setosa* as the leaf-rolling process and the increased lignification were positively correlated. Increased lignification in the leaves may also play a role as an independent defense mechanism in response to drought stress.

dos Santos et al. [42] studied the effects of water deficit on lignin metabolism in the stem of two commercial sugar cane (*Saccharum* spp.) genotypes. Expression patterns of various genes related to lignin biosynthesis were observed to be induced under water deficit and varied based on developmental phase of the stem, internodes, tissue type, and water regime. They found that severe water stress effected lignin deposition in the stem internodes in both genotypes. Compared to the respective control plants, the water stress plants featured significant accumulation of lignin. Additionally, it was observed that lignification was dependant on internode maturation (immature or mature) and the tissue region in the stem (rind or pith). The mature internodes exhibited more lignification than immature internodes, and the rind more lignified than the pith. The authors postulated that increased lignification in the rind may be a genetic trait functioning to enhance xylem vessel resistance and reduce water loss to manage the higher tension under water deficit. They also note that differences in study methodologies, tissue type, age, genotype variation and other factors could account for differences in lignin content reported in literature regarding sugarcane. Curiously, in mature internodes an overall reduction in lignin related enzymes but a

30% increase in lignin content in the mature rind in one of the genotypes was observed under drought stress as compared to control plants. The authors defer to the high complexity of the genomes of modern sugarcane cultivars, suggesting that functional redundancy in genes may explain the observed inverse correlation between lignin biosynthesis gene expression and lignin content observed in this study.

Srivastava et al. [43] mimicked drought and salinity stress conditions (often interconnected) by treating developing seedlings of *Leucaena leucocephala* (Vernacular name: Subabul, White popinac) with 1 % mannitol and 200 mM NaCl. They monitored enzyme linked immunosorbant assay (ELISA) based expression pattern of CCR protein, stained transverse sections of seedlings using Phloroglucinol/HCL for lignin, and qualitatively analysed structural features of Klason lignin using FT-IR spectroscopy to study lignin deposition in response to stress. Since CCR is considered a key gene for regulating lignin biosynthesis, comparing the abundance of this gene with lignin deposition in the stem and root tissues under drought stress may enable the linking of genotypes and phenotypes. While they observed no clear effect of drought stress on the developing root, they did observe a positive effect of drought stress conditions on lignin deposition and CCR expression in developing stems. FT-IR spectroscopy of Klason lignin was not able to identify differences in type/composition of the lignin deposited as a result of drought stress conditions; however, they note that the technique may have limited their ability to resolve these potential differences. In addition to being primarily a qualitative method, FT-IR spectroscopy is quite low resolution compared to other techniques such as NMR spectroscopy. Another important concern is the nature of the lignin studied. As will be discussed in Chapter 3, Klason lignin is known to be a relatively severe technique that is more often used for quantitative determination of total lignin in a biomass sample, rather than as a research grade lignin for which to make conclusions on the native lignin structure as it resides in the plant cell wall. The authors concluded that CCR accumulation, immune-cytolocalization of CCR protein and lignin deposition pattern of both root and stem tissues of treated and control samples further substantiated the suggestion that CCR is important for vascular tissue development under drought stress conditions in developing seedlings of *Leucaena*.

Xu et al. [44] investigated the transcriptional factors PtoMYB170 and its duplicate PtoMYB216 in Chinese white poplar (Populus *tomentosa*) for its functions in lignin deposition

and drought tolerance. They used gene cloning and sequence analysis, quantitative RT-PCR, histochemical staining, a scanning electron microscope, chemical analysis of the secondary cell wall components, and more, to reveal insights into these drought response mechanisms. Results showed increased lignification and thicker secondary cell wall in xylem in transgenic poplar plants with overexpressed PtoMYB170 as compared to wild type. In contrast, clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-generated mutation of PtoMYB170 showed reduced lignin deposition and resulted in more flexible and collapsed xylem phenotypes. Drought tolerance by reduced water loss through augmented stomata closure resulting from heterologous expression of PtoMYB170 in Arabidosis was also observed. The authors concluded that PtoMYB170 increases lignin deposition during wood formation in poplar by triggering lignin biosynthetic gene expression as well as promoting dark-induced stomata closure and thus drought tolerance in comparison to its divergent homologous PtoMYB216 gene.

Miscanthus is a perennial biomass crop with good potential as a renewable feedstock for biofuels (in particular cellulosic ethanol) because of its fast growth, low moisture, and high cell wall carbohydrate contents compared to other plants. Since drought is one of the most common abiotic stresses and can affect plant growth, physiological processes, and cell wall biosynthesis, Van der Weijde et al. [45] looked to investigate the compositional changes that occur as a result of drought stress and their impact on biomass quality for the production of biofuels. They investigated plant growth and compositional quality of the stem and leaf material of 50 different miscanthus genotypes which included *Miscanthus sinensis, Miscanthus sacchariflorus,* and interspecific hybrids under drought and control conditions. They used NDF/ADF and ADL method of Goering and Van Soest to determine the cellulose, hemicellulose, and lignin content of the biomass samples. Saccharification efficiency was evaluated through the conversion of cellulose into glucose with a mild alkaline pretreatment and enzymatic saccharification reactions. They also used near-infrared (NIR) spectroscopy and chemometric techniques to develop rapid prediction models of biomass properties. They obtained NIR absorbance spectra of 130 samples and processed the spectra with weighted multiplicated scatter correction and mathematical derivatization and smoothing treatments. A modified partial least-squares regression analysis was developed with a calibration set of 110 randomly selected samples (1:1 leaf to stem ratio) and the 20 remaining samples were used as an external model validation set. The models were used to

predict the cell wall cellulose, hemicelluloses, lignin contents, and cellulose conversion rate of the leaf and stem samples. Drought treatment was found to significantly impact almost all evaluated traits, and the different genotypes also showed significant differences in performance in response to drought conditions. Reductions in plant weight up to 70% and on average 45% compared to the control plants was observed. The authors also note that reduction plant weight under drought conditions varied significantly between genotypes, which indicates that there may be large variation in genotypic drought stress response mechanisms. While significant differences in most biomass quality traits were observed as a result of drought conditions, including a decrease in average cellulose content in the stem from 51% to 46%, and an increase in hemicellulose content in the stem from 42% to 47%, lignin content was only mildly affected in the stem with no significant effect on the lignin in the leaf tissue. This might be because lignin content is relatively low in grasses as compared to other biomass types (<10% lignin content in stem and leaf tissues), and therefore, may not play as significant of a role in drought tolerance for Miscanthus. The authors note that the minor effect of drought on cell wall lignin content and large effect on cellulose content was observed consistently over a diverse set of genotypes and three different miscanthus species. The authors observed significant genotypic variation in terms of cell wall composition and cellulose conversion, with drought-treated plants displaying more genotypic variation in terms of cell wall composition than control plants. Evaluation of saccharification efficiency revealed significant differences between drought and control treatments with plants under drought-conditions displaying higher cellulose conversion efficiencies. They conclude that based on the absence of strong correlations among drought tolerance and compositional characteristics, biomass quality characteristics are likely under independent genetic control and thus it may be possible to select for both drought tolerance and biomass quality traits using calculated breeding strategies.

Wildhagen et al. [46] used three *Pupulus nigra* L. genotypes from habitats of differing water availability to study the associations between wood anatomy, lignin content, and saccharification efficiency, as well as identify genes and co-expressed gene clusters linked with genotype and drought-induced variation in wood traits and saccharification potential. Their initial hypothesis was that drought would increase lignification and as a result decrease saccharification potential, genotypes originating from different environments will have different lignin content and thus will also have different saccharification potential, and that drought and genotype effects on

saccharification potential are supported by differences in wood anatomical traits and transcript abundances of genes involved in wood formation and cell wall metabolism, particularly lignin biosynthesis. In contrast with their initial hypothesis, their results showed that glucose release was in-fact moderately improved from drought conditions with a significant increase in saccharification potential under drought across all genotypes. Additionally, lignin content was not affected by drought stress – thus suggesting there may be a different mechanism of cell wall development for drought tolerance. Their investigation of transcriptional regulation of biosynthesis genes reveled that hemicelluloses and modifications of pectins were negatively correlated with saccharification potential, thus appearing as potential future targets of investigation for understanding drought stress and saccharification potential in *Populus nigra* L. Based on their work and discussion of several other previous studies, they suggested that lignin may not be a good predictor of saccharification potential.

After developing several engineered Arabidopsis plants for desired biofuel traits (altered galactan, xylan, lignin and/or acetylated xylan) in previous studies, Yan et al. [47] investigated the phenotypic traits of these plants in response to drought stress to better understand how they would perform in field conditions. Most of their engineered plants showed significantly better survival than wild type plants with the low xylan, low lignin, and low acetyl substitution degree of xylan plants demonstrating reduced levels of water loss and better drought tolerance than wild-type plants. They then investigated the dependency of the plant's drought tolerance to abscisic acid (ABA), which is an important mediator for triggering plant responses to common abiotic stresses including drought. The results suggested a dependency between the drought tolerance of the plants with low lignin and ABA; however, drought tolerance of plants with low xylan or low acetylation did not appear to be ABA dependant. Additionally, compared to wild type plants, engineered plants with low lignin content had stomata closures more responsive to ABA, which reduces water loss and could be key for conferring drought tolerance. Examination of the expression of drought-responsive genes in the plants also indicated that the plants engineered for low xylan, low lignin, and acetyl substitution of xylan have increased induction of stress-responsive genes, potentially accounting for their superior drought tolerance. The authors conclude that their study demonstrates potential for engineering plants with improved cell wall composition for processing objectives without compromising resiliency to environmental stressors.

Lima et al. [48] investigated the relationship between lignin, embolism resistance, and leaf life span in 22 tree species in a seasonally dry tropical ecosystem in north-eastern Brazil. They hypothesised that drought resistance would be determined by lignin content and composition. They used a thioglycolic acid and spectrophotometer method for determining the total lignin content of the samples and thioacidolysis with GC-MS for determining relative proportion of lignin monomer units. Their results showed that lignin monomer composition of the trees from their study was related to dry-season leaf life span and xylem embolism resistance. However, total lignin content was found to not to be correlated with any variable in their study, including wood density, which was also uncorrelated with lignin monomer composition. They propose two main lines of trait variation for drought tolerance based on their results, one for keeping their leaves around longer by increasing the S/G ratio, and the other for leafless trees to retain more water in the stem and to shed leaves to keep higher step water potential. They conclude that the two lines are related and that the mechanisms that link S/G ratio with xylem embolisms vulnerability requires further investigation.

Bang et al. [49] created overexpression and knockdown transgenic lines of OsTF1L, a rice homeodomain-leucine zipper transcription factor gene, to further investigate their function in drought tolerance mechanisms. They observed that overexpression of OsTF1L improved drought tolerance and grain yield of rice, with the overexpressed plants displaying reduced and delayed drought-induced damage, decreased water loss, normal photosynthetic efficiency under drought conditions, and recovered more rapidly after rehydration in comparison to the control plants. They found that overexpression of OsTF1L increased shoot lignification in the typically lignified tissues, with no ectopic lignification throughout the plant. These results were consistent with ribonucleic acid sequencing (RNA-seq) and real-time quantitative reverse transcription (qRT-PCR) analysis of lignin biosynthetic genes. Overexpression of OsTF1L also resulted in up-regulation of stomatal movement genes and higher ratio of stomata closure under drought conditions, thus reducing water loss. This indicates that OsTF1L could increase drought tolerance through multiple molecular mechanisms such as stomatal movement and lignin biosynthesis.

In the interest of understanding and improving drought tolerance of grapevine for the horticultural industry, Tu et al. [50] investigated lignins role and significance in drought resistance in grapevine. They generated VlbZIP30-overexpressed transgenic grapevine plants and analyzed

the plants response to drought treatment relative to control plants. qRT-PCR analysis allowed for the assessment of expression of VlbZIP30 in transgenic lines. Lignin content and monomer composition was determined using the acetyl bromide and thioacidolysis method, respectively. RNA-seq analysis was used to link the impacts of overexpression of VlbZIP30 and lignin biosynthesis. They found that the transgenic plants from their study with overexpression of VlbZIP30 displayed improved drought tolerance, demonstrated by the increased leaf relative water content, tuning of photosynthesis rate, and increased lignin content in the leaves. Increased leaf lignin content was also found to be primarily guaiacyl units. Consistent with those results, they found that VlbZIP30 regulates the expression of lignin biosynthetic (VvPRX N1, VvPRX4, and VvPRX72) and drought responsive (VvNAC17) genes, supporting lignin biosynthesis and improving drought resistance in grapevine, However, they also found that overexpressing VlbZIP30 in *Arabidopsis thaliana* did not result in the same regulation of lignin biosynthesis as in grapevine, which the authors speculate could be due to the difference in herbaceous versus woody biosynthesis mechanisms. Additionally, they found differences in lignin biosynthesis pathway regulation between stem and leaves, indicating that lignin biosynthesis may be regulated differently in certain tissues. The authors emphasize this is the first report of a bZIP transcriptional factor being directly involved in lignin biosynthesis and enhancing drought resistance in plants, which should support future molecular breeding strategies.

Liu et al. [51] investigated the roles of important enzymes for lignin synthesis and lignins function in drought tolerance by overexpressing *CmCAD1*, 2, and 3 in Arabidopsis *cadc cadd* double mutant and collectively and differentially silencing *CmCAD1*, *2*, and *3* in melon seedlings. They found that all three genes participate in lignification and that *CmCAD2* and *3* contribute more significantly to drought tolerance through their function's lignin synthesis. Additionally, *CmCAD2* and *3* positively recovered lignification in both floral stem and root development of Arabidopsis *cadc cadd* double mutant and speculated that this could reduce lateral water loss and increase water absorption thereby enhancing drought tolerance. Silencing of *CmCAD2* and/or *3* in melon seedlings significantly reduced drought tolerance of the melon plants which displayed less lignified but still tight and dense secondary walls. This is somewhat in contrast to the *cadc cadd* which were less dense, loose, and had disorganized secondary cell walls, suggesting differences in the lignification process between the two species. The authors conclude that lignification is

important for both melon and Arabidopsis in tolerating drought stress, and in particular *CmCAD2* and *3* are key to this mechanism.

Foxtail millet (*Setaria italica*) is a particularly drought resistance rice crop that may be a useful candidate for probing drought resistance mechanisms in plants on the molecular scale. Xu et al. [52] explored this mechanism by identifying a drought-induced R2R3-MYB transcriptional factor SiMYB56 in foxtail millet, and then comparing wild-type to transgenic plants with overexpressed SiMYB56 in normal and under drought stress conditions. They found that SiMYB56 can respond to abiotic stress in foxtail millet and that it had transcription repression activity. Drought stress tolerance analysis of SiMYB56 functions did not reveal any apparent growth performance difference between the transgenic and wild-type controls; however, there was significantly higher survival rate in the transgenic lines (50-80%) as compared to the wild-type rice (10%) under drought conditions. This indicated that SiMYB56 confers drought tolerance to the foxtail millet rice. In order to further explore the mechanism by which SiMYB56 confers drought tolerance, they simulated drought stress on the plants using 10% PEG6000. They found significantly higher lignin content in the transgenic rice as compared to the wild-type plants in this simulated drought condition, which the authors suggest may indicate that SiMYB56 confers drought tolerance in the transgenic rice by increasing lignin biosynthesis. Quantitative real-time PCR was used to measure the expression of several lignin biosynthesis related genes including PAL, 4-coumarate-CoA ligase (4CL), C4H, CCR, CAD, and F5H, which indicated that SiMYB56 overexpression can increase lignin content through the activation of lignin biosynthesis enzymes. Two years of field trails confirmed that the transgenic plants had improved drought tolerance compared to the wild-type controls with higher total grain weight in both years. Overall, the study showed that SiMYB56 may confer drought resistance in Foxtail millet by stimulating lignin biosynthesis under drought conditions.

Li et al. [53] investigated the functional mechanisms of the DREB gene, DcDREB1A, cloned from carrots (*Daucus carota* L.), in response to drought stress in transgenic DcDREB1A overexpressed *Arabidopsis thaliana*. Their results indicated that DcDREB1A is a nuclear protein involved in the regulation of plant drought stress tolerance. Isolation of DcDREB1A showed a significant increase in expression of the transcriptional factor gene under drought treatment. In response to drought treatment, transgenic *Arabidopsis thaliana* with overexpressed DcDREB1A

had decreased stomata opening and density, increased expression levels of lignin biosynthesis genes, increased lignin accumulation in the stem and leaves, and increased reactive oxygen species (ROS) -scavenging enzymes superoxide dismutase (SOD) and peroxidase (POD). These results indicate that DcDREB1A is involved in plant drought tolerance by reducing water loss through stomata regulation and lignin deposition, decreasing oxidative lipid damage from ROS by increasing SOD and POD activities, and increasing the expression of other stress-responsive genes in plants.

Sharma et al. [54] looked at lignin deposition in the legume crop chickpea (*Cicer arietinum* L.) in response to drought stress. They focused on the root xylem as roots can be the first organs to interact with environmental changes that cause stress. Drought treated plants showed an increase in both primary root length and total lignin content of ~25% as compared to the control samples. Of which the increased lignin deposition primarily occurred in the xylem tissue. Laccases are multicopper oxidases that play a role in lignin biosynthesis via oxidative coupling of monolignols. Analysis of the expression of six laccase genes (which are linked with lignin biosynthesis) after drought treatment and relative to the control plants showed alterations in the expressions of some laccase genes. The authors concluded that chickpea root length and lignin content may be increased under drought conditions and that the laccase gene family may play a role in this stress response mechanism.

Gu et al. [55] used tandem mass tag and liquid chromatography-tandem mass spectrometry to study the proteomic profiles of stress proteins in tea plants under drought stress, specifically how they impact lignin, flavonoids, and fatty acids. Of the 4789 proteins identified, 11 were up- and 100 were downregulated, respectively. These included many proteins related to lignin biosynthesis. This was contrasted by the contents of lignin in the plants showing an increase under drought stress, indicating that tea plants respond to drought stress by inhibiting the accumulation of enzymes that catalyse lignin biosynthesis, while still promoting the accumulation of lignin, potentially through encouraging the activity of these enzymes.

Li et al. [56] performed a genome-wide analysis of laccase genes in moso bamboo using a bioinformatic approach. They identified 23 *PeLACs* in moso bamboo and then narrowed their focus on PeLAC10 as it was estimated to be the most relevant candidate to study its relation to

lignin. Their results positively correlated the expression of PeLAC10 with lignification in bamboo. They then used PeLAC10 overexpressed transgenic *Arabidopsis* to further investigate the function of PeLAC10. Histochemical staining and qualitative analysis of lignin content in the stems indicated that PeLAC10 may be associated with lignification as shown by increased lignin content in the transgenic plants. Different stress treatments also showed that the transgenic plants were more resistant to phenolic acid stress and had normal growth compared to the control plants under drought stress treatment. The transgenic plants had better survivability, relatively normal growth, and reduced malondialdehyde (MDA) content (MDA concentration is an indicator of membrane damage from ROS) compared to the control plants. These traits might be related to the higher level of lignification in the transgenic plants as a result of the overexpression of PeLAC10. The authors conclude that they provided information on laccase family members in bamboo and demonstrated that overexpression of PeLAC10 could be a promising strategy for molecular breeding of bamboo with improved stress tolerance.

Hori et al. [57] studied the cell walls of a model woody biomass, poplar (*Populus trichocarpa*), and their response to drought and high-salt stress conditions using both analytical chemistry and transcriptomic analysis. They looked at the young shoot and mature stem (xylem). Lignin content was determined with the thioglycolic acid method. The results showed that for both the young shoots and xylem, abiotic stress influenced cell wall polysaccharide content but not total lignin content. More detailed structural analysis of the plant cell wall lignin was investigated using solution-state 2D heteronuclear single-quantum coherence (HSQC) NMR spectroscopy. They found that lignin composition in the young shoot was significantly altered as shown by a decrease in S and increase in G lignin monomer units under both drought and salt stress conditions. They did not observe the same change in the xylem tissues. Transcriptional analysis revealed significant alterations in the expression of cell-wall related genes in young shoots in response to abiotic stress conditions. In addition, genes involved in cellulose, hemicellulose, and the core lignin biosynthesis pathway genes (F5H and COMT) were significantly decreased. The authors note that since the down regulation of F5H and COMT has been observed in rice under salt and osmotic stress conditions, it may be a common abiotic stress response among plants. The authors have demonstrated here that lignin composition can be altered in response to drought stress without influencing lignin content, indicating it may be a strategy in drought response. They propose that

decreasing the S/G ratio could decrease cell wall damage and increase hydrophobicity of the plant cell wall. They have also identified some promising targets of molecular engineering strategies for tailored lignocellulosic breeding and feedstock traits.

Cao et al. [58] investigated the role of lignin in cell physiology during plant growth and drought stress tolerance by generating transgenic *Populus* with cell-specific downregulation of the 4CL gene to supress lignin biosynthesis in vessels and normal fibres. Transgenic plants with reduced lignin content in the vessels but not fibres had significant reduction in sap flow and hydraulic conductance in the xylem transportation system compared to wild type plants. These transgenic plants were also more vulnerable to drought, displaying dwarfism, low survival rate, and reduction of aboveground biomass yield. In contrast, transgenic plants with reduced lignin in the fibres instead of the vessels did not exhibit significantly reduced performance of xylem transportation system or vulnerability to drought stress, although they did show more significant reduced mechanical strength. These results underscore the importance of lignin for water transport and mechanical strength in plants and in particular highlights that lignin may play various physiological functions depending on tissue and cell type. These results suggest that discussions of lignin's role in drought stress response mechanisms must be discussed in a tissue specific context.

Zhao et al. [59] isolated the full-length cDNA of a caffeoyl-CoA O-methyltransferase gene (CCoAOMT) from P. *ostii* which is an important enzyme for lignin biosynthesis and is suggested to play a role in abiotic stress tolerance. They studied its expression patterns in P. *ostia* and its subcellular localization in tobacco leaves. They also used CCoAOMT overexpressed transgenic tobacco plants to study the drought stress response mechanism associated with this enzyme. Using transcriptome sequencing data of P. *ostii* under drought stress and bioinformatic techniques, they determined a protein PoCCoAOMT of which they then studied the expression pattern under drought stress. Under drought treatment, the expression levels of PoCCoAOMT in P. *ostii* was found to be overexpressed as compared to the control plants. The PoCCoAOMT overexpressed transgenic tobacco plants were then used to further understand its role in drought stress response. The results showed that the transgenic lines had significantly increased drought tolerance compared to the wildtype tobacco plants as demonstrated by the better growth rate, higher relative leaf water content, decreased relative electrical conductivity, and decreased accumulation of

reactive oxygen species $O_2^{\cdot-}$ and H_2O_2. In connection with the decreased accumulation of ROS, the activity of antioxidant enzymes SOD and POD were found to be significantly higher in transgenic plants while APX was found to be significantly lower. Curiously, the degree of stomata opening under drought stress in the transgenic tobacco plants was found to be higher than the wildtype. Stomata closure is a common drought stress response adaptation that allows plants to regulate their transpiration/CO_2 uptake through their leaves which is known to be induced by ABA. Here, the transgenic lines that were more drought tolerant had greater stomata opening, indicating that the drought response mechanism in transgenic plants with overexpressed PoCCoAOMT allowed for the larger stomata opening, encouraging better overall photosynthetic activity for the plant. To further understand this mechanism, they looked to investigate the role that lignin may play by determining the expression of lignin biosynthesis genes, lignin content and lignin monomeric composition in the plant's roots, stem, and leaves. In the transgenic lines they found PoCCoAOMT was expressed in all three tissues, however, the roots showed the highest expression level. Consistent with this finding, lignin content was significantly higher in the roots, stem, and leaves than in the wildtype plants, and highest in the roots. Four genes involved in lignin biosynthesis PAL, CAD gene, 4CL and caffeic acid-O-methyltransferase gene (COMT) were found to have increased expression in the transgenic lines compared to the wildtype lines in all tissues, although inconsistently. Quantitative analysis of lignin monomer content revealed that guaiacyl and syringyl units were the predominant components of the plant's lignin, and the transgenic lines had increased levels of both monomer units, with a larger increase guaiacyl units. Thus, they speculated that drought stress tolerance could be specifically associated with higher levels of guaiacyl lignin units. The authors conclude by proposing that a potential basis for improving drought stress tolerance in P. *ostia* could involve exploiting the expression of PoCCoAOMT to promote lignin synthesis and ROS scavenging activity.

Yan et al. [60] investigated the molecular mechanism of drought resistance in Cassava (*Manihot esculenta*), a woody shrub important for both food and energy. Specifically, they looked at the role of Related to ABI/VP1 (RAV) which is an important subfamily of the APETALA2/Ethylene responsive factor (AP2/ERF) transcription factor family, in drought response. They found that silencing the drought stress-responsive transcriptional factor RAV5 significantly reduced drought stress resistance in Cassava, with both control and drought plants

showing higher H_2O_2 and lower lignin content, but to a greater extent in the RAV5-silenced plants. Additionally, they showed MeRAV5 physically interacted with MePOD and lignin-related CAD 15 (MeCAD15), promoting their activities to influence H_2O_2 and lignin content. To further explore this relation, they developed MeCAD15 and/or MeRAV5 co-silenced cassava lines and discovered that these silenced lines showed lower lignin content as well as a drought-sensitive phenotype. However, treatment of these lines with external alkali lignin improved drought resistance, indicating that MeRAV5- mediated CAD activity and resultant increased lignin content could be responsible for the enhanced drought resistance conferred through MeRAV5. They conclude by proposing a model for the molecular mechanism of MeRAV5, MePOD and MeCAD15 action in modulating plant drought resistance, whereby drought stress induces the expression of these transcriptional factors, with MeRAV5 interacting with both MePOD and MeCAD15, promoting their activities and influencing the accumulation of H_2O_2 and lignin, conferring drought resistance to the plant.

Wen et al. [61] isolated a transcriptional factor gene, MsWRKY11, generally considered to be stress-inducible, and investigated its potential mechanism in drought tolerance in alfalfa. Transgenic alfalfa with overexpressed MsWRKY11 or dominantly repressed MsWRKY11 in seedlings or hairy roots were used to probe this mechanism. They found that the MsWRKY11-overexpression lines were more resistant to drought stress than the wildtype under drought treatment, with a significant increase in stomatal density (14-20%). To understand how MsWRKY11 may affect lignin biosynthesis in alfalfa under normal and drought conditions, the lignin content of the stem, leaves, and hairy roots were determined. Additionally, the lignin composition in the stem in terms of its relative abundance of monomer units were analysed using the TA followed by GC-MS. After identifying the transcription factor MsWRKY22 as potentially binding to promoter sequence of MsWRKY11 under various abiotic stresses, transgenic planted with altered expression of MsWRKY22 were generated. They found that the MsWRKY11 or MsWRKY22 overexpressed transgenic lines had increased lignin content in the stems, while the MsWRKY11 or MsWRKY22 dominantly repressed transgenic lines had decreased lignin contents in the hair roots, indicating that MsWRKY11 positively regulates lignin biosynthesis in stems and roots of alfalfa. The MsWRKY11 overexpressed lines had increased levels of all three main monomer units in the stem, with the greatest increase found in the syringyl monomer units. Overall,

the authors showed that MsWRKY11, positively activated by MsWRKY22, increases stomatal density and lignin accumulation, conferring drought tolerance to alfalfa.

Liu et al. [62] investigated the transcriptional regulatory network involved in wood formation by screening for interacting transcription factors as well as for transcription factor DNA interaction. In particular, they focused on the transcription factor PtrMYB074, previously identified as key for regulating wood formation. After identifying PtrWRKY19 as the most highly expressed xylem-specific interactor, transcription factor-gene analysis of PtrMYB074 and PtrWRKY19 revealed that these two transcription factors co-target a set of genes, most notably *PtrbHLH186*. Genetic analysis demonstrated that overexpression of *PtrbHLH186* resulted in abnormal lignification, enhanced vessel cell development and altered wood composition, in particular, early and increased lignification, increased total lignin, increased guaiacyl units, decreased polysaccharides, and more drought-tolerant phenotypes.

2.3 Summary and findings

Work to elucidate lignin's role in drought stress response mechanisms of plants has intensified in the last few years as reflected in both the number of studies and increasing sophistication of the approaches. Many of the recent studies have focused on the transcriptional regulatory networks, which have been made more accessible by advances in "omics", bioinformatics, and new techniques such as RNA-seq which allows quantification of gene expression levels [63]. These findings can be connected with phenotypic observations to develop qualitative models of drought stress response mechanisms. Thus, very recently more elaborate information has been published regarding lignin's role in drought stress tolerance mechanisms, expanding our understanding and offering new directions for investigations. While there have been major advancements in our knowledge, this has also revealed the complexity of these drought response mechanisms, demonstrating the substantial work still to be undertaken in this area. Section 2.2 has provided an overview of the literature regarding lignin its relationship to drought stress response. This section will aim to summarize the findings and offer perspectives based on the summation of the literature. First, the general conclusions regarding lignin's content and structure changes in response to drought stress will be reviewed. Subsequently, the effect of drought stress on lignin biosynthesis at the metabolic level will be discussed. Next, proposals

concerning the function of lignin in drought stress tolerance will be explored. The section will conclude with a discussion of limitations, general comments, and future opportunities.

Regarding lignin and drought stress tolerance of plants, this section considers four main questions: Is lignin content and/or composition altered in response to drought stress conditions? Does lignin content, composition, or the metabolic intermediates involved in its biosynthesis, acting either alone or in combination, impact drought stress tolerance in plants? If lignin is associated with drought stress response, is this the result of the properties it confers to the plant cell wall or rather the by-product of a greater drought stress response mechanism. If lignin is associated with drought stress response, what is the spatio-temporal nature? For example, which plants, which tissues/cells, what stage of development, and what are the kinetics and mechanism of the response. A challenge with making broad conclusions regarding lignin's role in drought stress tolerance is the immense genetic variation among plants. Results may be heavily dependent on the plant in question, the tissue and cell types being assessed, the drought treatment, and the techniques used for analysis. Despite there being some conflicting evidence, in general it appears that more often than not, studies have found a positive association with lignin content and drought stress. This finding has been observed in a variety of tissues, including those in the roots, stem and leaves, and has been shown to be dependent on tissue maturity [42].

Research regarding lignin composition and drought tolerance has lagged compared to that of lignin content; however, more recently lignin composition has been increasingly considered. These efforts have mainly been limited to evaluating lignin monomer composition, a detailed assessment of lignin structure has not been observed. In response to drought stress, some studies have found an increase in syringyl over guaiacyl units [39] [48] [61], and an increase in guaiacyl over syringyl units [57] [59] [62]. The conflicting evidence here may suggest that the relationship regarding drought stress tolerance and lignin composition is complex. More work will be required to further understand this matter. The findings may have key implications for biomass processing as S/G ratio is commonly associated with processing efficiency and the manipulation of monomer units could be the target of breeding strategies for improved feedstock properties.

Many studies have noted the increased expression of lignin biosynthesis related genes and transcription factors under drought stress and have connected them with lignin deposition and other

observations such as leaf rolling, stomata closure, defenses to reactive oxygen species and more [41] [41] [43] [44] [49] [53]. These results clearly affirm lignins involvement in the stress response mechanism of many plants. It is interesting to note that there have been multiple observations of increases in tissue lignification despite decreases in lignin related enzymes [42] [55]. Functional redundancy in genes has been proposed as a potential explanation for this observation [42], but the observation is a good reminder of the complexity of these mechanisms. Lignin has been linked with reducing membrane damage via the regulation of ROS [56] [59] [60]. Studies have also found that overexpression of genes found to induce drought tolerance in one plant species may have little or no effect on drought tolerance in another, indicating differences in drought tolerance mechanisms or lignin biosynthesis pathways between plant species [50][51]. Additionally, differences in lignin biosynthesis pathway regulation among various tissues have been observed, indicating that lignin biosynthesis may be regulated differently in certain tissues [50]. Transgenic plants with tissue specific gene regulation have demonstrated the importance of lignin for water transport and mechanical strength in plants and in particular highlights that lignin may play various physiological functions depending on tissue and cell type [58]. More recently, the importance of interacting transcription factors regarding lignin formation and drought tolerance has been demonstrated [64][62].

There are a variety of proposed mechanisms by which lignin may confer drought tolerance to plants; yet, due to the great complexity of plant cells and the high evolutionary variability among plants, there is no consensus on its function in drought tolerance. It is frequently proposed that more lignified plant cell walls increase hydrophobicity and thus acts as a barrier against the evaporation of water to the environment, which may be extremely valuable for both productivity and survival during severe drought. Additionally, the increased strength conferred to the plant cell walls via lignification is also proposed to protect vulnerable plants under drought stress. Increased cell wall strength may also prevent cavitation resulting from hydrostatic pressure imbalances during drought. Increased lignification could be the result of alternative mechanisms for the regulation of ROS that are produced during drought stress and cause toxic oxidative stress. Changes in lignin composition, in particular changes in monomer unit concentrations will influence cell wall properties such as hydrophobicity and rigidity, and thus could impact drought stress tolerance of plants. Thus, plants may be adapted to alter lignin composition in response to

drought, either in addition to lignin content or irrespective of lignin content, altering the properties of the cell wall to confer drought tolerance. In particular, guaiacyl units have been related to the activity of water transport and could be selectively promoted as a drought tolerance mechanism, but the evidence is not strong in this regard [57]. While most studies that considered lignin composition in relation to drought stress have focused on the relative proportion of monomer units, other structural features such as inter-unit linkages, molecular weight, and degree of branching should also be considered.

Given the compiling evidence, it appears increasingly likely that lignin plays a role in the drought stress response of many plants. As work continues in this area, the nature of this mechanism will be gradually revealed. It is likely that lignins role in drought tolerance varies among different plants, tissues, cell type, stage of development, and severity of drought. Future studies should consider strategies to cover this variability in their approach. In addition, it is important to connect physical characteristics with molecular level responses to consider the results in full context. Larger scale investigations can use data-based approaches considering a wide range of variables to more comprehensively examine these mechanisms, including identifying various interacting variables. This approach may also provide clues for more focused research in the area. The field of bioinformatics already takes this approach for efficiently narrowing down key areas of interest given large and complex data sets. Given the extreme complexity and variability of plants, data-based approaches will likely be necessary as we look to further understand these multiscale problems. Chemometric approaches, which are described in detail in Chapter 5, may play a key role in this regard as they involved the application of mathematical and statistical tools to chemistry and are very suitable for problems that involve large data sets.

Chapter 3: Lignocellulosic biomass fractionation

3.1 Introduction

Efficient fractionation of lignocellulosic feedstocks into its respective components is a top priority for the subsequent conversion to energy, materials, and chemicals. In this thesis, the term fractionation is used generically to describe the approaches, typically occurring as one of the first processing steps, used to: separate lignocellulosic feedstock into their main components, to isolate a desired component, or to remove structural and composition impediments enabling further processing into value-added products. Thus, fractionation is used in a similar sense as the also common term "pretreatment" but is preferred for its generality. The exception to this definition would be when the term fractionation is applied to lignin specifically, in which case fractionation would indicate the breakdown of lignin macromolecules into smaller lignin fractions or monomer units. Lignocellulosic biomass fractionation is well studied and there are a wide range of methods due to the economic significance of this step in any commercial lignocellulosic industry. The fractionation step is typically one of the first steps in most biomass processing strategies and often considered the costliest step [65]. Additionally, the fractionation approach will have significant implications for the subsequent processing or analysis and thus is a particularly important area of research.

The complex and recalcitrant assembly of the lignified plant cell walls poses significant challenges for accessing the components and conversion into valuable products. There are a wide variety of methods that have been developed to breakdown the plant cell wall matrix and access the components. Fractionation strategies use physical/mechanical, chemical and biological treatments, alone, in sequence, or in combination. Physical treatments could include milling or the manipulation of variables such as temperature, and pressure. Chemical treatments use chemicals to target both covalent and non-covalent bonds to selectively breakdown the plant cell wall matrix while biological treatments employ microorganisms to selectively breakdown the plant cell wall. Different combinations of these approaches can be favourable for tuning the fractionation processes and the vast majority of fractionation strategies are physiochemical treatments. All these methods offer certain advantages and limitations depending on the mechanism by which they fractionate the plant cell wall matrix. Lignin is also degraded to some extent during lignocellulosic

fractionation, primarily through cleavage of its most common linkage type ($\beta - O - 4$) along with other more accessible and vulnerable bonds. In addition, lignin often participates in inconvenient condensation reactions during treatment which involves lignin fragments forming strong carbon-carbon bonds with itself, reducing its reactivity, and rendering the end product less desirable for further processing.

In this section, the main lignocellulosic biomass fractionation strategies will be reviewed with a focus on the impact of the treatment on the chemical structure of lignin. As interest in lignins role in contemporary biorefineries increases, more focus will be placed on the impact of these fractionation strategies on the composition of the lignin stream. In order to provide a structured overview of fractionation strategies, the methods discussed in this section will be categorized based on their primary application. Based on this, there will be two broad categories: commercial processes and laboratory methods. Commercial processes include the fractionation approaches intended for larger industrial processing while laboratory methods include the methods that are used to isolate lignin for further analysis. Laboratory methods typically focus on preserving lignins structure from its native state during the extraction process which contrasts with commercial methods that are more concerned with processing objectives.

3.2 Commercial fractionation

Commercial fractionation can be further sub-divided into two main categories: conventional processes and emerging methods. Conventional processes are those processes that are used in commercial pulping mills and bioethanol approaches. Emerging methods include the less developed methods that have been proposed as promising alternatives for conventional biomass fractionation approaches.

3.2.1 Conventional processes

3.2.1.1 Pulp and Paper

The three main industrial pulping processes include the kraft, soda, and sulphite processes. These processes generally operate via the dissolution of lignin at high temperature and pressure with wet chemicals [66]. The typical objective of pulping processes is to remove as much lignin as possible without reducing the strength of the fibres; however, not all lignin is removed with about 3-10% by weight remaining in the pulp [67] [68]. The kraft process is the most common

process for paper pulp production in the world, accounting for 80% of the total chemical pulping industry [22] [69]. The kraft process employs a $Na_2S/NaOH$ mixed solution (white liquor) to treat biomass material at temperatures of 155-175℃ over several hours [22]. In a conventional kraft process the white liquor and wood chips would be reacted in a large heated pressure vessel referred to as the digester [70]. This process degrades the lignin-polysaccharide linkages and leaves lignin fragments dissolved in solvent forming a lignin-rich fluid referred to as black liquor [71] [22]. The soda process is very similar to that of the kraft process, with the difference being that the lignins separated from the black liquor with Na_2CO_3 and Na_2SO_4 are named soda and kraft lignin, respectively [71]. Therefore, the soda process offers a sulphur-free alternative to the kraft process. Using sulphur in the treatment procedure is sometimes also used as a criterion for the classification of extraction processes. After extraction, the lignin is then burned in the recovery boiler producing steam for the process of recovering some inorganic chemicals [72]. It is suggested that this inefficient usage of lignin in the kraft process may change as the market value of lignin increases and valorization techniques develop [70].

Sulphite lignin is recovered by the sulphite pulping of biomass material (see Figure 3). The sulphite process utilizes an aqueous solution of SO_2 and a sulphate salt to cleave the lignin-polysaccharide linkages at temperatures of 140-170℃ [22]. Lignosulphonates can then be formed using a variety of bases, which are different for various pulping processes [71]. A comparison of the production pathways for kraft and sulphite lignin is shown in Figure 3. The resulting lignosulphonates are isolated by precipitation using alcohol, dialysis and a variety of other methods [71]. The harsh conditions used in these three main pulp and paper industrial processes produce extracted lignin that is severely degraded from its native structure. Soda lignin processes are typically used on annual fibre feedstocks, while kraft and sulphite lignin processes are mostly used on wood [22]. The increased development of the kraft process has led to a significant decrease in the sulphite lignin production because of the versatility of the kraft process [73]. Additionally, the efficiency of the sulphite process is dependent on the absence of bark and the type of raw material [74]. This has influenced the versatility of the sulphite process and thus contributed to the dominance of the kraft process.

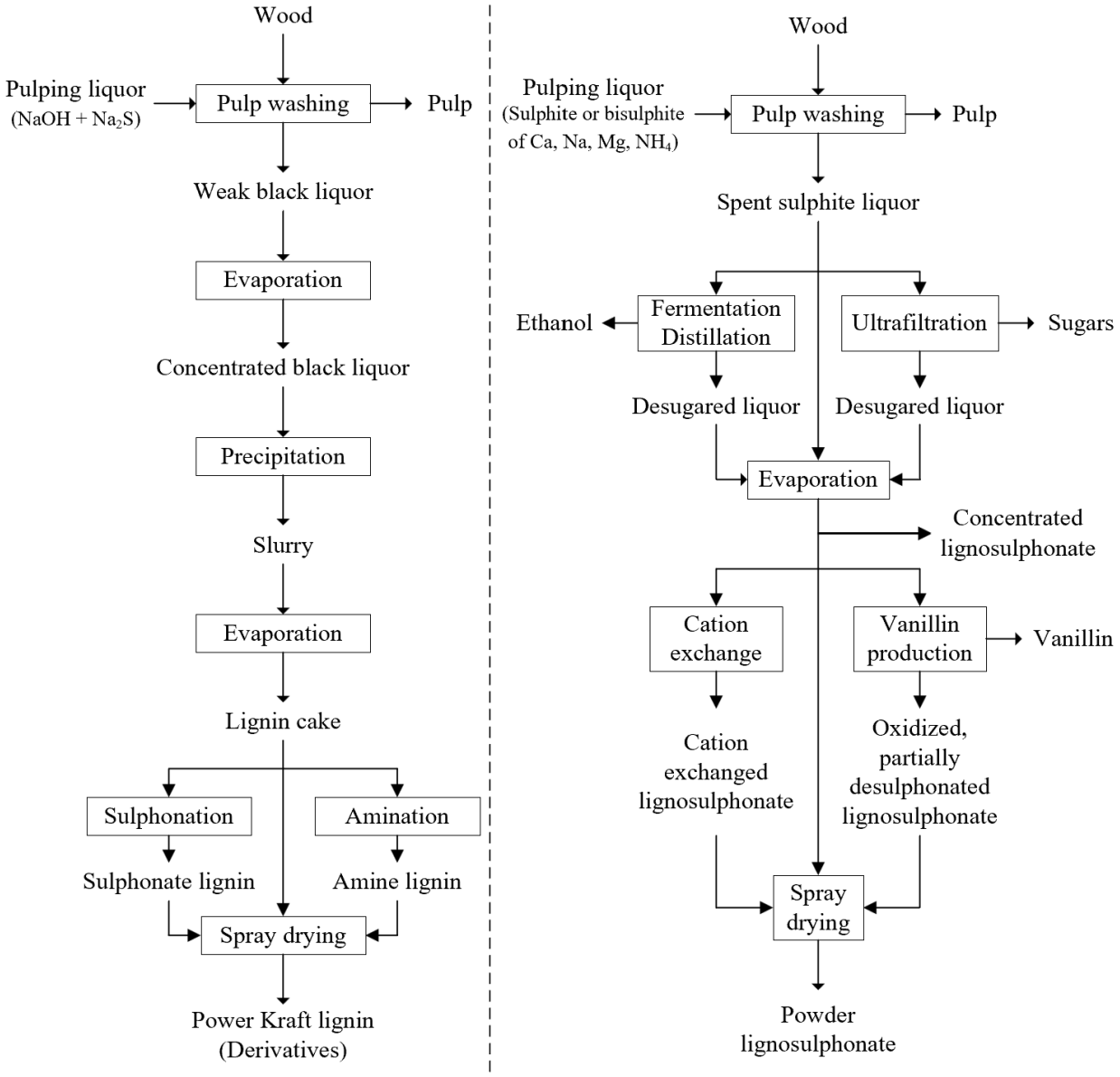


Figure 3. Production pathways for kraft (left) and lignosulphonate (right) lignin including selected modified pathways. Adapted from Lora [75].

The pulp and paper industry has processed lignocellulosic biomass at scale for many years. Therefore, there is opportunity to use the principles and infrastructure applied in the current pulp and paper industry to modern biorefinery designs. However, the harsh nature of the pulping processes may limit their application in future biorefineries. Additionally, the kraft and sulphite pulping processes release sulphur gases into the atmosphere, as a result some suggest that alternative extraction processes such as the organosolv process could be more beneficial in the

long term [76]. Table 2 summarizes the three main industrial pulping processes along with features of their lignin products.

Table 2. Summary of industrial pulping processes and their respective lignin products. Adapted from Luo et al. [77].

	Kraft Pulping	**Sulphite Pulping**	**Soda Pulping**
Generic Definition	Aqueous solution of sodium hydroxide and sodium sulphide (white liquor) reacted with wood chips in large pressure vessel called a digester [70].	Dissolution of lignin and release of cellulose fibres using bisulphite ions such as calcium, potassium sodium, magnesium, or ammonium [78].	Aqueous solution of sodium hydroxide (leaving out sodium sulphide used in kraft pulping) reacted with non-wood fibres [78].
Lignin Product	Kraft lignin	Lignosulphonate	Soda lignin
Lignin purity	High	Low-medium	High
Lignin Sulphur content (%)	1-3	3-8	0
Solubility	Water, Alkali	Water	Alkali
Separation techniques	Precipitation (pH change)	Ultrafiltration	Precipitation (pH change)

3.2.1.2 Bioethanol production

Bioethanol is a promising renewable energy source produced from feedstocks rich in carbohydrates such as sugars, starch, lignocellulosic biomass and algae, with lignocellulosic biomass being one of the most abundant and promising candidates [79]. Lignocellulosic biofuels are produced via fractionation of the cell walls into their three main components (cellulose, hemicellulose, and lignin), the subsequent hydrolysis of cellulose and hemicellulose into soluble sugars, and the fermentation of these sugars to ethanol. A schematic diagram of the general process for converting lignocellulosic biomass to ethanol is provided in Figure 4. The fractionation/pretreatment of the slurry is the main step for removing lignin from the fermentable sugars. There are a variety of different physical, chemical, physiochemical and biological approaches for the fractionation step in bioethanol production, many of which will be discussed in more detail in the emerging methods section. However, these approaches applied to bioethanol production should meet the processing objectives focusing on the yield of soluble sugars and would be less concerned regarding the quality of the residual lignin stream. Contemporary bioethanol production strategies may seek to balance the yield of soluble sugars with higher-quality lignin streams for value-added applications.

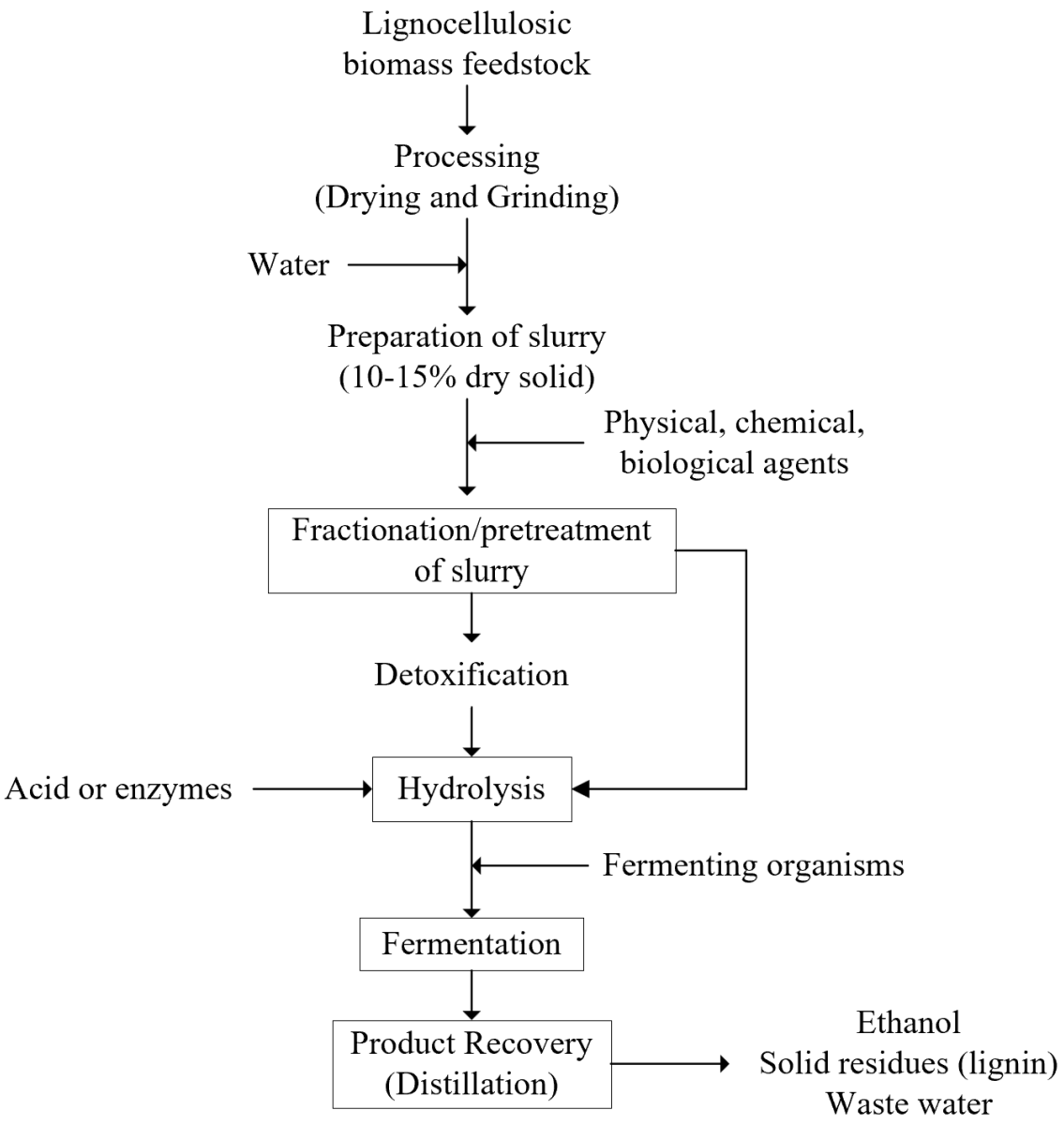


Figure 4. Schematic diagram overviewing the major steps in the conversion of lignocellulosic biomass into ethanol adapted from Zabed et al. [80].

3.2.2 Emerging methods

While there are many mature fractionation processes developed in the pulp and paper industry, there are a wide range of emerging methods under development that offer advantages compared to the conventional methods. Conventional fractionation approaches have concentrated

primarily on the value of the carbohydrate components, leaving residual lignin severely degraded and irreversibly altered, greatly reducing its potential for further upgrading into valuable products. Many of the emerging methods are modifying their fractionation strategies to avoid using sulphur and to realize higher value from all lignocellulosic components. There are a diverse range of emerging approaches with different advantages and limitations. Here we will highlight some of the main emerging fractionation strategies.

From the point of view of process chemistry and the resulting lignin product, most fractionation approaches can be generally classified as either acid- or alkali-based processes [81]. Hydrolysis is the process in which water is used to cleave chemical bonds [82]. Acid-Hydrolysis (AH) is a well-known approach for fractionating lignocellulosic biomass typically involving the use of a dilute acid to disturb the crystalline cellulose structure and hydrolyse lignin and polysaccharides into smaller fractions [83]. In particular, acidic treatment attacks hemicelluloses enabling the enzymatic hydrolysis of cellulose [84]. Acid hydrolysis of lignin primarily degrades the $\beta - O - 4$ linkages producing so-called Hibbert ketones, the resulting lignin possesses high relative contents of carbonyl groups and saturated aliphatic structures [81]. In contrast to acid, alkali-based fractionation uses various alkaline compounds (commonly NaOH/ammonia) to disrupt the lignin structure making the carbohydrate fractions more accessible [85]. Alkali pretreatment is generally performed under less severe conditions and is more effective with lower lignin content [86].

Steam explosion (SE) is a developing fractionation method that uses steam (150-230℃) and high pressure (1.38-3.50 MPa) for short periods (1-20 minutes) and then a sudden release of pressure to produce a water-insoluble lignin extracted from biomass [22] [87]. It is the most extensively studied physiochemical method of biomass pre-treatment and is sometimes referred to as “autohydrolysis” as a more accurate description of the chemical changes that occur during the steam explosion pre-treatment method [85]. It is reported that the most significant change to the supramolecular biomass complex resulting from steam explosion is related to the separation of hemicellulose and allowing the access of enzymes to the cellulose fibrils [65]. Typically, a small fraction of lignin will remain in the soluble phase and will require alkaline delignification for complete fractionation [73].

Liquid hot water (LHW) is a fractionation method akin to SE except is employs liquid water (pH 4-7) at high temperatures (106-230℃) and pressure (up to 5 MPa) instead of steam [88] [89] [90]. The treatment dissolves hemicelluloses into oligosaccharides and partially dissolves lignin enabling enzymatic access to cellulose [89]. LHW fractionation operates at a controlled pH to reduce the formation of monosaccharides which can degrade into toxic by-products. LHW treatment does not require a release of pressure like SE, it does not need a catalyst or chemicals, and toxic by-products are low, however, the large amount of hot water makes the process highly energy intensive [90].

Ammonia fibre explosion/expansion (AFEX) [91] fractionation uses high temperature (60-120℃) and pressure (up to 11 MPa) and ammonia to disrupt the cell wall structure by reducing cellulose crystallinity, partially degrading hemicellulose, targeting acetyl groups, and some lignin bonds to support the access of enzymes to the cellulose for digestion [92]. Similar SE, AFEX utilizes a rapid release of pressure. AFEX has been shown to target lignin-carbohydrate linkages while preserving lignins structure [93]. AFEX is more appropriate for lignocellulosic feedstocks with lower lignin content [90]. Ammonia must be recovered and reused for the process to be practical.

Ionic liquid (IL) fractionation can isolate lignin from biomass in a selective and environmentally friendly way; however, the procedure is still in development and can come at high costs with some challenges in the separation of products after pre-treatment [22]. Ionic liquids are organic salts with melting points under 100℃ composed entirely of organic cations and organic or inorganic anions with desirable properties negligible vapor pressure, non-flammability, and thermal stability among other characteristics [94] [95]. While IL extraction is still in its development stages, it is proposed that a possible mechanism for their degradation of biomass is related to the weakening of intra-molecular hydrogen bonds of the cellulose causing dissolution [95]. It is reported that lignin recovered from IL extraction have similar properties to that of organosolv lignin (described later in this section) and therefore IL extracted lignin might be suitable for similar applications [73].

Deep eutectic solvents (DES) have emerged as an environmentally friendly alternative to IL's with many shared properties. DES are eutectic mixtures of Lewis or Brønsted acid and bases

containing a range of anionic and/or cationic species while IL form from systems of primarily one type of discrete anion and cation [96]. Research on DES is still in its early stages and the exact mechanism of fractionation is still not fully understood; however, DES are of particular interest for lignin valorisation strategies due to the high solubility of lignin in DES. DES are easily synthesised, and their low cost, low toxicity, and biodegradable properties are favorable for the 'green' fractionation of lignocellulosic biomass. Challenges such as thermal instability and susceptibility to contaminants have limited DES in practice [97].

Derived from the principles employed by the pulp and paper industry, the organosolv extraction process is based on extracting lignin and hemicellulose from biomass using organic solvents [73]. Aqueous organic solvents along with some additives or catalysts at temperatures of around 140-220℃ break down the aryl ether linkages of lignin by hydrolytic cleavage and leave the fragments dissolved in the solvent [22]. The benefits of the organosolv method include its environmentally friendly conditions (sulphur free) and its high-purity product with a lower severity compared to traditional pulping methods, making it likely to have less chemical/structural alterations [22] [98]. These benefits have made it a promising candidate for biomass utilization, however, at the current stage of development the high uncertainty in capital investments [99] [100] [73] required to implement the organosolv process at an industrial scale have prevented the process from reaching commercial application and it currently remains in at the pilot/demo scale [98].

3.3 Laboratory fractionation

The previous sections concerning lignocellulosic biomass fractionation have focused on methods used for commercial purposes. These strategies are aimed at the efficient processing of lignocellulosic components to meet various downstream processing objectives. In this section, lignocellulosic fractionation methods used in the laboratory settings will be reviewed. In particular, methods aimed at quantification and analysis of lignin will be considered. These approaches can be essentially divided into two distinct categories based on their research objective: quantitation of lignocellulosic components as determined via the isolation of the components and evaluation of the fractions, or the strategic isolation of the components for further analysis. While the former approach is focused on effective fractionation of the components, the latter approach is more concerned with the mechanism of fractionation and the preservation of the lignin structure. This is

because the lignin structure can be significantly altered during extraction from the plant cell wall matrix depending on the mechanism and severity of the fractionation method. Thus, laboratory methods for quantifying lignin are generally quite severe to ensure effective separation of components, while laboratory methods for further lignin analysis are generally quite mild.

3.3.1 Quantitative analysis of lignin content

There is an assortment of methods and variations for the quantitative analysis of lignin content by fractionation, but the three main techniques include the Klason/acid-insoluble lignin, Van Soest, and acetyl bromide methods. These methods do not produce lignin suitable for further analysis because of the harsh nature of fractionation that severely alters the lignin structure. It is also important to note that while these methods are often used as the primary means of assessing lignin content of lignocellulosic samples, they all have various issues and limitations which include requiring toxic chemicals, relatively large sample amounts (300 mg), very laborious procedures, and mechanistic differences leading to varying results [101]. Thus, often the lignin content determined using these techniques is better understood as a relative measure. Consequently, lignin content is better reported as prefixed by the method of analysis.

The Klason method is considered as the standard for determining lignin content and is the most well-known lab-scale method [22] [102]. Biomass is treated with 72% H_2SO_4, followed by dilute H_2SO_4, leaving lignin as an insoluble residue after the hydrolysis of cellulose and hemicellulose [102]. This residue is dried, weighed and reported as lignin content. Limitations of the procedure result from the potential for alternative components such as proteins condensing and ending up in the 'lignin' residue, covalently linked lignin and carbohydrates ending up in the wrong fraction, and some acid-soluble lignins remaining in solution [103]. Acid-soluble lignins can be corrected for using spectroscopic techniques applied to the filtrate from the second stage of the Klason process by relating the intensity of the spectrum at a characteristic lignin wavelength to lignin concentration via Beer's law [104]. Klason lignin is highly condensed and altered, therefore deemed unsuitable for chemical characterization [66]. The Klason-type procedures can also estimate cellulose and hemicellulose components from the difference, as lignin and ash components are removed. Ryan et al. have compared two Klason-like techniques for determining the approximate carbon fractions of the three main components of biomass - forage fibre

techniques [105] [106] and forest products (see Figure 5) techniques [107]. Ryan et al. credit Van Soest [105], and Goering and Van Soest with the development of the forage fibre analysis [106], and TAPPI and Effland with the forest products techniques [108]. Ryan et al. broadly divide the organic compounds from biomass into three categories: extractives which include nonpolar compounds such as fatty acids and lipids and polar compounds such as sugars and phenolics, polymer carbohydrates or holocellulose (cellulose and hemicellulose), and acid-insoluble aromatic compounds (lignin) [107]. The forage fibre analysis and forest products techniques work by separating the biomass components into different categories. In the case of the forage fibre analysis an estimate of the ash free (acid – detergent) lignin is found and other extractives such as cellulose and hemicellulose are determined by the difference [107]. The forest products technique is similar in method but is more complicated. Goering and Van Soest describe the neutral-detergent fibre (NDF) and acid-detergent fibre (ADF) analytical procedures for estimating the lignin, cellulose, and hemicellulose contents in plant material. The NDF procedure for cell-wall constituents is a fast method for analysing total fibre in vegetable feeds, while the ADF procedure can provide a rapid determination of lignin and cellulose in feeds. Taking the difference between the ADF and NDF allows the estimate hemicellulose content [106].

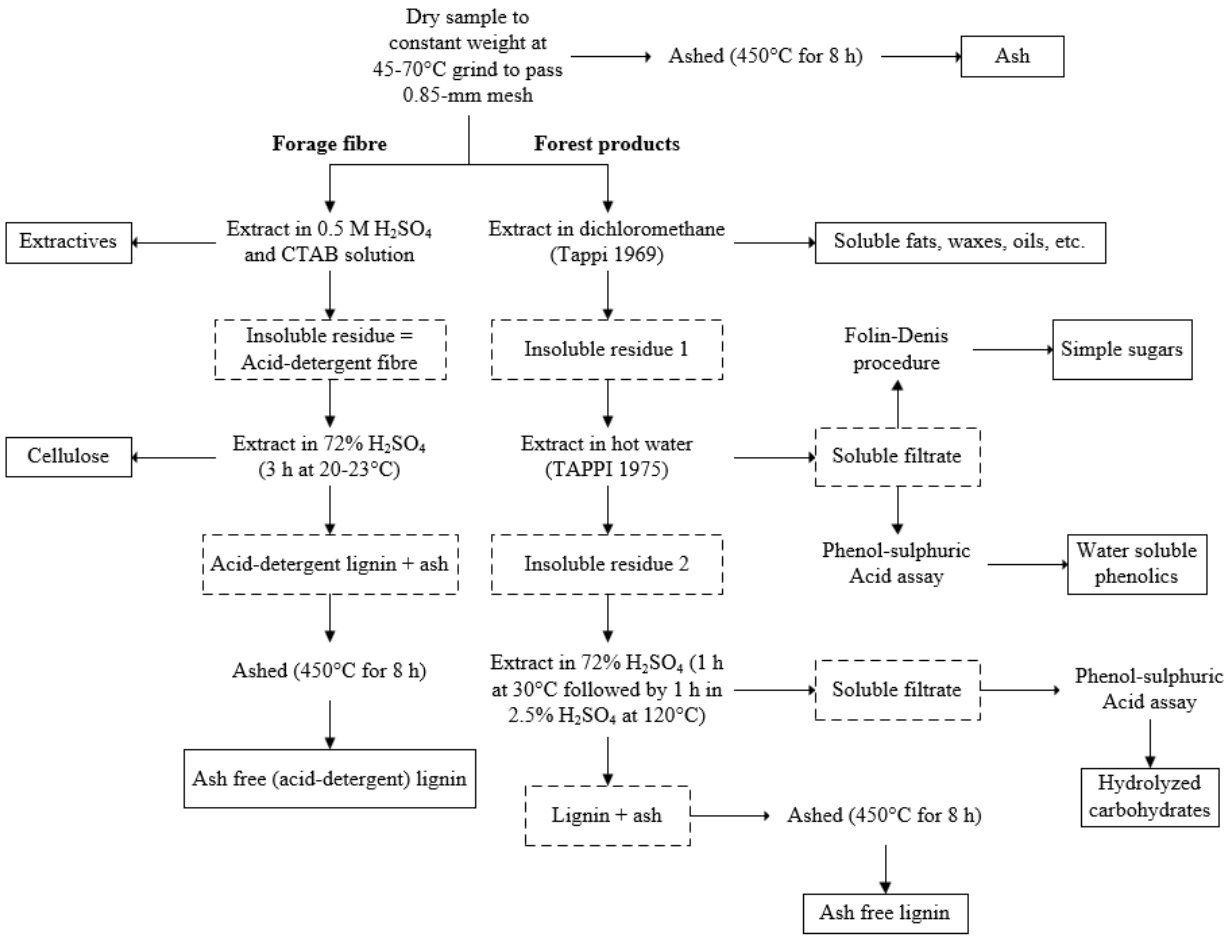


Figure 5. Flow diagram for the forage fibre and forest products determination of proximate carbon fractions, where the cationic surfactant cetyltrimethylammonium bromide solution is labelled as CTAB. Adapted from Ryan et al. [107].

While not necessarily a fractionation method, the acetyl bromide method is one of the most common approaches for quantitatively assessing lignin content from lignocellulosic material and thus will be described here. The acetyl bromide method is the most popular of a number of spectroscopic-based methods for quantitative lignin content analysis. It works via the dissolution of a sample in a solvent and subsequent measuring the absorbance of the solution at a distinct wavelength characteristic of lignin of which a previously developed calibration curve can be used to estimate the lignin concentration in the solution and thus the lignin content of the sample [104]. The acetyl bromide method has the advantage of rapid analysis time compared to its laborious counter parts (e.g., Klason, Van Soest), it requires only milligrams of sample, and should in theory

account for all the lignin present in the sample, including the acid-soluble lignin. However, since the procedure relies on a calibration curve it is an in-direct measure.

3.3.2 Lignin extraction for analysis

In contrast to the methods described in the previous section with the objectives of accurately determining the lignin content of lignocellulosic materials, this section will discuss the methods used to isolate research grade lignin. Detailed investigations of lignins structure require pure and unaltered lignin to be extracted from the other lignocellulosic components of which it is entangled with inside the plant cell walls. Various techniques have been developed for this purpose; however, it is not so clear to what degree these approaches truly preserve the lignin structure. Thus, like the quantitative lignin content determination methods, often the structural analysis performed on isolated lignin is considered relative to the isolation procedure. Yet, inferences regarding the structure of native lignin are often extrapolated from the analysis of isolated lignin, despite there being some degree of alteration from extraction. The terms “in situ lignin”, “native lignin”, “protolignin” and “technical lignin” are widely used in the lignin extraction literature. In situ, proto- and native lignin refer to unextracted lignin that is inside the plant cell wall, as found in nature. Technical lignins are those that have been obtained as a result of lignocellulosic biomass processing, and therefore, technical lignins are prepared from natural lignin [22]. Note that it is also common for technical lignins to be named after the extraction process used to derive them (MWL, kraft lignin, organosolv lignin, etc.).

The most common laboratory technique for isolating lignin from wood is done via the classical Björkman method, which uses extensive grinding, followed by solvent extraction (see Figure 6) [109] [102] [110]. The product of this procedure is called a milled wood lignin (MWL). If a larger amount of solubilized lignin is preferred, the Björkman process can be modified either by treating the finely ground wood with hydrolytic enzymes for the removal of associated polysaccharides prior to solvent extraction or by treating the insoluble material from the first dioxane extraction with the enzymes to remove carbohydrates and subsequently solubilizing the material in dioxane [111]. The products of the altered Björkman method are referred to as cellulolytic enzyme lignin (CEL) [110]. The altered Björkman method yields increased lignin and therefore the yield of CEL is greater than that of MWL. This increased yield is considered to be

lignin associated with carbohydrates that may be participants in LCC's. The additional step in the CEL process is able to release more of the lignin associated with carbohydrates. The CEL has been shown to be preferred over MWL, as the lignin yield is higher and the chemical degradation is decreased [112]. However, the altered Björkman method that produces CEL is more tedious than the classical Björkman method that produces MWL, as it adds an additional step and time to the procedure [113]. Refining the CEL process and including a mild acidolysis step can yield another modified product called enzymatic/mild acidolysis lignin (EMAL) [22] [114]. In the EMAL process, the mild acidolysis step is used to cleave the remaining lignin-carbohydrate bonds to produce high yield and purity lignin [114].

The Brauns' lignin isolation process extracts lignin with organic solvents (EtOH/dioxane) which are then removed by evaporation [113]. The Brauns' procedure is suggested to cause such slight chemical structure changes to the sample that the product is referred to as native lignin, however, yields of 2-10% and contamination by extractives can make this method undesirable for certain applications [104] [113].

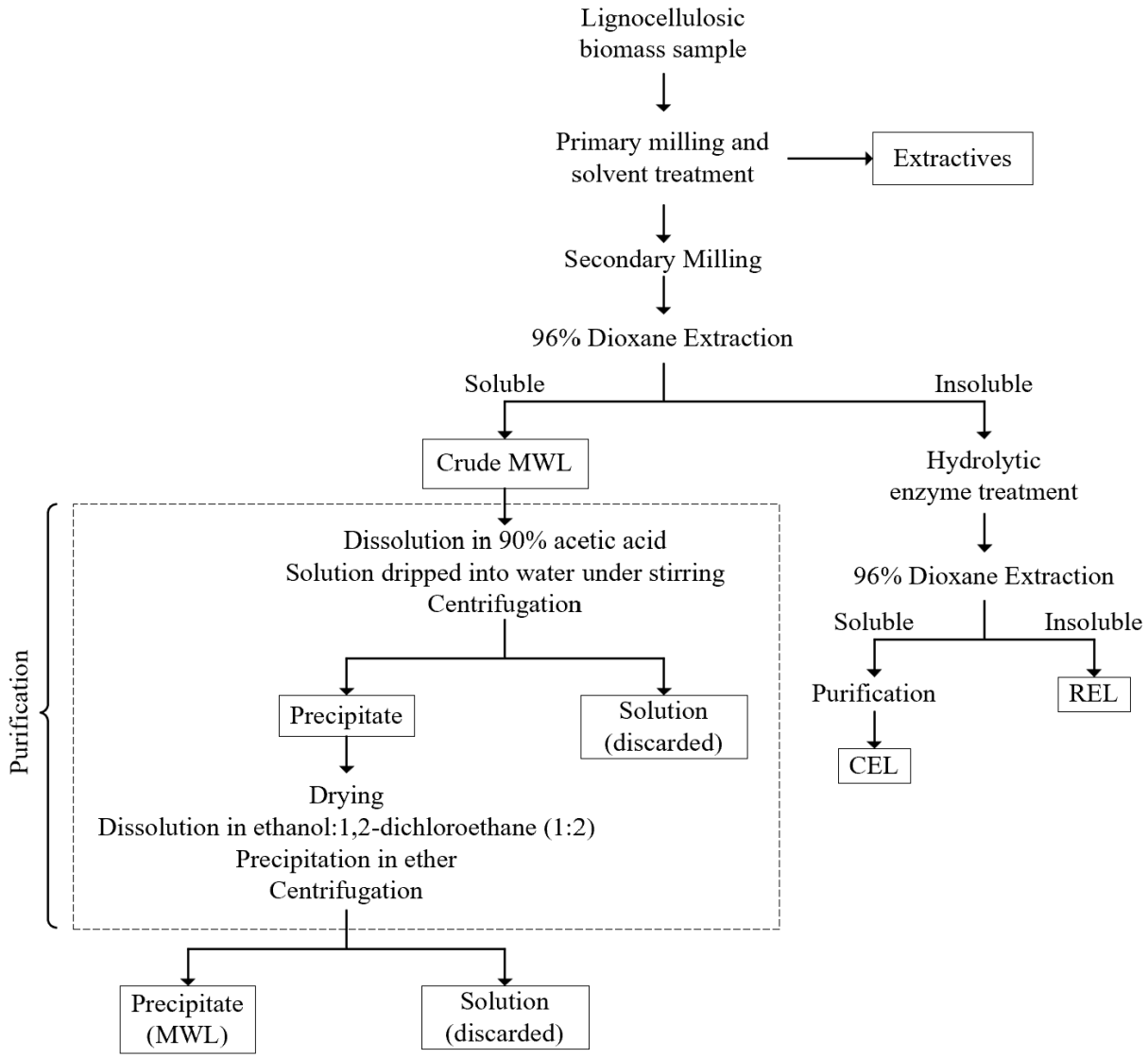


Figure 6. The MWL and CEL isolation procedures. Adapted from Brunow et al. [115] and Holtman et al. [116].

Hydrolysis is the process in which water is used to cleave chemical bonds [82]. Hydrolytic lignin extraction can take many forms; however, typically hydrolysis extraction methods utilize an acid, base, or enzyme to dissolve the cellulose and hemicellulose components leaving lignin as an undissolved residue [117] [22]. Since hydrolytic lignin is a broad category that can refer to multiple extraction techniques, the term is somewhat loosely used in literature which can lead to confusion. Hydrolytic lignin extracted from residues of the cellulosic ethanol production is commonly known as enzymatic hydrolysis lignin (EHL) and is usually used as a low value fuel to generate steam or

electricity [71]. The EHL can have minimal structural changes, however, EHL typically contains 5-15% carbohydrate impurities [118]. While these impurities can cause problems for analytical procedures, NMR methods have been reported to overcome these challenges [118]. Prior milling of samples, which significantly degrades the polysaccharides along with some breakdown of the lignin macromolecule, is a requirement of enzymatic hydrolysis of wood [118]. It is also important to note that enzymatic hydrolysis can also be used to investigate LCC linkages [118].

While there may be altered procedures depending on application, Lundquist [119] defines acidolysis as specifically referring to treating lignocellulose or lignin with 0.2 M HCl in dioxane-water (9:1, v/v) at reflux or heating at 100℃ [119]. The procedure has initially been applied in the field of lignin chemistry by Pepper et al. [120] to isolate lignin from plant materials [119]. The acidolysis procedure can also be applied as a tool for the characterization of lignin and its components [121]. It is reported that acidolysis lignin has only minor carbohydrate impurities and less severe chemical degradation of the native lignin than other methods [22]. Thioacidolysis (TA), where ethanethiol is used as a solvent instead of water, is proposed to provide a more effective degradation and an increased yield of lignin [22]. The TA procedure has been used to estimate lignin monomer composition [122].

Chapter 4: Lignin characterization

4.1 Introduction

Challenges to the efficient conversion and utilization of lignocellulosic biomass have motivated researchers to investigate the chemical structure of lignin. It is proposed that a greater understanding of the lignin structure will assist in the utilization of lignin as an abundant renewable resource. Moreover, elucidating the mechanism of interaction of lignin within the plant cell wall will provide insight into the degradation of raw lignocellulosic biomass into its main components. Lignin monomer composition and S/G lignin ratios have been demonstrated as important criteria for predicting recalcitrance of lignocellulosic biomass [101]. However, there is still some conflicting data on this topic with some literature suggesting higher S/G ratios are correlated with reduced recalcitrance [123] [124], and others suggesting higher S/G ratios are correlated with increased recalcitrance [125] [126]. In addition, the S/G ratio is not relevant for all lignocellulosic feedstocks, for example, softwood lignins are primarily composed of guaiacyl units and have only very minor concentrations of *p*-hydroxyphenol and syringyl units.

While many advanced methods for characterization of lignin exist today, much is still unknown about its structure. As described in the last section, lignin can only be studied after some mild extraction from the plant cell wall. Therefore, analytical procedures can only provide information on the extracted technical lignin, leaving the native lignin structure a key point of debate. The detailed description of lignin structure requires multiple pieces of information, including elemental composition, relative proportion of monomer units, linkages, functional groups, molecular weight distribution, and degree of branching of the macromolecular structures. As a result of this complexity, multiple analytical methods are required to obtain a detailed description of the structure of lignin. The chemical structure of lignin is dependent on a multitude of factors including the type of biomass material, part of the plant, location of tissue, cell type, age, and growth conditions [22]. Further, structural characteristics of technical lignin are influenced by the degradation process, extraction technique, and sometimes the analytical method used [22]. This section provides an overview of the main analytical methods used for the structural characterization of lignin.

4.2 Wet Chemical analysis

Wet chemistry is a type of analytical chemistry approaches that use observation, typically in the liquid phase, for chemical analysis. Wet chemical approaches have been vital in supporting our current understanding of lignins structure today. They have been used to detect functional groups such as carbonyl, carboxyl, aliphatic hydroxyl, phenolic hydroxyl, methoxy, condensed units, free phenolic groups as well as the main lignin monomer units [101]. Functional group analysis and degradation techniques allow the precise determination of lignin features but can be indirect, time-consuming, error prone, and too specific [127]. In-fact, wet chemical approaches in general are time-consuming, tedious, often require toxic reagents. Thus, while these conventional approaches have been fundamental, high throughput spectroscopic analysis in becoming increasingly preferred.

Wet chemical methods such as nitrobenzene oxidation (NBO), TA, permanganate oxidation (PO), and derivatization followed by reductive cleavage (DFRC) are well-established techniques and have been used extensively in literature [22]. NBO has been used for years as a procedure for examining the structure of lignins since its introduction in 1939 [128] [129]. In the NBO process, the non-condensed *p*-hydroxyphenol, guaiacyl, and syringyl units in lignin are oxidatively cleaved forming aromatic carbonyl compounds, such as phenolic aldehydes and phenolic acids [130] [131] (Figure 7). Lapierre [29] reports that NBO has a high uncertainty (standard deviation of 20 to 30%) on the yield of monomeric products and speculates that this low interlaboratory reproducibility arises from variation in reaction duration, temperature, and analytical complications [132]. The S/G ratio determined by NBO achieves a higher interlaboratory reproducibility (standard deviation 4 to 8%) but predicts a higher S/G ratio than is thought to be accurate, with this characteristic being attributed to the syringyl units being less involved in condensed inter-unit bonds as compared to the guaiacyl units [29]. The NBO method is reported to be less complex comparatively to PO in terms of taking less time and requiring fewer steps [129]. The PO method has provided significant contribution to the field of lignin chemistry [133]. The PO reaction selectively degrades aliphatic side chains attached to the aromatic moieties of lignin; a mixture of aromatic carboxylic acid structures remains as the product [133]. Information on the absolute phenolic group content of and relative proportions of individual lignin substructures can be derived from the identity of the resulting acids, providing information on the

structure of a lignin sample [133]. A major drawback of PO is that only around 10-20% of the lignin units in native wood can be identified. This is because only phenolic groups that are initially free and then become methylated can be analyzed [133]. If the lignin aromatic alcohol groups on the C4 carbon are initially etherified, they would not be analyzed [29]. There are methods that can improve the yield of PO but these add additional steps to an already long and complex process [134]. Overall, despite its contribution to lignin research, PO is less commonly used recently because of its low throughput and complicated multi-step procedure [29]. An adaptation from acidolysis (described at the end of Chapter 4), TA is defined as solvolysis in dioxane-ethanethiol with boron trifluoride etherate [135]. It is an acid-catalyzed reaction that targets the $\beta - O - 4$ inter-unit linkages to depolymerize lignin enabling the determination of functional groups, interunit linkages, and information regarding the parent polymer [29]. In particular, TA can provide information on the type and quantity of lignin units involved in $\beta - O - 4$ bonds. Similar to TA, DFRC is an analytical procedure for lignin characterization that selectively cleaves the $\beta - O - 4$ linkages in lignin, producing monomers and dimers for analysis via GC or GC-MS [136]. The approach operates via three steps which include the acetyl bromide treatment, reductive cleavage of $\boldsymbol{\beta - O - 4}$ using zinc in an acidic medium, and acetylation of monomers followed by GC quantification of the monomers [137]. It has not only been used as a stand-alone procedure but also modified or combined with other techniques to provide additional structural information about lignin [138].

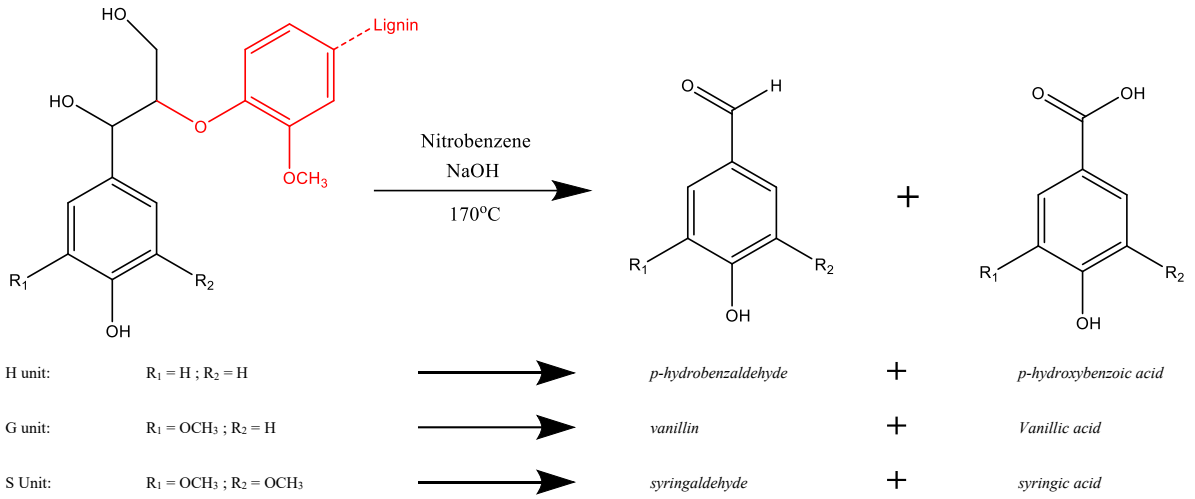


Figure 7. The major products of nitrobenzene oxidation of lignin adapted from Jiang et al. [131]. The NBO method provides the information of the monomeric composition and the condensation degree of lignin.

While wet chemical techniques can be precise in determining specific functional groups and structural moieties, each technique has its own limitations and does not provide a complete picture of the entire lignin structure [127]. Degradative wet chemical methods are indirect and would be best suited to provide information on the relative amounts of lignin and its monomer units. This is because the mechanism behind each wet chemical method may produce altered results compared to the values in native lignin, such as in the example of NBO discussed above where S/G ratio is overestimated because of the difference in inter-unit bonds between S and G monomer units. [101] Therefore, if more representative quantitative results are required, it is recommended to use multiple complimentary wet chemical and/or other methods [118]. The comparison and summation of information produced from these different methods would allow for a more informed determination of lignin structural features.

4.3 Vibrational spectroscopy

Spectroscopic methods (e.g., FT-IR, Raman, or near-infrared (NIR)) are useful for the analysis of the whole lignin structure (functional groups, linkage types, energy states of atoms) [139] and to identify lignin moieties essentially directly as compared to the wet chemical methods that require chemical modifications [22]. FT-IR spectroscopy is a non-destructive, rapid analytical technique applicable for both solid and liquid samples and is the most widely used technique for

the qualitative determination of functional groups [22]. In lignin sample analysis, FT-IR spectroscopy is challenged by strong bands of residual non-lignin compounds overlapping the characteristic lignin bands [22]. FT-IR spectroscopy can efficiently analyze lignin without sample pre-treatment [22] and enables the determination of lignin monomer composition, methoxy and carbonyl groups, and the calculation of the ratio of phenolic to aliphatic groups [140] [141]. For quantitative work, typically some variation of calibration is required. Different chemical groups may have shared absorption coefficients in the FT-IR measurement making the composition determination not absolute. Other advantages of FT-IR include its high signal-to-noise ratio and high optical throughput, which results in a high dynamic range of linearity, making it very apt for quantitative work [141].

Raman spectroscopy shares some spectral assignments of absorption bands of FT-IR spectroscopy but can provide complimentary information [22]. Raman spectroscopy had not been applied to lignin or lignocellulosic material until 1984, despite being a known technique for a long time prior [142]. This is likely the result of challenges with laser-induced fluorescence (LIF) which is a background signal that interferes with weaker signals in the Raman spectrum [142]. A central problem with LIF is the separation of the Rayleigh scattered exciting radiation of air molecules from the fluorescence signal. To suppress the scattered excitation radiation, excitation is conducted at higher energy levels than those of fluorescence, the detected wavelengths are shifted and can then be separated from the excitation wavelength using wavelength filters. However, since the development of near-IR FT-Raman spectroscopy – which has allowed the avoidance (for the most part) of LIF and significantly reduced time requirements – Raman has become a more common technique for lignin and lignocellulosic biomass characterization [142]. It can be used to determine lignin S/G ratios both directly and indirectly using multivariate analysis [101]. Perera et al. [143] have introduced a new Raman micro-spectroscopic method for analyzing the structure of native lignin. The authors recovered lignin spectra of poplar, *Arabidopsis,* and *Miscanthus* and detected structural differences using chemometrics (a field of techniques which will be discussed in more detail in Chapter 6.).

4.4 Nuclear Magnetic Resonance spectroscopy

The NMR spectroscopy provides unparalleled structural detail and is routinely used by researchers for lignin monomer analysis and S/G ratio determination [101]. It also provides a way

to accurately study the chemical structures and nature of chemical bonds in the lignin macromolecule, including determination of the S/G ratio [140]. NMR is a significantly less sensitive technique than the other typical spectroscopic methods because the relatively small difference between energy levels essentially makes the ground state and excited energy levels equally populated [144]. This lower energy characteristic of NMR and small/fast transition between ground and excited state reduces the potential for large conformational changes and prevents physical or chemical change to the molecules (or at least prevents visible changes) and thus provides it with a major advantage compared to other spectroscopic techniques [144]. Compared to what the structure would look like if it were not being observed, the NMR would see a slightly less perturbed structures in comparison to other spectroscopic methods [144]. Another factor influencing the sensitivity of NMR is the availability of the interrogated atom. If the relative abundance of a particular atom of interest is low, this can reduce the already low sensitivity of the technique. Despite the poor relative sensitivity of NMR, it can acquire a lot of information, often fully identifying compounds and resolving structure and bonding patterns even in complex molecules [144]. There are a large number of solid and solution state NMR techniques used for investigating specific structural information. When selecting a method suitable for a particular application, one will have to consider the information that they are expecting to gather from the experiment in order to select an appropriate NMR technique.

A remarkable example of one NMR technique in lignin research is the determination of hydroxyl groups by phoshorous-31 nuclear magnetic resonance (^{31}P NMR) spectroscopy [145] [146] [147]. The ^{31}P NMR spectroscopy enables the quantification of different types of hydroxyl groups, such as aliphatic, phenolic and carboxyl hydroxyl groups. These functional groups belong to the main lignin functionalities and therefore their quantification is important for the structural analysis of lignin. Since hydroxyl groups in lignin cannot themselves be detected by ^{31}P NMR (due lack of phosphorus atoms), they must be converted to the corresponding phosphites. In short, a common procedure for this analysis requires that the lignin be dissolved in a mixture of dichloromethane (DCM) and pyridine, in the presence of the internal standard triphenyl phosphorus oxide (TPPO) and a relaxant chromium (III) acetylacetonate ($Cr(acac)_3$), and then phosphitylated using a mixture of a derivatization regent (i.e. 2-chloro-4,4,5,5-tetramethyl-1,3,2-dioxaphospholane (TMDP)) and deuterated chloroform. The reaction of lignin hydroxyl groups

with TMDP replaces hydroxyl protons with a phosphorus-containing moiety, which are detected by ^{31}P NMR (as illustrated in Figure 8). The integration of signal response over different regions of the ^{31}P NMR spectrum allows for determination of phosphitylated hydroxyl concentration by group type (e.g., aliphatic, phenolic or carboxylic).

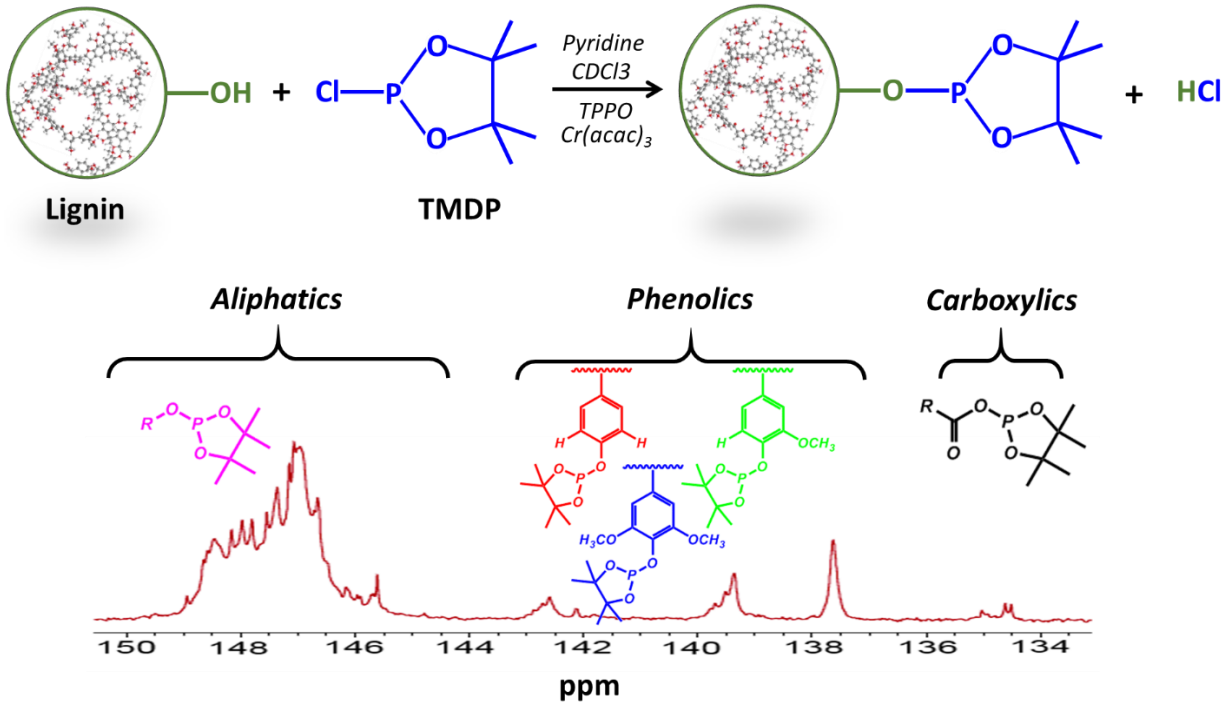


Figure 8. The scheme of the phosphitylation reaction of hydroxyls groups in lignin (top) along with the ^{31}P NMR spectrum of lignin derivatized with TMDP with identified regions (and structures) belonging to aliphatic, phenolic and carboxylic hydroxyl groups (bottom). Adapted from Meng et al. [145].

It is important to note that NMR cannot always provide quantitative information and so it is typically partnered with one or more quantitative methods. Compared to other spectroscopic methods, NMR has much higher resolution, allowing it to provide detailed information [127]. Depending on the technique employed, the sensitivity of NMR varies. To summarize, despite its insensitive nature, with sufficient sample quantities (around 10-100 mg) relatively detailed characterization of the structure of lignin can be derived using NMR [144].

4.5 Mass spectrometry

Mass spectrometry is an analytical tool that can determine exact molecular weights of sample components. Gas chromatography-mass spectrometry (GC-MS) is the most common

analytical approach for the determination and quantification of organic compounds in mixtures [148]. It is an extremely useful and highly accurate technique for confirming the identity and elucidating the structure of unidentified components in lignocellulosic biomass [149]. A conceptual illustration of the major components of the GC-MS system is shown in Figure 9.

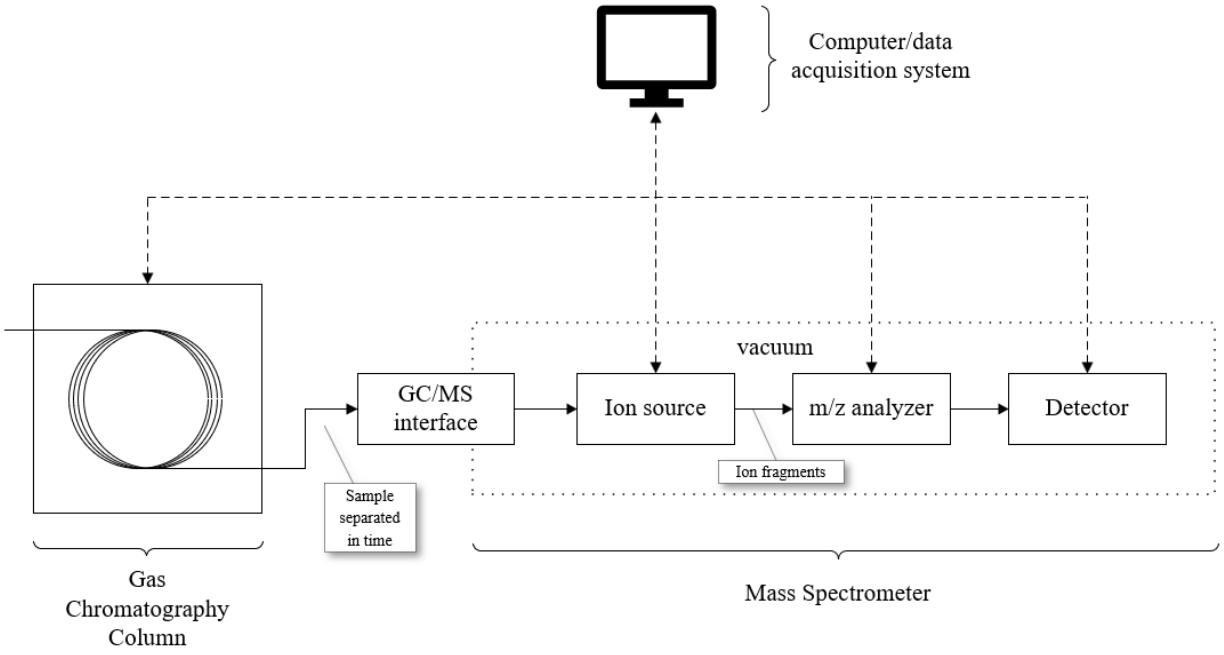


Figure 9. Flow diagram for the GC-MS system and its major components which include the Gas chromatograph, ion source, *m/z* analyzer, ion detector and the data acquisition system. Not shown are inlets to GC column (injectors, gas sampling valves, probe), the carrier gas phase, the CI reagent gas (chemical ionization gas), and direct inlet/direct insert probes for the ion source. Some elements of this figure are subject to changes based on certain procedural variables such as the type of mass spectral ionization. Adapted from Sparkman et al. [149].

As one would expect, the GC-MS system is comprised of a gas chromatograph, a mass spectrometer, and a computer data acquisition system [149]. The sample is volatilized, and gas chromatograph separates components using a selective stationary phase [150]. The separated compounds are ionized and then analyzed in the mass spectrometer [150] [151]. The mass spectrometer separates ions according to their mass-to-charge ratio (m/z) and the data produced from the GC-MS analysis are referred to as mass spectra [151]. These mass spectra of molecular ions and fragment ion peaks are used to determine the molecular weight and structural makeup of the corresponding compound [152] [153]. In cases of weak signals, complex mixtures, and or

strong background signals, before analysis, the mass spectra must be retrieved from the computer system and corrected by subtracting background spectra. Mass spectra are then used to identify components by matching with known spectra stored in computer systems [149]. It is important to note that the resulting structural assignments from GC-MS may require verification depending on the level of match and ionization method used [149]. Verification may include confirmation with the use of a standard of the compound using the same instrument employed in the complex sample analysis, or confirmation using an ancillary analytical method (e.g., IR, UV, and NMR) [149] [152]. Using the data collected from GC-MS it is possible to propose a potential macromolecular structure, as demonstrated by Lu et al. [154]. Analyzing the low molecular weight products from the pyrolysis of a complex polymer, such as lignin using GC-MS requires dealing with large volumes of raw data that requires both editing and interpretation [144]. Manual editing, identification, and assessment of mass spectra from GC-MS can be an extremely tedious task even for an individual experienced in the procedure [152]. Therefore, researchers are turning to mathematical algorithms and computer programs to assist with the analysis of large volumes of data [155] [156].

Gas chromatography (GC) is a very useful tool that concentrates on separating the organic species with lower polarity and lower boiling point, and has been used extensively in the characterization of lignin-related complex samples [22]. Recently, Py-GC/MS has become a common technique for lignin analysis, and can provide information on lignin content, molecular weight distribution, monomer composition, oligomer composition, frequency of linkages, and thermodynamic behaviors [22]. Using the Py-GC/MS method, a possible macromolecular structure for lignin can be proposed [154] [22]. Pyrolysis, separation and measurement occur within a single system, minimizing sample loss. Separated products can be identified using their mass spectra and insight can be gained into the initial macromolecular structure [157]. Figure 10 presents a schematic diagram of the analytical Py-GC/MS system including illustration of the sample probe where the lignin sample is located and pyrolyzed.

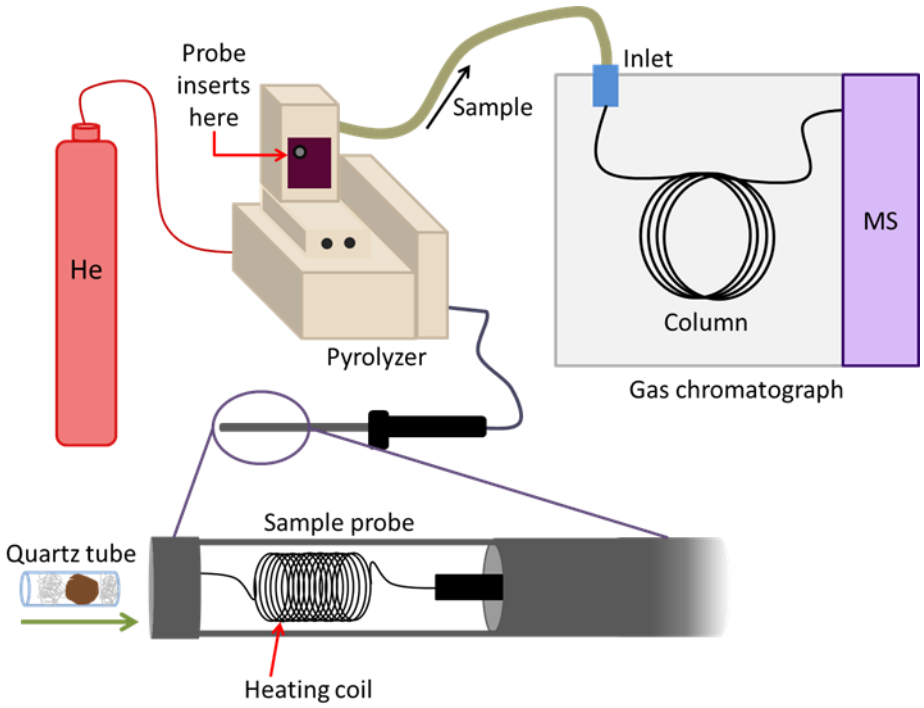


Figure 10. Schematic diagram of the pyrolyzer GC-MS system. Reproduced from Heshka et al. [157].

Py-GC/MS can provide a reliable technique for the analysis of lignin monomers, and more specifically the determination of S/G ratio. Pyrolysis lignin utilizes heat in the absence of oxygen to transform non-volatile compounds into volatile mixtures [158]. The mechanism of lignin pyrolysis is not understood because the violent conditions lead to multiple and parallel reaction mechanisms that have not been characterized [22]. However, lignin content and structure in lignocellulosic biomass have been evaluated by analytical pyrolysis. When conducting Py-GC/MS on lignocellulosic biomass samples, the degradation products of polysaccharides and lignin are separated by gas chromatography before being identified by the mass spectrometer. The mass spectrum of lignin is distinct and recognizable, even when analyzed in combination with carbohydrates, and therefore no pre-treatment/removal of carbohydrates is required before the analysis [158]. Fast analysis time and requirement of modest sample size makes this method favourable in structure investigations [113]. Other benefits include minimal sample preparation and no requirement of fractionation/pre-treatment [158].

Analytical pyrolysis techniques can be briefly described as: rapid heating (e.g. 0.1 to 20 °C/ms) the sample under inert atmosphere to an optimum temperature (700 °C) which results in

breaking up the complex, heterogeneous solid into smaller molecules that are characterized and measured (as demonstrated in typical pyrogram in Figure 11) [158]. Scholze and Meier [159] have compared five methods to elucidate the composition of pyrolytic lignin using Py-GC/MS, various wet chemical techniques, and FT-IR, and concluded that pyrolysis lignin have similarities to milled wood lignin. However, pyrolysis lignin does not require the same sample preparation requirement.

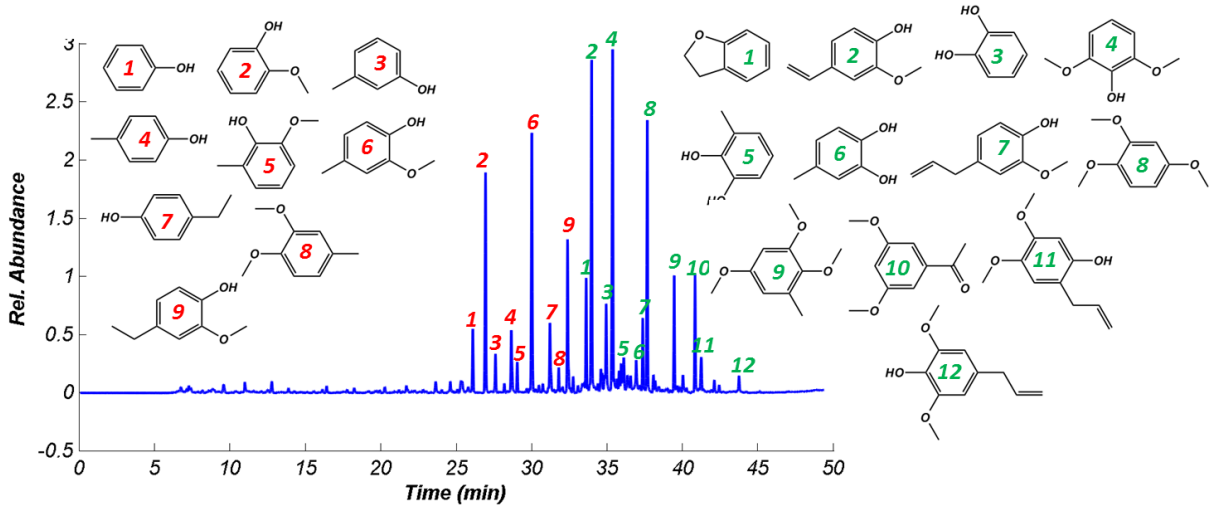


Figure 11. An example of Py-GC/MS pyrogram of softwood lignin along with identified pyrolysis products [160].

4.6 Size Exclusion Chromatography

Molecular weight is a key parameter of polymers with significant implications on their mechanical and thermodynamic properties. Molecular weight is often reported in terms of various average molecular weights in terms of arithmetic means of molecular weight distribution (MWD) because it is easier than characterizing the whole MWD of a sample. In size exclusion chromatography (SEC), a polymer sample is dissolved in a suitable solvent and injected into a column full of stationary porous particles. A mobile phase, which is typically the same solvent used to dissolve the polymer, carries the sample through the column. Higher molecular weight macromolecules in a polymer sample will exhibit larger hydrodynamic volumes and will be unable to access some of the volume inside the porous particles and thus will exit the column more quickly relative to the smaller particles which undergo more convoluted paths before they reach the end of the column. It is by this mechanism that SEC separates a polymer sample based on hydrodynamic

volume. Molecules will exit the column at different retention volumes based on their hydrodynamic volume. The profile of the separated components exiting the SEC can describe the molecular weight distribution of the sample. A schematic representation of the SEC mechanism is shown in Figure 12.

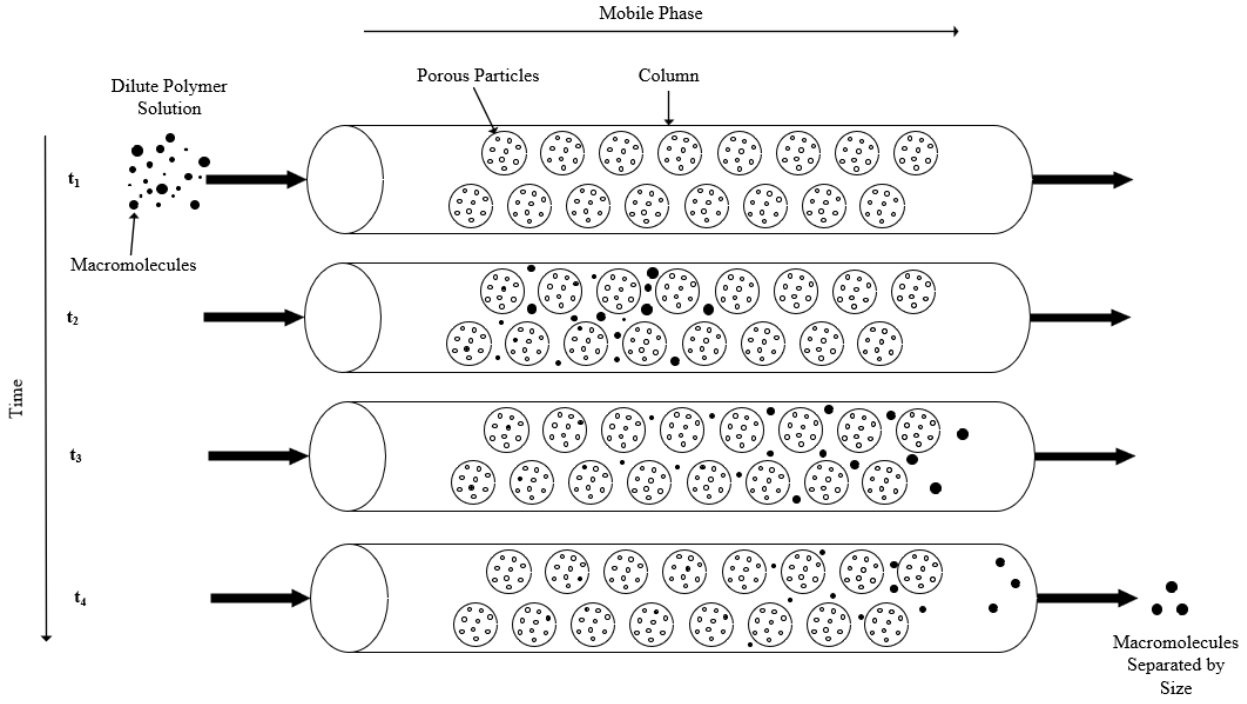


Figure 12. Schematic of SEC separation of a polymer sample. A dilute polymer sample is injected into the column filled with stationary porous particles and a mobile phase. Smaller molecules will pass through the porous particles and extend their pathlength. Larger molecules which cannot penetrate the smaller pores within the packing will find their way through the column faster. This figure is intended provide a qualitative visualization of SEC and is not to scale by any measure.

SEC is a valuable tool for determining molecular weight averages and distributions of polymers by separating molecules based on hydrodynamic size. Using SEC for the determination of molecular weights of complex branched polymers such as lignin adds additional complexity and requires special considerations to achieve precise and accurate results. The increased number of functional and end groups resulting from its variable structure and degree of branching pose challenges to its solubility and increase the potential for undesirable interactions with the stationary phase. Additionally, the structural heterogeneity, non-uniform degrees of branching, and even

conformational variations between macromolecules increases the uncertainty in the translation of hydrodynamic volume to molecular weight. Two lignin macromolecules of truly equal molecular weight could potentially assume large variation in hydrodynamic volume given that one is linear, and the other is highly branched and/or crosslinked. This concept is demonstrated in Figure 13.

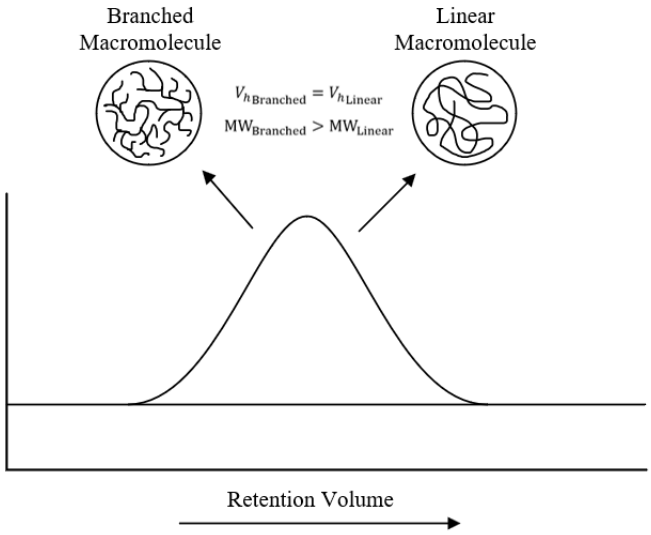


Figure 13. Visualization of branched and linear macromolecules with similar hydrodynamic volumes eluting in the same retention volume. The molecular weight of the branched macromolecule is higher than that of the linear macromolecule, however, they exhibit the same hydrodynamic volume. Here V_h is hydrodynamic volume and MW is molecular weight.

At the current state of the field, researchers seeking to determining the molecular weights of lignin using SEC should first identify suitable mobile and stationary phases to reduce axial dispersion resulting from undesirable interactions from lignin. Then, based on their desired information and application, they can review the detectors and or combination of detectors available that is most suited to reducing uncertainty in the results. It would be valuable to abide by the available standards that seek to reduce interlaboratory variation in the technique so that the results would be comparable and relative to literature. While reducing uncertainty is always desirable, sometimes achieving consistent and comparable results could be more advantageous, depending on the application. All experimental techniques have an associated uncertainty, and although this section focused on the complications that challenge the application of this technique of SEC, this technique is used routinely for relative determination of lignin molecular weight. SEC

dominates the industry for molecular weight determination and has done its part in advancing our understanding of lignin and countless other polymers.

4.7 Summary

Much like lignin extraction, there is no single perfect technique for lignin analysis. These approaches have improved significantly with the developments of new instruments, techniques and procedures. While there are many methods available when attempting to analyze lignin structure it is also important to consider the research objective, nature of sample, samples availability. When selecting the experimental method, consideration of method accuracy might be just as important as the sample availability/number or time requirement. If the number of samples is large, faster methods such as vibrational spectroscopy or Py-GC/MS that rapidly characterize large amounts of data would be useful. If more specific information is required, one can consider partnering methods to offer a more comprehensive image of the lignin sample. Additionally, a more comprehensive method, such as high-resolution NMR, can be employed with smaller sample quantities (40-70 mg) [118]. If properly planned, samples sizes of just 100-200 mg of wood (20-30 mg of lignin) and a few days of experimental work is all that is needed for the isolation and NMR analysis of lignin [118]. It is also worth noting that differences in data from multiple analysis methods and extraction techniques could be used to make comments on the native lignin structure [118].

Chapter 5: Chemometrics and lignocellulosic biomass – Advanced strategies for the accelerated analysis and valorization of lignin

5.1 Introduction

Fossil fuels are a major contributor to anthropogenic climate change, which has become a primary global interest. Growing concerns about the effects of climate change are incentivizing scientists to find new means for energy and materials production that could reduce emissions. There are many challenges to finding and implementing an alternative to fossil fuels that can renewably and sustainably provide a source of both energy and materials. Lignocellulosic biomass provides a unique and promising opportunity to fill this void as it holds the potential as a renewable source of both fuels and carbon-based materials and chemicals [25] [161] [162].

Lignocellulosic biomass is derived from organic plant matter and primarily composed of three main biopolymers: cellulose, hemicellulose, and lignin [163]. Cellulose is a semicrystalline linear polymer solely composed of D-glucopyranose (glucose) units linked via β-1,4-glycosidic bonds [24]. Hemicellulose is an amorphous branched polymer composed of pentoses (D-xylose, D-arabinose), hexoses (D-mannose, D-glucose, D-galactose), and occasionally uronic acids and acetyl moieties as side-chain groups [24] [25]. Lignin is a complex amorphous polymer composed predominantly of three 4-hydroxyphenylpropanoids distinct in their number of methoxy groups on the 3 and 5 positions on their phenyl ring [164]. These are known as *p*-coumaryl (4-hydroxycinnamyl), coniferyl (3-methoxy-4-hydroxycinnamyl), and sinapyl (3,5-dimethoxy-4-hydroxycinnamyl) alcohol, or as *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units, respectively, in the context of the lignin polymer [164] [165]. The organization, structure, and interactions of cellulose, hemicellulose, and lignin in the plant cell wall pose challenges to the efficient conversion of lignocellulosic biomass to refined products [166] [167]. As a result, scientists are focusing on understanding the biochemistry of the plant cell wall and its components especially lignin composition and chemical structure [168]. Lignin's native structure is fundamentally unknown because of the inherent challenges in studying it in situ [169]. However, it is thought that a more developed understanding of lignin's molecular structure and interactions

with other plant cell wall components will be fundamental for efficiently processing and refining the resource [170].

A modern-day biorefinery must efficiently utilize lignocellulosic biomass from all main plant cell wall components, including lignin, for economic viability [171]. It may be useful to note that there are slightly inconsistent and indistinct definitions of the term “biorefinery”. The European concept is somewhat broad and includes both biofuel processes and pulp and paper processes as biorefineries. In contrast, the North American concept does not always recognize current pulp and paper processes as biorefineries [162]. We consider a biorefinery to mean any integrated industrial process that utilizes biomass feedstock(s) to produce bio-based fuels, chemicals, or materials [9]. The carbohydrate polymers are often considered valuable biomass components, and biorefineries concepts have primarily focused on carbohydrate-based biofuels. Historically, lignin has been treated as a waste stream rather than a marketable product, largely due to economic considerations and the inherent challenges in its processing. Yet, lignin has the potential as a feedstock for a wide variety of materials, including thermoplastic and thermoset composites, carbon nanofibres, lignin reinforced rubbers, foaming materials, and aerogels (3D solid nanoporous polymeric network materials) [172]. Lignin based micro- and nanoparticles are one of the most promising new materials for the application in biomedicine [173], including their use as a drug delivery agents [174] [175] [176]. Lignin presents many opportunities for use in polymeric materials, resulting from a variety of properties relating to its environmentally friendly nature, universal availability, antioxidant, antimicrobial and biodegradable properties. Recently, a review of forty-two peer-reviewed life cycle assessments concerning lignin and its derived products found that lignin-based products offer improved environmental performance compared to their fossil fuel-derived counterparts, most notably concerning their impact on climate change [177]. While this is important, to compete with conventional petroleum-based feedstocks and synthetic polymeric materials, understanding lignin’s structure and organization within the plant cell wall is required to synthesize value-added lignin-based materials. For lignin products to be of high value, they have to exhibit properties for specific applications; therefore, lignin valorization pathways will be vital for the success of these strategies. Recently, it has been demonstrated that technical lignin, containing considerable quantities of residual crystalline cellulose, exhibits enhanced performance in comparison to high-purity lignin for some applications, therefore leaving

opportunities for less complex and less costly lignin valorization strategies [162]. This is a more recent vision of the second-generation biorefineries, which may reduce costs and environmental impact by employing processing strategies with lower complexity and higher-value products. Regardless of the specific valorization pathway, it is estimated that if lignin is fully utilized for high-value applications at the industrial scale, the revenue from the lignin stream would be greater than that of the other components, which would transform our understanding of biomass utilization [162].

Isolating the main biomass components is critical for the efficiency of current and future industrial biomass operations [178]. There are a range of chemical/biochemical/thermal/physical processing approaches for isolating lignin from the plant cell wall. Often the recovered lignin is termed after its respective isolation treatment. For example, lignin isolated using the kraft process is termed kraft lignin, while lignin isolated using organic solvents is term organosolv lignin. Challenges with the effective isolation of these components from their native state within plant biomass are commonly associated with two main variables: lignin content and structural composition [179]. Recalcitrance is often associated with a higher lignin content and a lower ratio of syringyl to guaiacyl lignin monomer units (S/G ratio), which is a metric for communicating the relative proportions of its two most abundant monomer units. Linkage types, functional groups, branch density, and interaction with other plant cell wall components are also factors that could play a role in biomass recalcitrance. The plant cell wall is amorphous and heterogeneous, and its composition is dependent on numerous factors, such as plant type, age, location, and environment [180]. Furthermore, studying biomass components is challenging because of their resistance to degradation and the complex supramolecular organization of its three main components [181]. The nature of lignin and its interactions make research efforts to understand the plant cell wall, as well as industry efforts to efficiently isolate and refine its main components, face similar challenges. Figure 14 shows the three main components of lignocellulosic biomass and the known lignin-lignin linkages. Lignin's variable and complex structure makes it difficult to characterize its polymeric architecture and properties compared to other more uniform materials. In fact, lignin's structure, both in situ and in isolated states, is somewhat controversial, with long-time experts in the field holding different opinions on the extent of branching of the lignin polymer [27] [28].

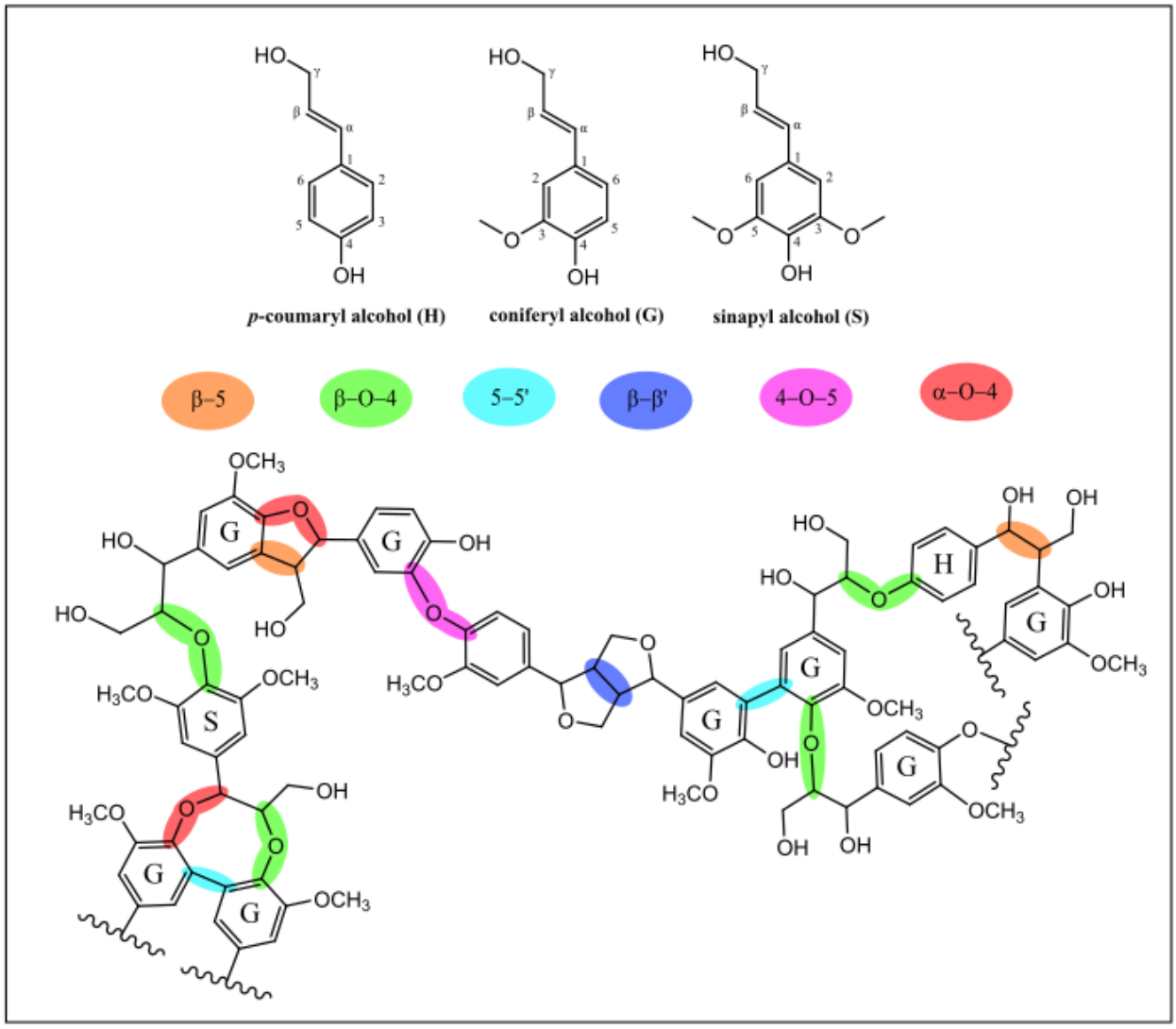


Figure 14. The three main components of lignocellulosic biomass, *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol, as well as their known lignin-lignin linkages. The linkages are highlighted in the model lignin polymer where the carbon – carbon linkages $\beta - 5$, $5 - 5'$, and $\beta - \beta'$ correspond to orange, light blue, and dark blue, respectively, and the carbon – oxygen linkages $\beta - O - 4$, $4 - O - 5$, and $\alpha - O - 4$ correspond to green, purple, and red, respectively.

Rapid and accurate techniques to study and analyze lignocellulosic samples and feedstocks are desired at both the research and development and industrial scales. Lupoi et al. have provided an in-depth review of the innovations in analytical methods for the qualitative and quantitative assessment of lignin [101]. They discuss the different analytical tools based on their use in the

analysis of lignin structure, lignin-carbohydrate and lignin-lignin linkages, lignin molecular weight, total lignin content, and lignin monomer composition. A wide variety of analytical techniques assist in the characterization of biomass, including spectroscopic methods, wet chemical methods, chromatographic methods, and microscopy. Wet chemical analysis approaches have been relied upon heavily to determine lignocellulosic biomass features, such as the relative contents of cellulose, hemicellulose, and lignin; however, wet chemical analysis can be extremely laborious, costly, and time-consuming. Spectroscopic analytical techniques are an alternative approach, providing powerful tools to generate specific molecular-scale information at significantly reduced time and cost. They offer both rapid qualitative and quantitative structural information on lignin and biomass with routine techniques [182]. While spectroscopic approaches can provide detailed information, the complexity of lignocellulosic material can convolute the results, making their interpretation complicated. Moreover, while spectroscopic techniques can rapidly analyze samples, they produce large amounts of complex data, especially when characterizing complicated natural organic materials such as lignocellulosic biomass. No single technique can provide all the desired information, and often a combination of techniques is employed for thorough characterization and structural validation. While researchers have focused a lot of their efforts on understanding the details of lignin's structure, more application-based investigations may instead focus on lignin's properties. In a recent review, Balakshin et al. suggest that although conventional lignin structural analysis is actually relatively effective and reliable, the real bottleneck for lignin engineering is related to the evaluation of lignin's performance for specific applications [162]. Correlation between lignin structure and properties may be essential for accelerated screening of lignin performance, which may help speed up the current process at the multiple scales of commercial application.

Chemometrics is a research discipline concerned with methods that can extract useful information from raw chemical data sets and was initially defined by Kowalski and Wold as the application of mathematical and statistical tools to chemistry [183]. Historical perspectives of chemometrics from some of the pioneers of the field can be found elsewhere [184] [185]. Chemometrics originated in analytical chemistry and are involved in the analysis and interpretation of experimental data. In particular, many chemometric techniques have developed around spectroscopy because of its rapid analysis that generates large amounts of data [186]. The

discipline of chemometrics is broad and mature with various tools and approaches that are all essentially means to support the statistical design of experiments and multivariate data analysis. Most real-world systems are multivariate and stochastic; thus, scientists turn to statistical approaches for interpretation. While chemometrics focuses on multivariate data in chemistry, many other scientific disciplines face similar challenges with large multivariate datasets (e.g., econometrics, psychometrics). These related fields may share the same tools used in chemometrics, and many chemometric techniques are not necessarily limited to chemistry applications. Regardless of their field of application, these approaches rapidly evaluate large data sets with intricate internal relationships that would otherwise be difficult to interpret.

Recognizable under multiple definitions and syntax – chemoinformatics (or cheminformatics/chemiinformatics) – is a broad term for a broad discipline that encompasses numerous different areas of chemistry [187]. The broadest definition for chemoinformatics may be that of Gasteiger and Engel: Chemoinformatics is the use of informatics (computer science) methods to solve chemical problems [188]. Since most of the mathematical and statistical tools applied in analytical chemistry today utilize computers, it is clear that there is a considerable overlap between the fields of chemoinformatics and chemometrics. In fact, it is common to see chemometrics classified as a subfield of chemoinformatics [189] [190] [191] and perhaps equally as common to see chemoinformatics classified as a subfield of chemometrics [192] [193]. Cheminformatics generally focuses on the efficient extraction of knowledge from data, storing and evaluating this information in chemical databases, and making predictions [194]. We see the value of chemoinformatics as aiding in knowledge handling and synthesis for researchers and providing advanced pathways for molecular modeling applications, providing key insights into the complex material. While chemometrics provides an analytical chemist with a variety of tools for interpreting and dealing with large complex data sets, it also facilitates a connection between experimental methods and computational chemistry. Data organization into matrices, preprocessing steps that can assist with the interpretation of data, and models that accelerate sample characterization are some advantages for the transfer of information to the computational space. Chemoinformatic approaches for storing, transforming, and analyzing this data in computers are the next steps for enabling the smooth transfer of this information for computational investigations.

Research on lignocellulosic biomass has increased in recent years as the modern lignocellulosic biorefinery edges closer to real commercial implementation. Lignin valorization efforts will likely be central to the success of these ventures. There is already a great deal of attention on lignin engineering, but application-based research efforts for the cost-effective transformation of the raw material into valuable products with desired properties will become critical [162]. The technical challenges for lignocellulosic biorefineries are multiscale and interdependent. Understanding the variable nature of lignocellulosic feedstocks, the effects of different unit process operations and conditions, as well as the varying chemical structures and properties is a challenging task. These problems need to be understood in a multivariate context, and chemometrics offer tools for this purpose. They can identify subtle patterns in large complex data sets or reveal interdependence among variables. Advanced regression modeling strategies can rapidly predict the physical and chemical properties of lignocellulosic materials with high-throughput analytical techniques. These techniques also have potential to directly correlate the structural features of lignin with performance characteristics. This review highlights a wide range of problems that can be tackled using chemometric approaches. A conceptual illustration of using chemometrics strategies for analyzing lignocellulosic feedstocks is presented in Figure 15. Chemometrics have often been applied for solving problems in research and development, but are also promising for implementation into industrial settings. With the advancement of chemometrics concurrently with computational efficiency, chemometric techniques can be integrated into computer process control schemes for online monitoring of a lignocellulosic biorefinery.

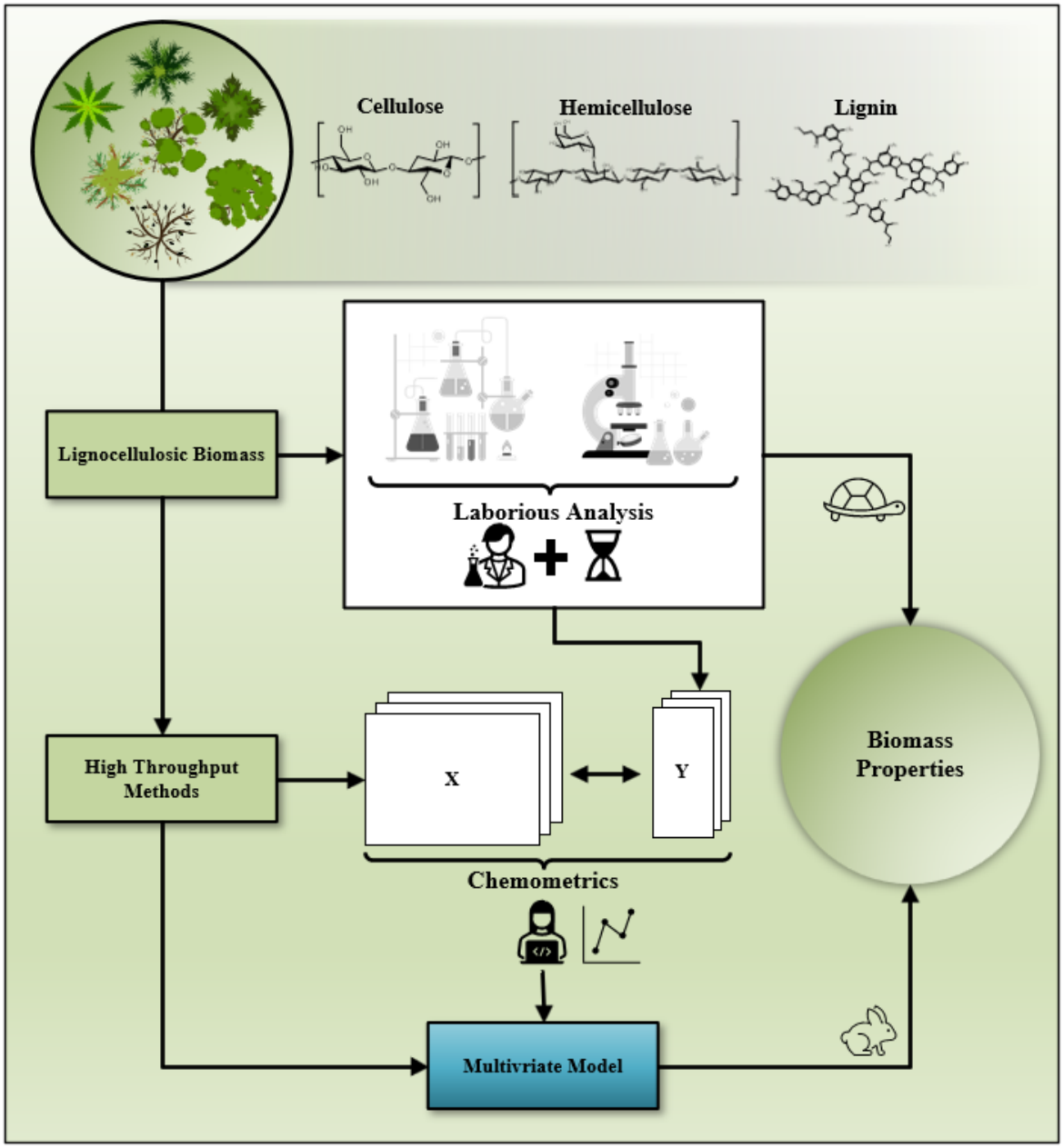


Figure 15. Conceptual illustration of using chemometrics strategies for analyzing lignocellulosic feedstocks. Biomass properties determined using laborious chemical analysis can be used to develop multivariate models. These multivariate models can then be used for future cases to rapidly predict biomass properties using higher throughput analysis methods, in this way, future cases can bypass the laborious chemical analysis. Here, the **X** matrix may represent any measured or observed variables, which are used as the independent variables, while the **Y** matrix represents the dependent variables.

Categorizing the applications based on analytical technique and including tables (found in the Appendix B) that summarize the papers and their multivariate techniques used will provide inspiration and identify gaps in innovation. Despite a large number of resources applying chemometrics available today, the ability to perform chemometric techniques does not guarantee that the outcome or interpretation of the results will be valuable [195]. Therefore, we provide a general introduction to the main techniques to encourage more understanding of these techniques prior to application. It may be tempting to carelessly navigate in chemometric software packages when analyzing data; however, this can lead to misunderstandings or misuses of these techniques. A review of common misunderstandings in chemometrics may guide where and why these mistakes can occur [196].

This review introduces the most popular chemometric techniques for interpreting data in the literature and discuses their applications to lignocellulosic biomass research, focusing on lignin valorization. We provide a comprehensive and contemporary overview of this area, demonstrating the field's growth, achievements, and trends. This paper will suit a general audience, ranging from scientists/engineers experienced in these techniques but looking for areas of opportunity, to a novice looking for a straightforward but detailed introduction.

5.2 Experimental Design

Experimental design is an important area of chemometrics that supports efficient acquisition of data sets containing desired information at lower cost [197]. Intuitively, the quality of the information being attained from an experiment depends heavily on the quality of the data used for analysis. Approaches for experimental design screen out irrelevant factors, optimize experiments using systematic methods, save time, and promote optimal quantitative modeling [198]. Experimental design can be turned into a mathematical or statistical problem where the optimal number of runs and the type of experiments can be determined efficiently [199]. Multivariate calibration, which will be covered in more detail in subsequent sections, can encounter major problems if the calibration strategy is poorly designed. Undesired correlation in data sets can invalidate a calibration model, but experiments can be designed so that components are uncorrelated or orthogonal to each other. Standard techniques for this application include multilevel fractional factorials, Plackett-Burman and Taguchi designs [198]. While the

experimental design is an important element of chemometrics, it is often omitted in the literature related to lignocellulosic biomass. Not all chemometric data sets are available in a formal statistically designed form, yet, this does not invalidate all of these data sets. Information can still be attained, but one should be conscious of the implications on the conclusiveness of their findings. There are many books and reviews [198] [200] [201] that cover experimental design techniques in more detail, as well as examples of experimental design in the literature on biomass analysis [202] if one seeks to explore this area further.

5.3 Basic Raw Data Preprocessing

Data preprocessing plays a pivotal role in chemometric data analysis and therefore must be covered in this review [203]. Analytical techniques, such as spectroscopic/chromatographic analysis, are influenced by uncontrollable factors that can affect the resulting data set. A proper data preprocessing approach can typically improve the performance of chemometric techniques by reducing the impact of these factors [204]. There are a variety of preprocessing techniques used to "clean up" the data by removing undesired variation, and to simplify subsequent multivariate data analysis methods; however, the method of selecting the optimal preprocessing technique for a specific application is subjective which can have major implications on the subsequent data analysis [205] [206]. It is common for data preprocessing techniques to be chosen using a trial-and-error approach, if there is no widely accepted technique for a specific situation [207]. However, this does not mean that these operations should be performed without understanding the fundamental effect on the data structure [208]. More recently, a complementary fusion of preprocessing techniques and their combinations has been an emerging approach [209].

Chemical data is almost always multivariate, therefore it is commonly represented using matrices. The rows and columns of these matrices correspond to the objects (e.g., samples) and features (e.g., variables measured on samples), respectively [189] [210]. For this reason, it can be useful to categorize data preprocessing steps as either "column-wise" or "row-wise" methods, where row-wise methods are applied sample by sample and column-wise methods require all samples to perform the preprocessing (Figure 16) [200]. Sometimes preprocessing is applied to individual elements rather than row-wise or column-wise. These preprocessing steps can be called the transformation of individual elements and are commonly used to achieve better data symmetry,

which is sometimes useful for skewed data [189]. Common approaches include logarithmic, power, and box-box transformations; however, these types of data preprocessing techniques are not so commonly used in chemometrics. This is because multivariate data sets often have distinct distributions among variables [211]. There are also dimensionality reduction/variable selection methods that could be considered preprocessing approaches but do not necessarily fit within the main categories. Such approaches include Orthogonal Signal Correction (OSC), Orthogonal Projections to Latent Structures (O-PLS), and Genetic Algortithms (GA). These are broader approaches acting on the data set to remove the unimportant data and select for valuable data. Brief descriptions of the main chemometric data preprocessing techniques can be found in Appendix B.

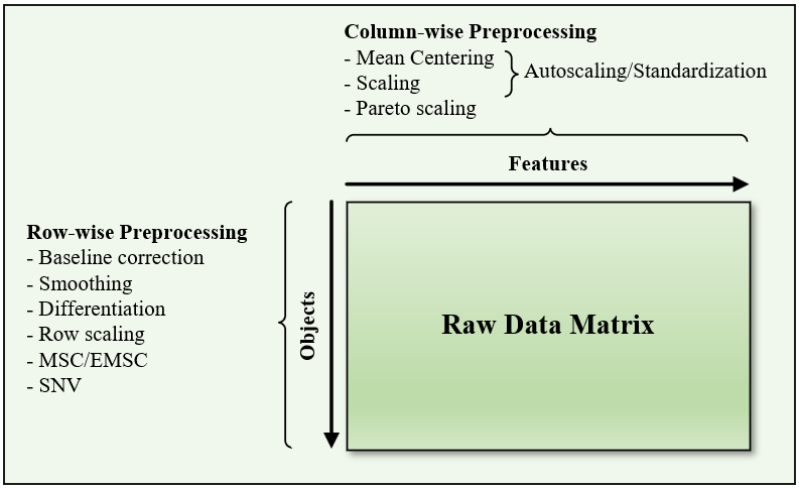


Figure 16. Overview of the two main classes of data preprocessing methods applied to a raw data matrix.

5.4 Multivariate Data Analysis – Pattern Recognition and Calibration

A major element of chemometrics is data analysis using multivariate statistical approaches. In contrast to univariate statistics, which relate a single independent variable x to a single dependant variable y, chemical data is often multivariate in nature and thus requires analysis that can consider many variables [189]. This section briefly describes the two leading families of multivariate techniques: pattern recognition and calibration. The term pattern recognition is self-explanatory; however, there are three main groups of methods for pattern recognition in chemometrics: exploratory data analysis, unsupervised pattern recognition, and supervised pattern recognition. All these approaches share the goal of recognizing or identifying patterns in data;

however, each approach takes a different methodology or has different objectives. On one hand, exploratory data analysis methods focus on the key characteristics to be analyzed and identified without any bias toward a particular outcome; on the other hand, unsupervised pattern recognition (e.g., cluster analysis) is used to discover patterns and identify similarities between data sets [198]. That being said, exploratory data analysis and unsupervised pattern recognition are sometimes used synonymously in the literature as a distinct category in contrast to supervised pattern recognition techniques. Supervised pattern recognition differs from unsupervised pattern recognition as it is mainly intended for classification, and therefore require training or calibration to allow for the classification of future samples [198]. In this way, supervised pattern recognition is related to calibration, and one could perhaps classify calibration as a subsection of supervised pattern recognition. Calibration, which will be introduced in more detail in the ensuing sections, involves the development of a mathematical model between measured properties and a property of interest.

5.4.1 Exploratory Data Analysis

5.4.1.1 Principal Components Analysis (PCA)

Perhaps the most fundamental multivariate exploratory data analysis method in chemometrics is PCA [198]. The PCA method essentially transforms a data set containing several interrelated variables to a new set of uncorrelated variables called principal components (PCs), where the first few PCs capture most of the variation from the original variables [212] [213] [214]. These PCs are determined through linear combinations of the original variables so the variance of these components is maximized [215]. In this way, PCA reduces the dimensionality of a data set, allowing for the interpretation of all variables simultaneously [216]. As PCA is commonly used to create graphical representations of the data that provide easy-to-understand visual insight in two or three dimensions, it can be considered an exploratory first step toward multivariate analysis [217]. Score and loading plots may be used to visualize the PCA results, often the scores relate to the objects or samples and the loadings relate to the measured variables [198]. The PCA method can be used for a variety of objectives such as outlier detection, dimensionality reduction, and graphical clustering with different varieties and extensions of the technique for these purposes [216]. Wold et al. describe the primary scope of PCA for multivariate data analysis in chemistry as getting an overview of the dominant “patterns” in the data tables, which represent the

relationships among objects and/or variables [218]. Note that different algorithms are available for performing PCA, including Jacobi rotation, nonlinear iterative partial least-squares (NIPALS), and singular value decomposition (SVD), the most commonly used algorithm in software packages [189].

5.4.1.2 Other Bilinear Decomposition Methods

Expressing an original data matrix as a product of two matrices is termed bilinear decomposition [203]. The PCA method also takes this form and thus falls under bilinear decomposition along with the related methods such as independent component analysis (ICA) and multivariate curve resolution (MCR) [219]. Instead of finding components that maximize the explained variance as done in PCA, ICA finds statistically independent components [220]. Both methods are often used in an exploratory context; however, ICA is much less popular than PCA. The MCR [221] is a family of methods that can resolve a data matrix of a complex mixture into pure physical/chemical profiles of its components [222] [223]. These methods are based on the concept that a measured data set of a mixture of components can be accurately modeled as an additive bilinear model of the individual components using only the information from the original measurement of the mixture [224]. While not all data sets follow a bilinear model, MCR can still be used as an exploratory tool for identifying patterns or variations in a data set [225]. The MCR methods are distinct from PCA and ICA in that their objective is to recover the solutions reflecting the true components' contribution to the data, as opposed to abstract solutions that represent sources of variation within the data. In this way, MCR helps provide valuable insights into the underlying nature of chemical systems.

5.4.2 Unsupervised Pattern Recognition (Clustering)

Clustering allows samples to be organized or grouped based on their measured properties without prior grouping information [198] [189]. There are a variety of cluster methods and algorithms available such as partitioning, hierarchical, fuzzy, model-based, PCA, factor analysis, Kohonen mapping, Sammon’s mapping, and Chernoff faces [189]. Particularly, hierarchical cluster analysis (HCA) and *k*-means clustering were frequently applied in the literature regarding lignocellulosic biomass. The *k*-means clustering is the most widely known partitioning algorithm of cluster analysis, and it works by assigning each object to one of *k* different clusters [189]. In

contrast, hierarchical clustering methods generate a set of cluster solutions by hierarchically ordering $k = 1, \ldots, n$ clusters, where n is the number of objects [189]. A dendrogram is used to visually present the results of hierarchical clustering. A more in-depth description of cluster analysis in the context of chemometrics can be found here [189].

5.4.3 Supervised Pattern Recognition (Classification)

Supervised pattern recognition requires training to classify an object or sample. Classification methods include linear discriminant analysis (LDA) [226], *k*-nearest neighbor (*k*-NN) algorithms [227], soft independent modeling of class analogy (SIMCA) [228], artificial neural networks (ANN) [229], and support vector machines (SVM) [230], among others.

5.4.4 Multivariate Calibration

In chemometrics, calibration can be defined as mathematically relating, correlating, or modeling a measured response based on the amounts, concentrations, or other physical or chemical properties of a set of analytes [231] [232] [233]. Martens and Naes simply define multivariate calibration as the means for determining how to use measured variables (e.g., $x_1, x_2, \cdots x_n$) simultaneously for quantifying the target variable(s) (e.g., y) [234]. The wide application and popularity of calibration can be attributed to its tangible advantages, such as reducing sample preparation requirements, removing systematic errors due to interferences in measured data, and rapidly providing desired information for a system [235]. A schematic of multivariate calibration is shown in Figure 17, where a matrix **X** corresponding to a data set of observed variables on a set of objects is mathematically related to a **Y** matrix corresponding to another set of known properties. Typically, once a model from calibration/regression is developed, it requires proper validation to ensure that it has good performance and predictability power. If the model is not accurate enough, it may require more training with more samples, a different type of model, or it may be found that there is not a relationship between the data obtained for the **X** and **Y** matrices, and therefore modeling is not a suitable approach. Model validation and optimization will be discussed in more detail in section 5.

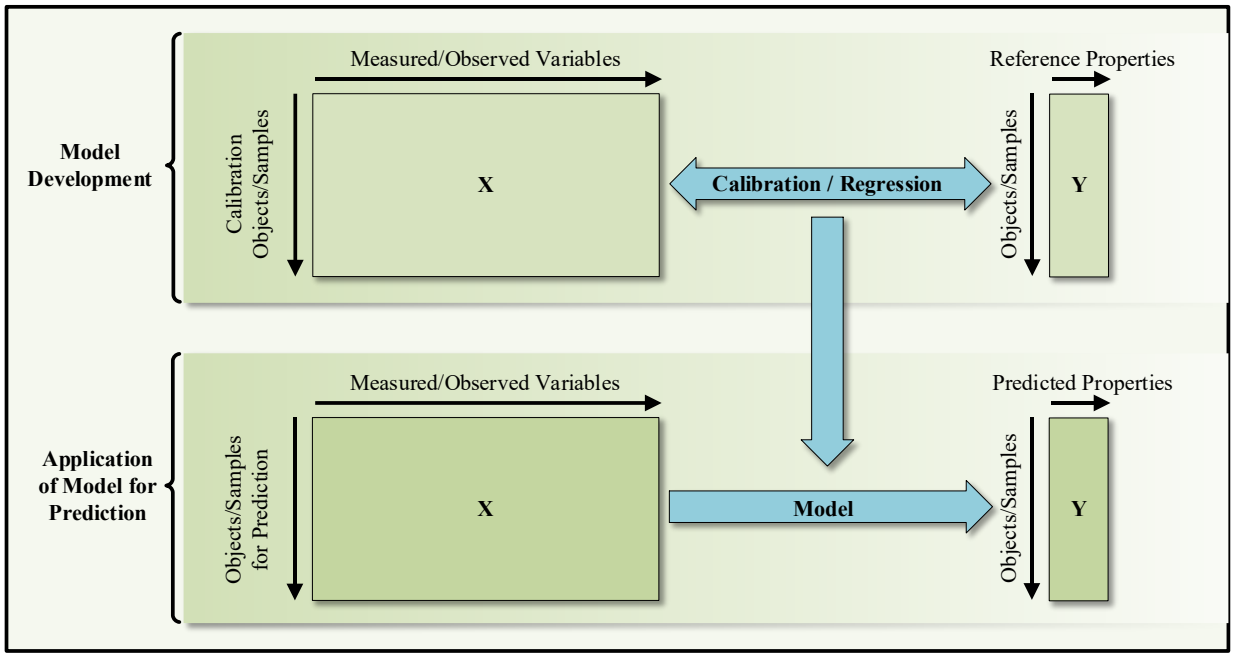


Figure 17. Visual representation of the calibration process. A predictor matrix (**X**) and a response matrix (**Y**) are used to develop a calibration model. The model can then be applied to a new data set to predict the properties of a new matrix based on the relationships found among the original data sets. Note here that preprocessing and validation may be key elements of this process but are not illustrated in this figure.

5.4.4.1 Multiple Linear Regression (MLR)

Linear regression relates two variables or vectors in the form of a straight-line equation (e.g., $y = mx + b$) by adjusting the values of m and b to provide the best prediction of the response variable y given x. The MLR method is an extension of the simple linear regression; however; it aims to establish a relationship between a single response variable (y-variable) and multiple predictor variables (x-variables) by fitting a linear equation to the data. Multivariate regression should not be confused with multiple regression since it refers to situations with more than one dependant variable [236]. MLR is the simplest approach for multivariate modelling and is a fundamental regression technique that is widely used [237].

5.4.4.2 Principal Components Regression (PCR)

Principal components were briefly introduced in Section 5.4.1.1 for PCA; however, PCs have also been used extensively in regression analysis [214]. The PCR method is a combination

of both PCA and MLR developed by replacing original regressor variables with their principal components, simplifying the computation, and adding stability with new orthogonal variables that are statistically independent [238] [239] [240]. While multivariate regression models are typically designed to minimize the sum of squares, PCR relates the concept of principle components with regression models to solve for multicollinearity, a situation where two variables/predictors are highly linearly related [215].

5.4.4.3 Partial Least Squares Regression (PLS/PLSR)

Originally developed by Herman Wold [241] in 1975 for modeling complicated data sets in terms of chains of matrices but introduced to the field of chemometrics by his son Svante Wold, among others, partial least squares regression (PLS/PLSR) is an approach for modeling relationships between observed variables [189] [242]. The process of PLS is similar to that of PCR in that the original data is transformed to new variables that are subsequently used for regression; however, as opposed to PCR, which derives the new variables exclusively from the predictor variables, PLS uses maximum covariance criteria for this transformation based on information from both the predictive and target variables [189]. This approach is useful because maximizing covariance merges the high variance in the predictor variables and the high correlation with the target property(s) [189]. The PLS is a simple approach to relate two data matrices with a linear multivariate model, and it is also reportedly [189] the most widely used method for multivariate calibration in chemometrics [242]. There are multiple algorithms for PLS and some important terminology to note. For example, PLS1 and PLS2 are distinct in that they refer to the situation where the response variable is a vector and matrix, respectively [189]. Therefore, PLS1 has only one response variable, whereas PLS2 has multiple.

5.4.4.4 Artificial Neural Networks (ANN)

Artificial neural networks are an exciting artificial intelligence tool for developing non-linear relationships between input and output variables based on the concept of emulating the function of the human brain [243]. ANN's are networks of artificial neurons that are nonlinear, parameterized, bounded functions [244]. A neurons output is calculated from a function called an activation function that takes inputs (x_i) and parameters that are assigned to the inputs as weights (w_i) [229]. Typically, an ANN model is organized in a three-layer configuration: the input layer, hidden layer(s), and an output layer. The ANN can be highly sophisticated approaches of modeling

complex systems and therefore may be useful for complex data produced from lignocellulosic biomass. The application of ANN include mapping, regression, modeling, clustering, and classification, and they are especially useful in non-linear problems that are challenging for more traditional statistical methods [229]. Additionally, while chemometric techniques are generally useful for large data sets, there can be a point where the traditional techniques, such as MLR, PLS or PCR, become less effective with too much information; in contrast, more information is generally favourable for ANN methods [245]. Some limitations of ANN include potential issues with overfitting if limited samples are available in practical applications, as well as time-consuming computation, the requirement to optimize many parameters, and difficulties with the interpretation of ANN models [246].

5.5 Model Validation, Optimization, Performance Criteria, and Testing

While chemometric models have great potential to provide insights and make predictions, they are inductive in nature and only useful to the degree that they can accurately represent the phenomena they are employed to describe. So how then could one derive confidence that the model they have developed based on their current observations will provide them with reliable predictions of future observations? This is a very important question, and the answer is that we can establish a level of confidence in a model by optimizing, testing, and evaluating it using performance criteria. Model validation approaches use specific criteria to measure the performance of a model. For regression models, this is generally based on observations that have not been used in the development of the model to determine residuals (i.e., the actual/observed values less the predicted value). Model validation can be classified as either external or internal, with test sets and cross-validation, respectively, being the most common techniques.

A major component of model validation is data splitting. If sufficient data were available, the best practice would be to split the data set into distinct subsets referred to as 'training/calibration', 'validation/optimization', and 'test/prediction' sets [247]. A distinction should be made between the terms model optimization and validation [248]. Optimization allows the determination of the optimum model, while validation determines how well a model fits the data; these are separate although related procedures. With that said, optimal models are selected

through the evaluation of how well the models fit the data and this is why validation and optimization are often combined into one subset. A general principle of modeling is parsimony, meaning that the best model is that which has the best prediction ability with the least complexity [249]. This can be accomplished using optimization. Figure 18 shows a conceptual illustration for taking a data set, splitting it into subsets, and then a flowchart strategy for using these subsets to build a reliable regression model.

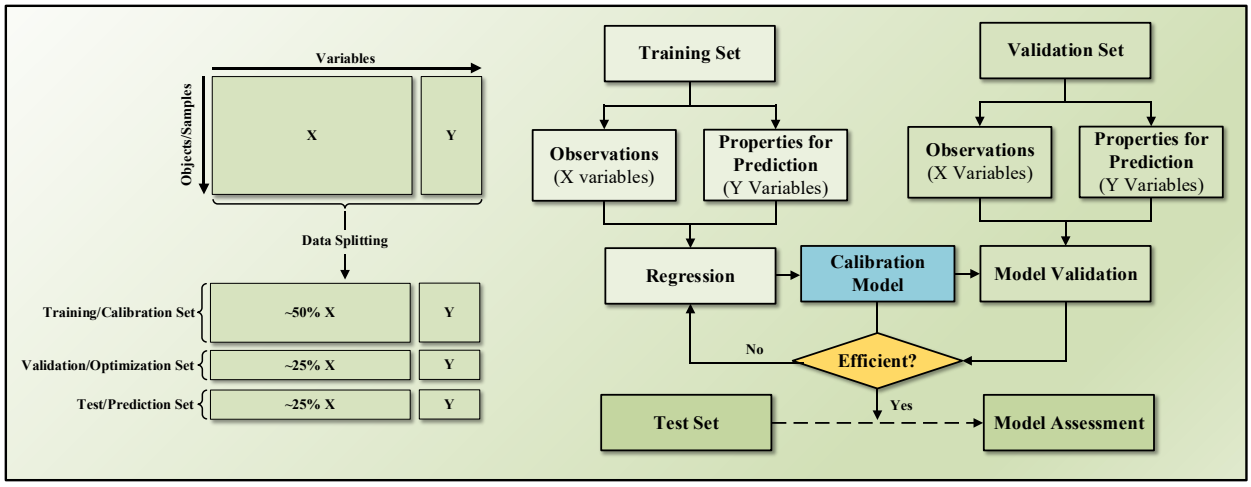


Figure 18. Illustration of data splitting and a flowchart strategy for building reliable regression models. Adapted from Xu et al. (Figure 3) [250].

5.5.1 Model Optimization and Validation

In test set validation, which is considered the most reliable form of model assessment, model performance is evaluated using an external data set, independent of the data used to develop the model. If a proper test set were always available, there would be no need for any other forms of validation; however, in real-world applications, there are often limited observations available and sufficient data to form both a representative model and a test set that can adequately assess the model are not obtainable. Cross-validation (CV) is another model validation approach based on resampling. It is essentially an iterative procedure where part of the data is left out and subsequently predicted using a model developed from the remaining data. In this process, all objects are used at some point to build the model and to test the model, yet never simultaneously. There are many variations of CV, as well a variety of objectives for the technique, such as determining the suitable number of components for a PCA model or regression model optimization

[251]. The CV is somewhat controversial in the chemometric literature, particularly as it applies as an alternative to test set validation. Esbensen and Geladi described CV approaches as sub-optimal simulations of test set validations and heavily criticized the technique [252]. Despite this criticism, CV is very popular and defended as being useful for various purposes, such as when the number of objects is very limited, when the purpose of the model is to understand the inherent data structure of the system and not primarily for prediction or classification, or when the objects in the dataset could be grouped based on contextual information about the samples [253]. A rule of thumb that has been proposed is to apply CV if the number of samples is smaller than 40 [253].

5.5.2 Performance Criteria

Performance criteria are typically based on residuals (e_i) obtained by subtracting the predicted value from its true value. Performance criteria are derived from the residuals but use different mathematical strategies to define distinct criteria. These criteria can be used for different objectives, for example, they could be used to select the best model by adjusting its parameters to reduce the magnitude of the residuals, as well as to assess a model by determining its expected error for new cases. There are a variety of abbreviations used for communicating performance criteria which can be confusing; however, they are relatively straightforward once they are clearly outlined. For a more detailed overview of these performance criteria, Varmuza and Filzmoser provided a clear introduction, here we will provide a brief summary [189]. We also provide the equations for performance criteria in the Appendix B. The standard deviation of prediction errors, known as the standard error of prediction (SEP), provides an estimate of the spread of the error distribution. If the predicted values are from the calibration set, it is sometimes referred to as the standard error of calibration. While the acronym SEC is often used for the standard error of calibration, here we do not use it to avoid confusion with size exclusion chromatography (SEC). Alternatively, one can indicate which data set the SEP is calculated from using an index, for example, SEP_{Test} or SEP_{Cal}, which would refer to the standard error of prediction for a test set or a calibration set, respectively. Another performance criteria commonly used is the mean squared error (MSE) which is the arithmetic mean of the squared errors. Taking the square of the MSE provides root mean squared error (RMSE) which is almost the same as the SEP if bias can be neglected. The RMSE criteria are preferred for practical applications because their units match the original data, however, the non-squared MSE are useful for model optimization purposes [200].

Similar to MSE is the predicted residual error sum of squares (PRESS). Like the SEP, either indexing or altering the acronym are common approaches to specify the data set used to calculate the performance criteria. Sometimes R^2, the coefficient of determination is used to represent the spread of the predictions.

5.6 Overview of Chemometric Applications to Lignocellulosic Biomass

To date, there have been no reviews fully dedicated to the application of chemometrics to lignocellulosic biomass and only five papers that partially cover the topic [254] [255] [256] [257] [258]. Thus, this work will be the most recent and comprehensive overview of this field. Krasznai et al. provide a good historical context of biomass characterization and discuss new opportunities for chemometrics and spectroscopy for lignocellulose characterization [254]. They discuss many of the key data preprocessing techniques and present recent applications of chemometrics in lignocellulosic biomass research. Moreover, they outline several key points of improvement for future work, including the mindful selection and clear reporting of data preprocessing, modeling, and validation parameters, surrogate mixtures which emulate properties of materials to improve the model building, and the development of databases that could facilitate both more comprehensive models and or more collaboration in this field. The authors conclude that these chemometric techniques are expected to become routine for analysing lignocellulosic material in both research and industrial settings.

Alzagameem et al. have reviewed the availability and composition of different plant biomass as potential feedstocks for second-generation biorefineries and dedicated the second portion of their review to the potential of multivariate data processing for biomass analysis and quality control [255]. They discuss chemometrics for spectroscopic data processing, lignocellulosic feedstock specification, and aspects for future applications in quality control. As pointed out, the number of studies using chemometric methods in biomass analysis compared to the total number on biomass analysis is very small, therefore potentially suggesting that chemometric analysis is still in its early stages. They comment that mostly classical chemometric techniques for modeling primarily FT-IR data have been employed in the field, leaving gaps for new advanced algorithms and more chemometric applications to different analytical techniques.

Additionally, they note that there are no examples of multivariate techniques for resolving overlapping peaks in 1D and 2D NMR profiling or multivariate modeling of specific ^{31}P and ^{13}C NMR profiles. Based on our review, most of these points still stand; however, it is worth noting that in the more recent papers, we have seen the application of more advanced chemometric techniques, especially for near-infrared (NIR) and Raman datasets.

Rump et al. have dedicated a short section of a book chapter on lignin and its characterization techniques to chemometrics and computational chemistry used in lignin analysis [258]. Based on their literature survey, IR techniques are the primary technique used with multivariate data analysis, an observation they attributed to the simplicity, low cost, and convoluted nature of NIR spectra. In parallel with Alzagameem et al., they note that methods with more complex data sets, such as the 2D NMR methods heteronuclear single quantum coherence (HSQC), heteronuclear multiple bond correlation (HMBC), diffusion-ordered spectroscopy (DOSY), have not been extensively explored using chemometric approaches, this indeed is an area with a high potential for innovation [255]. Data-based approaches for analyzing complex data sets, such as those derived from the methods listed above have the potential to greatly expedite the analysis time and lower the barrier of entry for new users of these techniques.

Lupoi et al. reviewed spectroscopic techniques for biomass characterization with a focus on high-throughput spectroscopic techniques [256]. They provide an informative overview of recent spectroscopic approaches for studying lignocellulosic biomass and discuss many examples where chemometrics are utilized. They concluded that these high-throughput approaches can reliably predict biomass characteristics and would be essential for second-generation biorefinery feedstock screening. Xiao et al. have reviewed NIR and pyrolysis-molecular beam mass spectrometry (Py-MBMS) methods coupled with multivariate data analysis for high-throughput biomass characterization [257]. They present a guide for the effective data analysis of NIR and Py-MBMS of biomass by summarizing the important structures measured by using these techniques and introducing, comparing, and evaluating the conventional and multivariate data analysis methods. They concluded that the high-throughput analytical techniques NIR and Py-MBMS in combination with chemometrics are efficient tools for exploring the features of biomass data sets, characterizing the cell wall chemistry, and predicting chemical, physical, mechanical, and fuel properties.

This section will present an overview of chemometric techniques applied within lignocellulosic biomass research. Although this paper is focused on reviewing the chemometric applications rather than the analytical techniques, this section was organized based on analytical techniques to showcase the environment and trends of application of chemometrics in the field of biomass analysis. Each subsection will have a brief introduction to the analytical technique and then a discussion of some of the key findings and applications. Detailed summary tables listing all the chemometric papers identified that fall under the respective subsection category can be found in Appendix B. Using this approach, we aim to identify gaps in the field that can be addressed in future studies.

5.6.1 Vibrational Spectroscopy

Vibrational spectroscopy is the study of the interaction between electromagnetic radiation and matter. Vibrational spectral data can contain detailed qualitative information on the structure, properties, and environment of molecules, even in complex systems [259]. Yet, complex multicomponent systems, such as lignocellulosic materials, limit the effectiveness of this approach; consequently, chemometric approaches have been employed to deal with such multivariate datasets both quantitatively and qualitatively. Several vibrational spectroscopic techniques can be used to analyze complex systems, including lignocellulosic biomass and its components. The most common being the near-infrared (NIR), mid-infrared (MIR), and Raman spectroscopies. These techniques are known to provide fast, non-destructive, qualitative, and quantitative analyses of lignocellulosic biomass and lignin [260]. Vibrational spectroscopy is one of the fundamental analytical chemistry techniques, which chemometrics has formed around, and thus there is a solid background of chemometric approaches for vibrational spectroscopic data. This is reflected in a large number of chemometric applications to NIR, MIR, and Raman spectroscopy of lignocellulosic materials identified in this review. Morais et al. have provided a detailed tutorial on multivariate classification for the vibrational spectroscopy of biological samples [261].

Infrared spectroscopy has been used in lignin chemistry since the early 1950s [141]. The IR spectrum can be divided into the near (12500 cm^{-1} ($\lambda = 800\ nm$) – 4000 cm^{-1} (2.5 μm)), mid (4000 cm^{-1} (2.5 μm) – 400 cm^{-1} (25 μm)), and far (400 cm^{-1} (25 μm) – 10 cm^{-1} (1 mm)) IR

spectral regions [262]. In IR spectroscopy, a sample absorbs electromagnetic radiation due to molecular vibrations and the absorption of IR energy produces IR absorption bands, which can be observed [263] [262]. The IR spectrum, which represents the number of IR absorption bands, their intensities, and their shapes, provides information on the molecular structure of a sample [262]. There are several protocols for IR spectroscopy in biochemistry [264] [265], including that of Baker et al., who have brought together several leaders in the field of IR spectroscopy to develop a consensus on the spectral preprocessing and data analysis of both traditional vibrational spectra and spectral images [266]. Xu et al. have provided a mini-review of qualitative and quantitative analysis of lignocellulosic biomass using IR techniques [250].

In contrast to IR spectroscopy where vibrational spectral information is collected based on the absorption of IR photons, in Raman spectroscopy vibrational spectra are produced from light scattering via the Raman effect [267]. Like the IR spectrum, the Raman spectrum includes the number of bands, their intensities, and their shapes, which are also directly related to the molecular structure of the sample [262]. The IR and Raman spectroscopies each provide complementary information of molecular vibrations based on the excitation of anti-symmetric vibrations, causing changes in dipole moment (vector quantity), and symmetric vibrations that change the polarization (tensor quantity), respectively [268] [262]. Therefore, IR spectroscopy is often used for the detection of functional groups of molecules, while Raman spectroscopy is often used for the detection of skeletal structures [268]. While IR and Raman spectra provide information on chemical bonds in the form of spectra, compositional information is not always clearly resolved in multicomponent systems. Chemometric multivariate analysis approaches are valuable for both building relationships between spectral features and their corresponding chemical components/bonds, as well as for relating the information on chemical components/bonds to biomass properties [250].

It is also important to draw attention to spatially resolved vibrational spectroscopy techniques which are known as vibrational microspectroscopy [269]. These techniques combine the principles of vibrational spectroscopy and microscopy to reveal chemical changes over microscopic dimensions. In principle, they can be used to probe the chemical microstructure of plant cell walls in their native state. These techniques create large 3D datasets called hyperspectral data cubes that require data analysis approaches for interpretation. The data preprocessing and

analysis of microspectroscopy data has been reviewed [270, 271] and Gierlinger has reviewed the vibrational microspectroscopy of plant cell walls [272]. They also discuss multivariate data analysis for vibrational microspectroscopy and focus on the promising methods of cluster analysis, vertex component analysis (VCA), and non-negative matrix factorization (NMF). Vibrational microspectroscopy has been demonstrated to provide insights into the plant cell wall structure and composition, facilitated by multivariate data analysis [273].

5.6.1.1 Near-Infrared Spectroscopy (NIR)

The NIR region is mainly composed of overtones and combination bands (stretching and deformation) of fundamental molecular vibrations which are more subtle, with lower intensity than the fundamental vibrational modes seen in MIR or Raman, and therefore require deconvolution, which typically is performed using chemometrics [141] [274] [275] [256]. Despite this, the NIR region has multiple advantages over the other IR regions which include the rapid analysis times, the ability to generate spectra for a variety of materials without complicated sample preparation, and good reproducibility [274]. Based on these advantages, specifically its speed and simplicity, NIR is very suitable for process control applications using online monitoring [274]. The broad and overlapping bands that convolute the data pose a major challenge for NIR interpretation, making it difficult to accurately assign the structural features of a sample based on its spectral features [276]. While the nature of the MIR spectrum allows for easy qualitative identification of organic compounds, it is also susceptible to instrumentation and sample preparation, therefore, NIR is more suited for quantitative analysis [277]. Thus, with the introduction of multivariate data analysis methods that could reduce overlapping signals or deconvolute spectral data, NIR has emerged as a useful quantitative tool. NIR applications to wood science and technology have been reviewed [278], and a tutorial for assessing trees, wood, and derived products with NIR has been made available [279]. A list of the papers identified in this review that use NIR and chemometrics to study lignocellulosic biomass or lignin along with the preprocessing methods and multivariate techniques employed can be found in Appendix B.

An early study by Wallbacks et al. compared Carbon-13 Nuclear Magnetic Resonance Spectroscopy Cross-Polarization Magic Angle Spinning (^{13}C NMR CP/MAS), MIR, and NIR combined with multivariate data analysis for the prediction of changes in pulping chemical composition during birch kraft pulping [280]. They used PLS to model the relationship between

the three spectroscopic data sets, and the wet chemical analysis of Klason lignin, glucose, xylose, arabinose, mannose, and galactose contents. They found that all PLS models were able to predict the changes in the main constituents Klason lignin, glucose, and xylose; however, the remaining sugars that were present in too small of amounts (<1% each) were not described properly. In their opinion, the model was insufficient to describe low-content sugars because of limited sensitivity of both the spectroscopic and chemical analysis. Of the three techniques, they found that NIR gave the best model in terms of predictability. In addition, they suggested that combining the data from the different spectroscopic techniques may provide a more complete picture of the changes occurring in the kraft pulping process than using a single technique. To test this hypothesis, they combined principal components from each spectroscopic data matrix into a single **X** matrix, scaled it, and then performed PLS using the chemical constituents as the **Y** matrix. This approach did not improve the standard error of prediction for any constituents compared to the NIR model. They then tested another approach using 19 original NIR variables and three principal components, each from both NMR and MIR. The PLS method was applied to the dataset, and a standard error of prediction for all three main constituents was reduced. The authors concluded that any of the three spectroscopic methods (^{13}C NMR CP/MAS, MIR, and NIR) could be used to describe the compositional changes (Klason lignin, glucose, xylose content) occurring during the kraft pulping of birch.

Poke et al. assessed the use of NIR in combination with PLS as an alternative to the costly and time-consuming wet chemical methods to predict lignin and extractives contents in wood as a measure of pulping quality [281]. A good agreement between model prediction and laboratory determined contents was observed. They concluded that NIR analysis was a reliable predictor of extractive and lignin content in *Eucalyptus globulus*. Stackpole et al. studied the natural genetic variation in monomer composition (S/G ratio), lignin, cellulose, and extractives contents using a 16-years-old field trial of an Australian tree species, *Eucalyptus globolus* [282]. They employed the model developed by Poke et al. to predict Klason lignin and extractives contents and developed additional models to predict cellulose and pulp yield, and S/G ratio [281]. Significant genetic correlations among wood chemical traits at population and additive genetic levels were observed. Particularly, the population differentiation in the S/G ratio of lignin was positively correlated with latitude, which may result from chemical adaptation to climate or associated biotic factors.

Alves et al. combined chemometrics techniques, such as PCA and PLS, with NIR spectroscopy and various reference methods to investigate lignocellulose materials and modeling their properties. They developed PLS models for H/G ratios from the analytical pyrolysis of pinewood [283], Kappa number of *Pinus pinaster* [284], S/G ratio of *Eucalyptus globulus* [285], wood extractives of *Eucalyptus globulus* [286], and wood density based on X-ray micro density data of *Pinus pinaster* and *Larix* x *eurolepis* [287]. Different preprocessing strategies were also investigated to develop refined models. Overall, the results confirmed that NIR and chemometrics could provide accurate predictions, demonstrating the versatility of the approach for a variety of important properties for biomass processing.

Schwanninger et al. used wet-chemical methods to determine the total lignin content of 200 Norway spruce wood samples and analysed the same samples with NIR [288] [289]. The PLS regression was then used to develop a mathematical correlation between the wet-chemical data and NIR spectra. The authors aimed to create a simple PLS model with lower error of prediction than those already published. A variety of preprocessing steps were tested, which included MSC and first and second derivative with the Savitzky-Golay algorithm. Additionally, different wavenumber ranges were explored. The authors were able to develop several good models with a low error of prediction. Yet, since there were many models with similar performance characteristics, the authors introduced an "evaluation step" to determine the model with the highest predictability over the most spectral variations. The evaluation step involved the use NIR spectra of 366 additional wood samples and was necessary to evaluate the pre-selected combinations of wavenumber ranges and preprocessing methods.

Horikawa et al. developed a model between NIR spectra and wet chemical analysis determined properties of pretreated rice straw [290]. They treated rice samples with alkaline, acid, and hydrothermal treatments. The resulting samples were then subjected to wet chemical analysis, enzymatic hydrolysis, and NIR spectroscopy to determine biomass composition, saccharification efficiency, and obtain NIR spectra, respectively. The PLS regression was used to develop a quantitative model between the Savitzky–Golay second derivative NIR spectra and the saccharification ratio and compositional properties of the pretreated samples. A schematic illustration of the calibration procedure can be found in Figure 19. The results showed acceptable predictions of saccharification ratio and some major structural components. In a later study,

Horikawa et al. applied a similar methodology using PCA and PLS to predict the saccharification efficiency and wet chemical properties of chemically pretreated *Erianthus* [291]. They were able to interpret the regression coefficients of their PLS model to correlate compositional variables with saccharification efficiency. These correlations were corroborated by the wet chemical data. Once again, they were able to develop an acceptable quantitative calibration model.

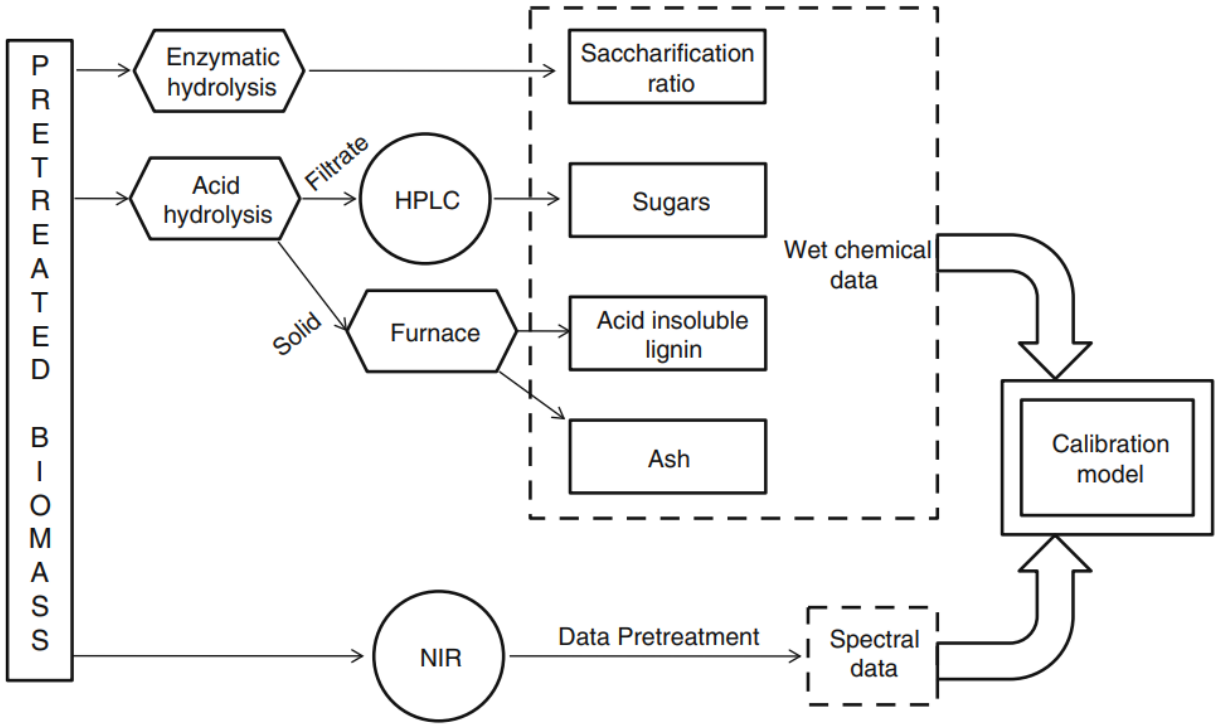


Figure 19. Schematic illustration of the procedure to construct the calibration model between the wet chemical and spectral data. HPLC is high-performance liquid chromatography. Reproduced from Figure 1 of Horikawa et al. [290].

Via et al. investigated NIR spectroscopy coupled with PLS and PCR in the applied to wood chemistry [292]. They focused specifically on the precision in factor loadings determination for PLS and PCR. The loadings from these models were used to interpret the relationship between the wood chemistry and/or functional groups and some key traits [293] [294] [295]. The authors objective was to determine if PLS introduced additional error in the loading plot as compared to PCR, because of the shifts in the loading peaks occurring during the process of optimization of the covariance between the **X** and **Y** data matrices. They developed PCR and PLS models on both the

raw spectral data and first derivative with the Savitzky-Golay smoothing to predict cellulose, hemicellulose, lignin, and extractives. It was concluded that the PLS model was superior for prediction while PCR was better for model interpretation and wavenumber selection. This study highlights the need for clear objectives when selecting a chemometric method for analysis because prediction and interpretation may differ among the techniques.

Krongtaew et al. monitored the changes in lignin, hemicelluloses, and amorphous, semi-crystalline, and crystalline regions of cellulose moieties of straw subjected to acid and alkaline pretreatments partly in combination with hydrogen peroxide [296]. Their goal was to reduce lignin content in order to increase the accessibility of enzymes to polysaccharides for digestion. The NIR spectra were processed with PCA to analyse the changes occurring during the pretreatment. The contributions of the second, third, and fourth PCs were used to generate a score plot (Figure 20). The first PC did not explain a significant distinction among all groups and was excluded from interpretation. The authors concluded that key properties of biomass influencing enzymatic hydrolysis were identifiable using NIR and could be discriminated qualitatively and explained using PCA.

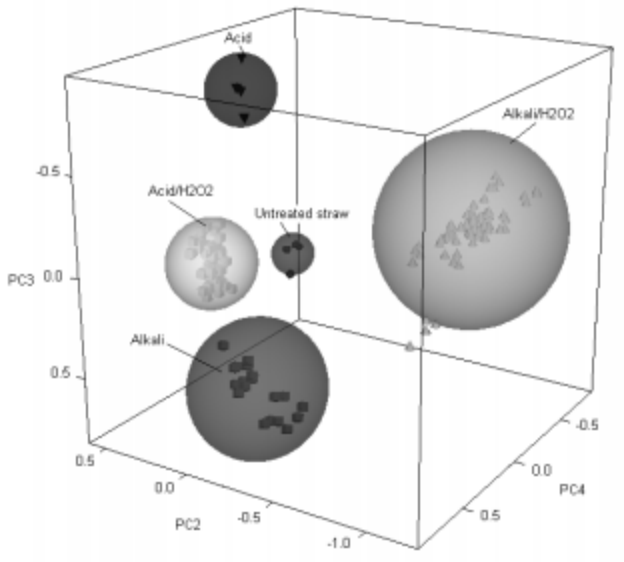


Figure 20. Clustering of untreated wheat straw samples (●; 4 samples) and samples treated with acid (▼; 4 samples), alkali (♦; 20 samples), acid/H_2O_2 (■; 68 samples), and alkali/H_2O_2 (▲; 68 samples). The spheres indicate the supreme spectral distances from the average spectra of straw samples from each cluster represented as the coordinate origins. Reproduced from Figure 6 of Krongtaew et al. [296].

Modern lignocellulosic biorefineries require feedstocks with predictable and desirable properties to compete in global markets with petroleum-based products. Drought is one of the most common plant stress agents and can severely alter the productivity and composition of biomass feedstocks. Understanding the impacts of drought on these feedstocks will be a high-priority variable for future biorefineries, particularly as climate change is predicted to accelerate. Moreover, while understanding the role lignin plays in the properties and value of lignocellulosic feedstocks is already a technical challenge in the field, lignin has also been associated with drought stress tolerance in some lignocellulosic feedstocks [48] [50] [59] [61]. Thus, drought tolerance adaptations that promote altered lignin content/composition must be considered for both upstream selective breeding and downstream feedstock engineering strategies. Van der Weijde et al. have utilized chemometrics to probe the influence of drought on Miscanthus and its use as a biomass feedstock for biofuel production [45]. They subject 50 diverse *Miscanthus* genotypes to drought conditions and monitor their biomass properties. Wet chemical analysis and saccharification reactions are used to determine biomass properties, such as cell wall, cellulose, hemicellulose,

lignin content, and cellulose conversion, respectively. The NIR spectra of the samples are then acquired and subject to multivariate data analysis. Prior to the development of these regression models, the data was preprocessed using MSC, first and second derivatives, and smoothing treatments. Modified PLS regression models are then generated from a randomly selected calibration set (110 samples out of the 400 available) to predict cell wall, cellulose, hemicellulose, lignin content, and cellulose conversion from the NIR spectra. An external validation set of 20 samples is used to evaluate the prediction quality of the models. The study has shown that drought stress has a considerable effect on biomass quality for biofuel production, in particular significant reductions in the cell wall and cellulose content and a significant increase in hemicellulose. However, lignin content is not significantly affected, which according to the authors contributes to the inconsistency among various studies regarding lignin content in different species and tissues under drought stress. In addition to the observed changes in biomass content, drought has been found to significantly increase cellulose conversion, suggesting drought may enhance the efficiency of biorefineries. This is found to hold true even for the tolerant genotypes that maintained good yields despite drought conditions.

More recently, Liang et al. have coupled NIR with PLS to predict the holocellulose and lignin contents of multiple wood species, including poplars, eucalyptus, and acacias as a rapid non-destructive analysis technique for the pulp and paper industry [297]. A comparison of variable selection methods for NIR spectral variables optimization is performed with competitive adaptive reweighted sampling, Monte-Carlo uninformative variable elimination (MC-UVE), successive projections algorithm, and GA. Data preprocessing steps include MSC, first and second derivatives. It has been found that MSC and second derivative preprocessing efficiently resolve undesirable scatter effect and overlapping peaks in the raw NIR spectra. The authors have also observed that variable selection methods based on regression coefficients rather than collinearity minimization or heuristic population search are more reliable and efficient. Particularly, the competitive adaptive reweighted sampling method is significantly more accurate and robust compared to the other models.

Yang et al. investigated a portable NIR spectrometer combined with chemometrics as a rapid tool to determine holocellulose and lignin contents in pulp wood [298]. They evaluated four preprocessing methods for removing noise and irrelevant information in the raw spectra, which

included first derivative, moving average filtering, MSC, and SNV. They also compared four modeling approaches, including PLS regression, least-squares SVM, back-propagation neural network, and kernel extreme learning machine. The PCA method was used for dimensionality reduction prior to calibration of the last three modeling approaches. Additionally, a particle swarm optimization algorithm was used to optimize the regularization parameter and kernel function parameter of the least square SVM and the kernel extreme learning machine. The authors found that the EMSC successfully reduced the noise and other irrelevant information from the data and that the kernel extreme learning machine was the best predictive model compared to other approaches. They concluded that their method had a great potential for the rapid and accurate assessment of pulp wood using a portable NIR spectrometer. Wolfrum et al. compared the performance of small, low-cost portable NIR spectrometer prototypes (a Texas Instruments NIRSCAN Nano evaluation model and an InnoSpectra NIR-M-R2 spectrometer) to a conventional laboratory spectrometer for the rapid assessment of biomass composition [299]. They analysed 270 well-characterized herbaceous biomass samples and developed calibration models using the PLS-2 algorithm to predict glucan, xylan, lignin, extractives, and ash content from the NIR spectra. Their results showed that although the portable spectrometers had smaller wavelength ranges (900-1700 nm) than the conventional laboratory NIR spectrometer (400-2500 nm), there was no statistically significant difference in the model within the comparable wavelength ranges between the portable spectrometers and the conventional one. They concluded that these results were encouraging, but challenges remain for these low-cost portable spectrometers. Some of these challenges include the need for a software "ecosystem" for data collection and model application, calibration transfer (transfer of models between instruments), and robust methods of sample preparation. The authors also noted that more advanced machine learning algorithms could be implemented to potentially improve the generated models.

Two recent studies combine NIR and chemometrics to determine lignin content in Korla fragrant pears [300] and 'Snow' pears. Wu et al. investigate different variable selection techniques combined with NIR for determining lignin content in 'Snow' pears, a popular fruit in China, because of lignin's unfavorable impact on the flavor and quality of the fruit [301]. As discussed above, a large number of spectral data acquired from NIR could contain collinearity issues and reduce the performance of a PLS regression model. Therefore, the authors propose a variable

selection method to identify influential variables before the PLS regression model is established, in order to improve the model performance. Their method is a bootstrapping soft shrinkage combined with the frequency and regression coefficient of variables (FRCBOSS) approach. In their work, the authors evaluate multiple data preprocessing steps, which include Savitzky-Golay smoothing, normalization, SNV, MSC, and first derivative both independently and in combination. They also compare five variable selection methods, including synergy interval partial least squares, competitive adaptive reweighted sampling, successive projections algorithm, bootstrapping soft shrinkage, and their own proposed FRCBOSS approach, both independently and in combination. The PLS regression models are employed predict lignin content in 'Snow' pears and are compared. The authors find that the synergy interval partial least squares-FRCBOSS method performed the best for prediction. It is concluded that their method is useful and accurate for the rapid determination of pear lignin contents.

Elle et al. investigated the use of NIR and chemometrics to predict lignin content in fine root as lignin is a key player in root decomposition, which is a basic element of the terrestrial carbon cycle and relevant for understanding global ecology and climate [302]. They used PLS to establish a calibration between NIR spectra and acetyl bromide determined lignin content in the fine root of 60 grassland species. The authors also combined PLS with the spectral variable selection method competitive adaptive reweighted sampling to identify and select the most relevant wavelengths for root lignin prediction. A schematic outlining their workflow for chemical analysis is shown in Figure 21. The PLS method combined with this variable selection approach resulted in a significant increase in model validation accuracies for all measures of model accuracy. They noted that these findings on PLS models with competitive adaptive reweighted sampling for variable selection were in agreement with other relevant studies. The authors concluded that NIR combined with chemometrics had a great potential for predicting lignin content in fine root and could be particularly useful for ecological studies with large sample sizes but limited sample amounts.

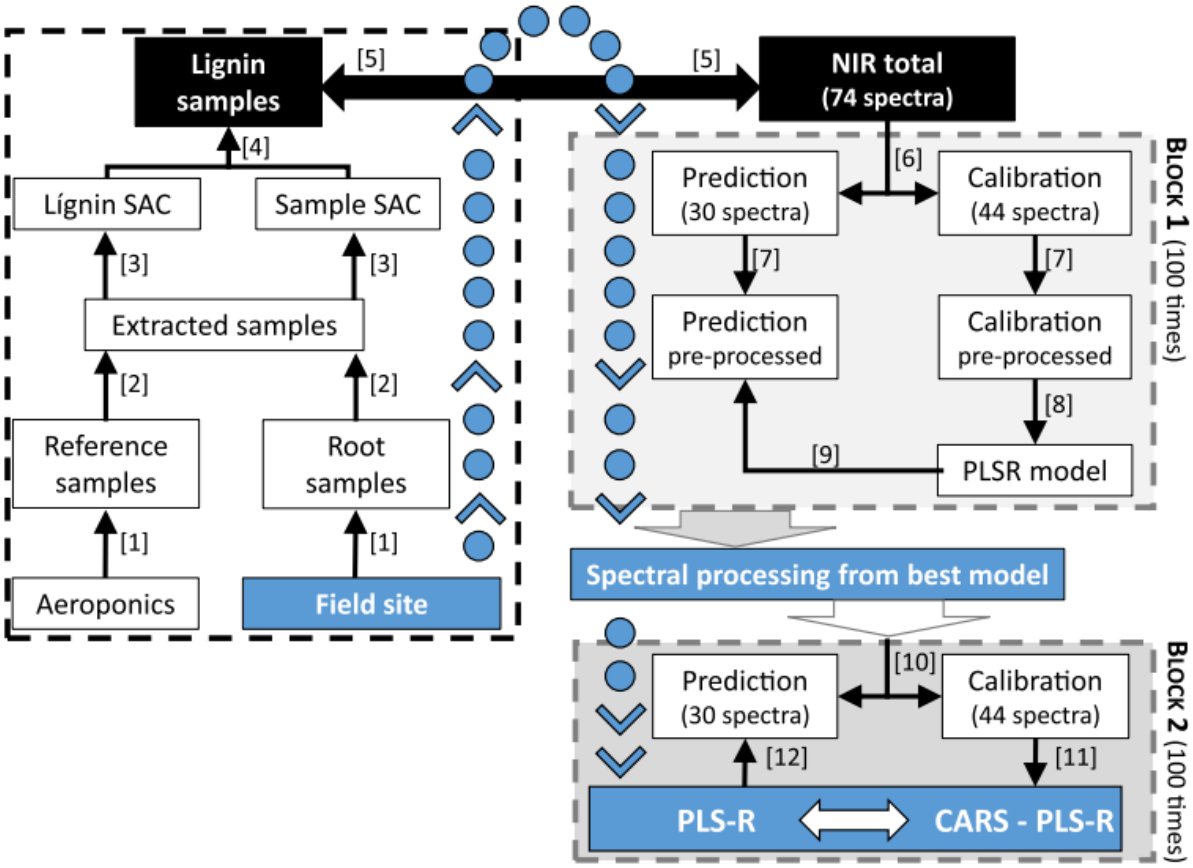


Figure 21. Schematic depiction of the workflow of chemical analysis from Elle et al. [302] (left box with steps 1–4 indicated in square brackets), spectral analysis (outlined in black boxes and step 5) and statistical analysis. The latter is subdivided in spectral pre-processing (block 1 with steps 6–9) and final analysis (block 2 with steps 10–12). Reproduced from Figure 1 of Elle et al. [302].

5.6.1.2 Mid-infrared Spectroscopy (MIR/FT-IR)

The MIR region consists of fundamental molecular vibrations and related structures [250]. MIR is comprised of sharper absorption bands that are highly distinctive to a specific compound. This high degree of spectral resolution leads to a more straightforward correlation between a sample IR spectrum and its inherent structural features [303]. For quantitative analysis, however, MIR is not as useful as NIR because of its high degree of sensitivity to instrumentation and sample preparation [277]. There are three most widespread MIR techniques which include attenuated total reflectance (ATR), photoacoustic, and diffuse reflectance FT-IR spectroscopy. A recent

comparison of these three methods for different bio-organic samples have been discussed [304]. A list of the papers identified in this review that use MIR and chemometrics to study lignocellulosic biomass or lignin along with the preprocessing methods and multivariate techniques employed can be found in Appendix B.

One of the early applications of chemometrics to FT-IR analysis of lignocellulose was done by Schultz et al. [305]. They studied the feasibility of the FT-IR procedure combined with stepwise variable elimination regression to accurately determine the cellulose, hemicellulose, and lignin content of lignocellulose samples as a rapid alternative to the conventional procedure which was lengthy and included many steps with considerable error. This methodology for reducing analysis time using regression modeling and chemical data sets has since grown in popularity and complexity.

Chen et al. employed FT-IR spectroscopy in combination with multivariate data analysis techniques to expedite the chemical composition analysis of wood [306]. They used the first derivative of the FT-IR spectra of the wood samples as the **X** matrix and the chemical determinations of the contents of the various components (i.e., lignin, cellulose, hemicellulose) as the **Y** matrix. Preprocessing steps prior to PLS regression also included mean-centering the **X** matrix and autoscaling the **Y** matrix. The PCA and HCA methods were used to qualitatively identify differences between samples, and the PLS model was used to simultaneously predict the three main components of wood (cellulose, hemicellulose, and lignin). It was concluded that PCA and HCA were useful tools for the discrimination of wood types (e.g., hardwoods *versus* softwoods) based on FT-IR spectra. Additionally, the PLS regression model was used to determine wood composition successfully. The authors note that this multivariate calibration model for FT-IR required 15 minutes per sample, which is efficient compared to the 3-4 days for the Van Soest [105] [307] analysis method.

Boeriu et al. evaluated FT-IR as an analytical technique to estimate the chemical composition and functional properties of lignin [308]. They used PCA on the IR-fingerprint spectra to classify lignins according to their origin and processing conditions by finding information on the dissimilarity between lignin samples and identifying which variables contributed the most to this difference. They also used a PLS model to predict the major components' concentrations and

radical scavenging activity. Their work showed that FT-IR spectra could provide a useful qualitative and quantitative characterization of the chemical structure and properties of lignins. In addition, they demonstrated that PCA is effective for the classification of lignin materials and PLS can be applied to accurately model lignin concentration and antioxidant and emulsion-stabilizing properties of lignin-based materials. The authors concluded that FT-IR combined with chemometric methods can fast and reliably characterize lignin, which could be especially useful as a non-destructive technique for the quality control of lignin-derived materials.

As lignin's structure varies, it has been difficult to predict its polymer properties like molecular weight and polydispersity, yet these characteristics have been fundamental for engineering lignins for targeted applications. Lancefield et al. proposed a chemometric approach to accelerate the characterization of lignin [309]. The authors demonstrated the use of ATR-FT-IR analysis combined with PCA and PLS modeling for the quantitative prediction of structural features of technical lignins that would normally require the use of SEC and 2D HSQC nuclear magnetic resonance (NMR). Calibration was performed by analyzing 54 different technical lignin samples covering kraft, soda, and organosolv processes with traditional SEC and NMR and correlating with their ATR-FT-IR spectra. The PLS models were used to correlate the ATR-FT-IR spectra with the SEC and NMR measurements. The authors concluded that their method could predict valuable structural information on lignin samples accurately.

Building on this finding, Khalili et al. have sought to demonstrate if the ATR-IR approach could be applied to the on-line monitoring of a lignin depolymerization reaction, which could be a useful strategy in a prospective biorefinery [310]. They use the off-line SEC measurements of weight average molecular weight (M_w) along with the operando ATR-IR spectra of the depolymerization of kraft lignin reaction to build a chemometric model for predicting future M_w and polydispersity values. Despite the many analytical challenges of acquiring the ATR-IR spectra of an operando depolymerization reaction (e.g., interferences from solvent, temperature, baseline shift, and light scattering from particles), they have been able to correct the spectra and develop a reasonable predictive model using an approach developed previously by the same team. In this previous work, Khalili et al. have demonstrated an operando ATR-IR spectroscopic method to monitor the aqueous phase reforming of Kraft lignin at 225°C and 30 bar over a Pt/Al_2O_3 catalyst

[311]. They use PCA to overcome water vapour and CO_2 gases interferences in the spectra, and PLS to correct for the differences in the heating rate of solvent and the reactions. Thus, using the aforementioned approach, Khalili et al. have been able to acquire higher-quality spectra to build their predictive model for molecular weight [310]. In addition to the spectral correction from the previous work, MSC is applied to account for any additive or multiplicative scaling effects. Subsequently, PCA is used for the exploratory analysis of ATR-IR spectra. The score plot from PCA distinguishes not only trends in M_w and polydispersity, but also information on saturated structures (Figure 22). Next, three quantitative multivariate regression models are built to predict the M_w of the continuous system using PLS and evaluated to select the optimum model for future predictions. The authors note that further model optimization could be possible with other chemometric techniques, such as variable selection and/or different preprocessing strategies. This would generate a large number of models, and thus it is suggested that this process should be automated. This work is an excellent demonstration of the power of chemometric approaches and their immediate potential benefits for speeding up the characterization of lignocellulosic materials.

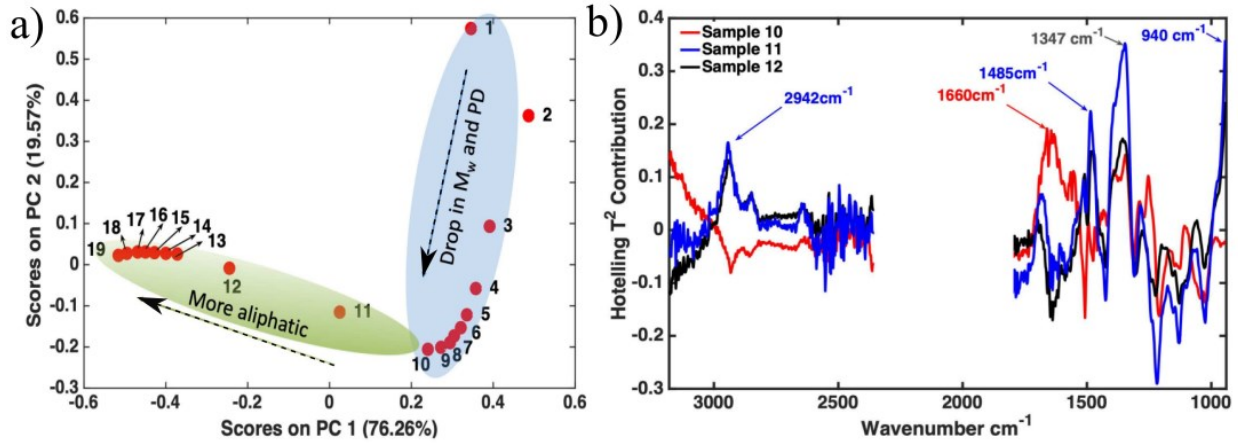


Figure 22. a) Score plot of PC2 versus PC1 of operando ATR-IR spectra corresponding to 19 samples acquired during aqueous phase reforming reaction. The dotted arrows show the direction of the trends in the data. The score plot explains 95.82% of the spectral variance. b) Scores contribution plot. Reproduced from Figure 7 of Khalili et al. [310].

Derkacheva and Sukhov applied chemometric techniques for resolving the spectral bands of isolated softwood lignins [312]. They used FT-IR spectroscopy to analyse mill wood lignin (MWL) and technical kraft lignins. A non-linear least-squares fitting of the FT-IR lignin spectra

was performed by the optimization of band parameters for a simulated spectrum of observed bands. Prior to the fitting procedure, the number of elemental bands was determined using Fourier self-deconvolution spectra of studied lignins and second derivative spectra. Their computer program optimized the elementary band parameters of the simulated lignin spectra until there was agreement with the experimental curve under the specified spectral range. Therefore, each studied lignin would have a simulated spectrum with a set of elementary spectral bands (Figure 23) which can be attributed to distinct molecular vibrations of lignin molecular units. From this, subtle structural changes of softwood lignins could be revealed. The authors report that the influence of isolation procedure on the lignin structures was observed.

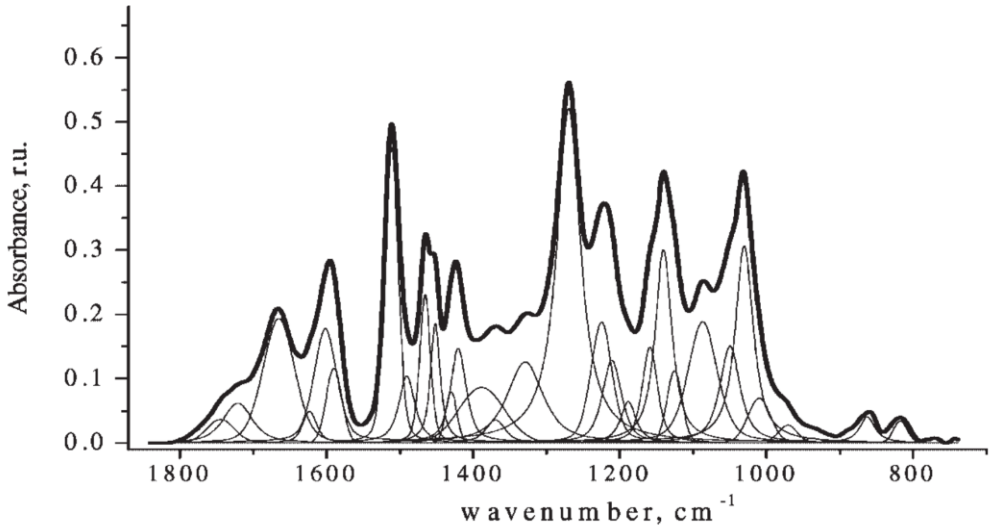


Figure 23. FT-IR absorption spectrum of Bjorkman's lignin (thick line) and its band model components (thin lines). Reproduced from Figure 3 of Derkacheva and Sukhov [312].

Kline et al. used PCA as a tool to assist in understanding the major and minor differences in the FT-IR spectra of different isolated lignin samples [313]. Preliminary PCA results identified significant structural differences between lignins isolated via hydrolytic and alkali methods and the other lignin standards isolated via the organosolv method. Thus, it was concluded that the chemical composition and structure of the lignin are strongly dependent on the choice of the isolation treatment. The authors then only selected the organosolv lignins for further PCA analysis, and produced score/loading plots (Figure 24). The score plot of the five organosolv lignin standards showed three distinct clusters of grass, hardwood, and softwood lignins. The loading

plot indicated the key differences responsible for grouping the samples along PC1, which were the higher S content for hardwood, the lower concentration of ester units in grass lignins compared to softwood, and the higher concentration of G units in softwood. In addition to the PCA of FT-IR data, Kline et al. also used PLS to determine the optimal wavelength for the lignin quantification using UV-Vis spectra [313].

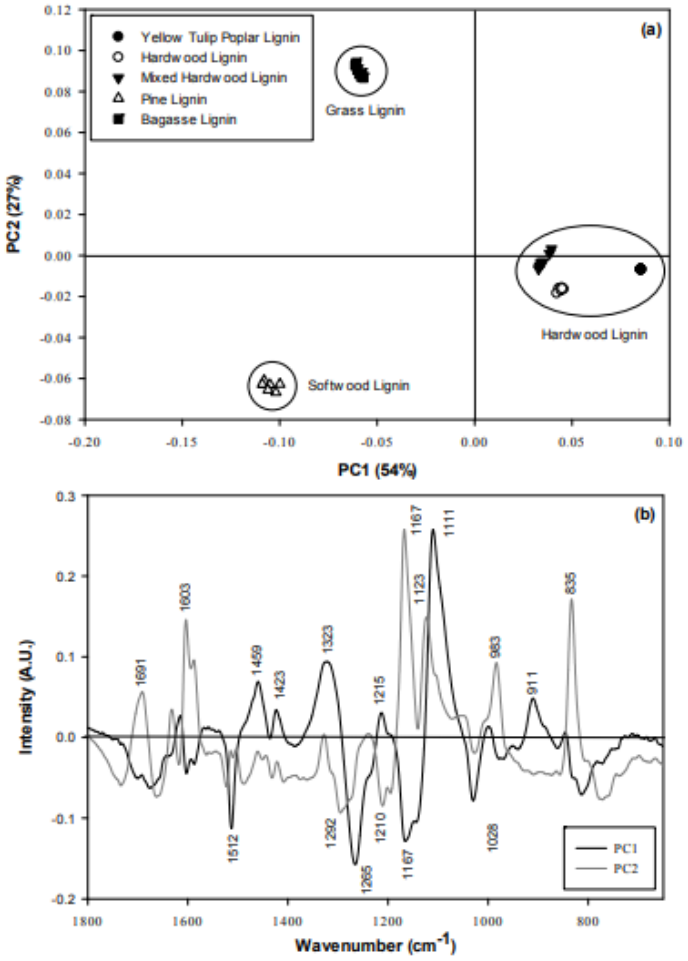


Figure 24. Results from the PCA of FT-IR spectra of 5 lignin standards isolated using the organosolv method: (a) Score plot as a function of the first and second principal components, PC1 and PC2. (b) loadings for PC1 and PC2. Reproduced from Figure 2 of Kline et al. [313].

Recently, Herrera-Diaz et al. have explored the structural changes occurring in precipitated kraft lignin with different oxidative enzymatic treatments to facilitate subsequent catalytic depolymerization for biorefining purposes [314]. Their main objective has been to select the most

suitable enzymatic dosage for achieving higher monomeric yields after the alkaline depolymerization process. They study the changes occurring in the lignin with pyrolysis gas chromatography-mass spectrometry (Py-GC/MS), high-performance size exclusion chromatography (HPSEC), and FT-IR. Despite confirming structural changes in the lignin samples from the enzymatic treatment with the different analysis techniques, no conclusive pattern could be identified and therefore it has not been possible to select the most useful pretreatment. To assist in their analysis, they employed chemometric techniques to the FT-IR spectral data to determine the most significant differences between the treatments. Data preprocessing steps that included filtering of raw signals using the Savitzky-Golay algorithm, scattering corrections, mean offset removal, smoothing, transformation, and centering, are applied. The PCA method is used to segregate the useful information in the data set. The HCA method enables the production of dendrograms to identify the most significant differences among treatments. The authors conclude that the combination of IR spectroscopy with chemometrics is a useful tool for discrimination among structural features and chemical properties, and could potentially be employed as a procedure for selecting the most appropriate chemoenzymatic treatment or kraft lignin valorization. The examples reviewed thus far illustrate the practicality of applying PCA and other chemometric techniques for providing effective insight into the chemical structure and composition of samples from complex MIR multivariate data sets.

5.6.1.3 Raman Spectroscopy

As a result of a significant laser-induced fluorescence from lignin, Raman spectroscopy has not been as widely applied to lignocellulosic materials as it might otherwise have been [142] [315]. However, the introduction of confocal Raman microscopy and later the 1064 nm NIR laser for sample excitation has mostly removed the issues with fluorescence in lignocellulosic material except for some technical lignins which can still produce significant fluorescent interference [315]. Recently, the NIR FT Raman approach has seen success in analysing lignocellulose samples with reduced time requirements and is likely to see more use in the lignocellulosic biomass field [316] [317]. Raman spectroscopy has been inaccurately interpreted to provide only the same information as IR techniques, but in fact it can provide complementary information. The tandem use of Raman and IR spectroscopy can enable the more comprehensive structural evaluations of lignin and biomass [260]. The Raman spectra of plant cells is somewhat limited by broad and overlapping

bands; however, multivariate data analysis approaches can help overcome this [318]. Raman microspectroscopy for Raman imaging has proven to be a powerful tool for probing biological materials [319] [320], and a large portion of the Raman papers identified in this review were micro-spectral studies. A list of the papers identified in this review that use Raman and chemometrics to study lignocellulosic biomass or lignin along with the preprocessing methods and multivariate techniques employed can be found in Appendix B.

Ona et al. tested the feasibility of FT-Raman spectroscopy for the rapid non-destructive determination of lignin S/G ratio in *Eucalyptus camaldulensis* and *E. globulus* of a variety of ages and colors [321]. They used second derivative after MSC prior to PLS regression to develop a model between S/G ratios determined by a modified thioacidolysis method and the Raman spectra of the samples. Based on a highly significant correlation between the wet chemical and Raman predicted values, they concluded that FT-Raman is valid for determining the lignin S/G ratio for *Eucalyptus* native wood meal samples, regardless of their age or color. A few years later, Ona et al. once again evaluated the feasibility of FT-Raman for the non-destructive quantification of wood constituents and wood anatomy [322]. The examined wood constituents were holocellulose, α-cellulose, hemicellulose, lignin, extractives, lignin S/G ratio, and hemicellulosic neutral sugar. The examined wood anatomy features included the proportion of cell types, cell length, cell wall ratio, cell width, as well as wall thickness of vessels and fibres. Similar to their previous work, they utilized both first and second derivative transformations after MSC prior to PLS regression. They observed a highly significant correlation between the predictions and conventional measurements, except for rhamnose and vessels ratio. It was concluded that FT-Raman was valid for assessing kraft pulp properties of *Eucalyptus* trees for both wood constituents and anatomy features.

Saariaho et al. used ultraviolet resonance Raman (UVRR) spectroscopy to define the characteristic vibration bands of model lignin compounds of H, G, and S lignin structures at three excitation wavelengths (229, 244, and 257 nm) [323]. Their results indicated that UVRR spectroscopy could be applied to determine chemical structures in lignin. They also concluded that the detection of the characteristic vibrational bands might be more challenging in pulps due to the altered lignin structures during processing, and that multivariate data analysis methods such as PCA and PLS could be useful in these cases. In a later study, Saariaho et al. [324] followed up on this conclusion by investigating if multivariate data analysis could assist in the interpretation of

the UVRR spectra of lignin model compounds, including the three main monomer units, C5 condensed and conjugated structures. They also sought to determine characteristic wavenumbers of different lignin structures with a PLS model. The complexity of the lignin model compounds made it difficult for the PCA model to detect all compound groups and their characteristic Raman bands. A PLS model between the **Y** values, which contained the quantity of the lignin structures, and the **X** variables, which were the measured spectrum variables, was constructed using the sub-structures of model compounds, quantified arbitrarily. Even though a single PLS model could theoretically predict all sub-structures in the compounds, their results indicated that simpler models are characterized by better predictions. They concluded that multivariate data analysis approaches could be instrumental in characterizing the chemical structure of a lignin sample. The authors also stated that quantifying additional lignin structures could be possible with the further development of PLS models.

Meyer, Lupoi, and Smith. used a home-built 1064 nm dispersive multichannel Raman instrumentation to show that Raman spectroscopy could be used to measure lignin composition [325]. Model lignin monomers and first derivative spectra or PLS analysis were used to determine sugarcane lignin monomer composition. Background subtraction was performed with a blank containing all matrix components except the analyte. The PLS analysis was used to predict the score coordinates for randomly selected monomer mixtures as a means of testing the model and quantitatively determine the monomer composition of the lignin. They concluded that using Raman spectroscopy, lignin monomer composition can be assessed qualitatively using first derivative spectra, and quantitatively using chemometrics, with PLS allowing for the determination of multiple monomer components simultaneously.

Lupoi and Smith et al. used Raman spectroscopy to characterize lignin content in different biomass feedstocks and validated their interpretations with CP-MAS ^{13}C-NMR [326]. They also quantified and compared lignin monomer composition measured with Raman spectroscopy to that of a modified thioacidolysis technique. Qualitative PCA accurately classified all but one feedstock, which had a unique lignin composition compared to the other sources. A PCR model was developed using the G and S lignin units determined by thioacidolysis followed by GC-MS and the Raman spectra. Overall, the PCR model was accurate in predicting G and S percentage; however, inaccurate results for two feedstock types indicated that PCR was not suitable for

biomass with a significant composition of extractable compounds. The authors noted that another PCR model could be developed for these samples with high extractables. In a later study, Lupoi et al. assessed and compared the three main vibrational spectroscopic techniques, MIR, NIR, and Raman spectroscopy combined with multivariate analysis to predict S/G ratios as determined using Py-MB/MS from 245 different eucalyptus and acacia trees across 17 species [260]. Different combinations of spectral preprocessing were tested to determine the most robust predictive model which included first and second derivative transformations with the Savitzky-Golay algorithm, smoothing, SNV, MSC, and EMSC. The reference set of wood samples for Py-MB/MS was selected using PCA that classified samples based on similarity. In this way, the PLS model could be developed to capture the variance likely to occur in future samples. The PLS model was developed from the spectral data determined using the vibrational spectroscopic techniques as the **X** matrix, and the S/G ratios determined using Py-MB/MS, as the **Y** matrix. It was observed that models constructed from the MIR and Raman spectra were more accurate at predicting S/G ratios, as compared to NIR. The authors concluded that the use of multivariate modeling with vibrational spectroscopy could reduce time and expenses with the accurate prediction of S/G ratios in a variety of feedstocks. Lupoi et al. also demonstrated the use of PLS to model acacia and eucalyptus lignin S/G ratios determined from Py-MB/MS and FT-Raman spectral data as a high-throughput and efficient biomass feedstock screening technique for the production of bio-based renewable fuels, chemicals, and materials [327]. They used first derivative Raman spectra and EMSC as preprocessing techniques based on the findings from their previous study [260]. However, as compared to the previous study, their goal in this analysis was to conduct a complete evaluation of the multivariate models for predicting S/G ratios using Raman spectra on a large unknown data set of acacia and eucalyptus samples in order to determine the practicality of the model for the assessment of future samples. The predicted S/G ratios were found to agree with those determined with Py-MB/MS, based on the mean values of each method within the 95% confidence interval.

Raman imaging and chemometrics have been used extensively to determine the distribution of lignin and other components in the plant cell walls [318] [328, 329, 330, 331, 332, 333]. Perera et al. [143] introduced a novel method for analyzing the structure of native lignin using Raman microspectroscopy and chemometrics. The data analysis involved data preprocessing, PCA, cluster analysis, and a deconvolution step to estimate the spectra of pure components. Wavelet

decomposition and second derivative transformations were used to reduce noise and fluorescent interference, respectively. The PCA technique was used for dimensionality reduction prior to *k*-means clustering to detect and classify compositional differences among samples. Average spectral identities were calculated for each cluster using a linear combination of the spectra pure components of the cell wall. A deconvolution methodology for estimating the spectra of pure components was done using a spectral entropy minimization [334] and simulated annealing optimization [335]. Lignin and carbohydrate spectral features were separated using this chemometric approach, which allowed for the recovery of information on different monolignol units that would be typically undetectable with routine data processing techniques due to signal overlap with cellulose and hemicellulose. Colares et al. [330] used the MCR-ALS to recover characteristic signals of cellulose and hemicellulose from the Raman spectra of mahogany and eucalyptus. The PCA method was used as a filtering approach prior to the MCR. The recovered characteristic Raman signals were used to estimate cellulose and lignin concentration distribution at the microscopic level, and represented as concentration maps (Figure 25). Their method successfully described the distribution of lignin and cellulose in two different wood species in a reasonable agreement with values obtained via the reference methods, high performance liquid chromatography and gravimetric wet chemistry. Recently, Zhang quantitatively visualized lignin in various plant cell wall layers during delignification using micro-Raman spectroscopy [333]. A sampling of the poplar tissues at varying periods during the delignification process and subsequent analysis with micro-Raman spectroscopy revealed unique insights into the mechanisms occurring in the plant cell wall during delignification process. This novel methodology could be employed in future studies that seek to further understand the changes occurring during delignification. Prats-Mateu et al. [318] investigated VCA, non-negative matrix factorization (NMF), and MCR-ALS for unsupervised spectral unmixing of Raman spectra of different plant tissues and two different species to compare the different algorithms, preprocessing steps, and other algorithm related input parameters. Raman spectra from a confocal Raman microscope were preprocessed by cosmic ray removal and background subtraction prior to multivariate data analysis. Background subtraction and the number of end members were assessed, and the analysis revealed that the multivariate data analysis was dependent on these factors; therefore, the careful selection of these strategies was required for every algorithm. The authors conclude that Raman spectroscopic imaging in

combination with multivariate data analysis could provide useful insights into the plant cell wall design.

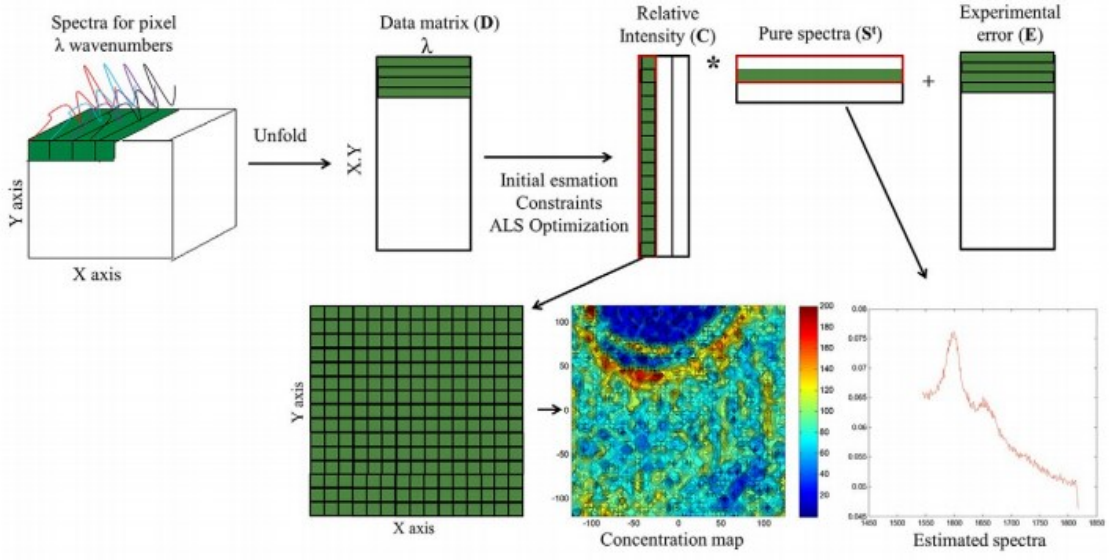


Figure 25. Schematic representation of bilinear matrix decomposition of a hyperspectral image using the MCR-ALS method. The MCR-ALS method is based on the bilinear model, $D = CS' + E$, where D is the raw data matrix, C is the matrix of relative intensity profiles, S' is the estimated pure spectra, and E accounts for the experimental error contained in the raw data. Reproduced from Figure 2 of Colares et al. [330].

Recently, Gao et al. have used chemometrics and FT-Raman spectroscopy to investigate lignocellulose materials [336, 337, 338]. Gao et al. [336] have collected Raman spectra excited using a 1064 nm laser to avoid the impact of water and auto-fluorescence, and employed a chemometric approach to assess the lignin content of poplar in large-scale breeding and genetic engineering programs. The rapid and non-invasive technique is valuable due to the high variation in lignin content among different poplar genotypes. Two different baseline correction strategies, two different data types based on scattering peaks intensity and area, and four different models are constructed using different algorithms, including PCR, PLS, ridge regression, and least absolute shrinkage and selection operator (LASSO) regressions to predict the lignin content of lignocellulosic materials. They found that the ridge and LASSO regressions were more suitable for predicting lignin content than PCR and PLS. It is concluded that the models are effective for

predicting the lignin content of poplar genotypes based on their Raman spectra. The authors highlight this approach as promising for large-scale breeding and genetic engineering programs. In an ensuing paper, Gao et al. [337] have proposed a novel strategy for quantitatively predicting holocellulose content using chemometric methods. Prior to model development, Savitzky-Golay smoothing, adaptive iteratively reweighted penalized least squares are performed for preprocessing. The authors evaluate a variety of algorithms (PCR, PLS, ridge regression, LASSO regression, elastic net regression) and 12 datasets to develop 60 different models. It was determined that FT-Raman spectroscopy and regularization algorithms, such as ridge regression, lasso regression, and elastic net regressions are successful in quantitatively predicting poplar holocellulose content. The PCR and PLS method are found to be insufficient for the prediction of poplar holocellulose content, which the authors propose could be the result of multicollinearity. Gao et al. [338] have also recently proposed a strategy to predict the Kappa number of bleached *Eucalyptus globulus* kraft pulps using FT-Raman spectroscopy and chemometrics.

5.6.2 Nuclear Magnetic Resonance

The NMR spectroscopy has been used extensively for characterizing lignocellulosic biomass samples and has greatly assisted in our understanding of the lignin structure [339] [340] [341]. Compared to other spectroscopic techniques used for characterizing lignocellulosic biomass and lignin, NMR has a much higher resolution and, therefore, can offer more information than the other techniques [127]. Specifically, 2D NMR methods are the most powerful method for analysing lignin, due to the advantages of the 2D spectra in allowing for the assignment of very complex lignin signals [162]. There are multiple solid and liquid state NMR techniques that offer different advantages and challenges. Common techniques include Proton (^{1}H) NMR, Carbon-13 (^{13}C) NMR, Phosphorus-31 (^{31}P NMR), 2D ^{1}H-^{13}C HSQC, and CP/MAS NMR [342]. A recent comparison of the NMR techniques for the determination of lignin S/G ratios has been published [343].

While NMR has the advantage of producing very detailed structural information, it also presents problems for data interpretation because of its large and complex data sets resulting from overlapping resonances and thousands of data points [344] [345]. Consequently, chemometrics approaches have the potential to be integrated into the analysis of NMR data to assist in identifying

underlying patterns in the data. Reviews of chemometrics applied to NMR spectroscopy can be found here [222] [345] [346] [347]. It is also interesting to note that quantitative NMR without chemometrics has been critiqued because oversimplified data analysis of such complex data sets can result in confirmation bias [222]. While NMR has been used extensively to characterize lignocellulosic biomass and its derivatives, the number of publications related to applying multivariate analysis to NMR data of biomass or lignin is noticeably lesser compared to the other popular analytical techniques such as vibrational spectroscopic methods or mass spectrometry (MS). We speculate that this may be due to the costly and time-consuming nature of the NMR technique as compared to vibrational spectroscopy, making it less suitable for rapid or high-throughput applications. Many of the chemometric approaches we have discussed focus on developing fast and robust calibration models for the rapid characterization of biomass such that they could be implemented for online processes monitoring or even in a portable instrument. This is currently less feasible for an instrument like NMR; however, it does not mean that chemometric approaches are not helpful for interpreting NMR data, in fact, the opposite appears to be the case. It has been pointed out recently [255] that there are no examples of using multivariate techniques for resolving overlapping peaks in 1D and 2D NMR profiling of lignocellulosic feedstocks, or multivariate modeling of specific ^{31}P and ^{13}C NMR profiles. From what we have been able to ascertain, this still holds. That being said, 2D NMR PCA loading spectra can facilitate interpretation of structural/compositional differences that may be undetectable by traditional peak picking approaches, and in that sense, these multivariate techniques have been used for resolving NMR peaks [348]. A list of the papers identified in this review that use NMR and chemometrics to study lignocellulosic biomass or lignin along with the preprocessing methods and multivariate techniques employed can be found in Appendix B.

Hedenström et al. [348] demonstrated how multivariate chemometric data analysis techniques could be utilized to gain more compositional information from the 2D NMR spectra of dissolved cell wall samples than could be attained with traditional spectral analysis (Figure 26). They showed that molecular level changes among sample types occurring in both lignin and polysaccharides could be visualized as high-resolution 2D NMR loading spectra. Two examples were investigated to evaluate the method, the first being a comparison of *Populus* tension wood to normal wood and the second being genetically modified *Populus* wood compared to wild-type

trees. Data preprocessing strategies, such as scaling and mean centring, were evaluated. In the first example, following preprocessing, the data was subjected to PCA to evaluate the variance among spectra and consequently differences in wood composition. It is important to note that the PCA of the entire spectral range could not deduce the compositional changes of lignin because of the relative decrease in total lignin amount. However, an isolated spectral region containing the aromatic peaks could be modeled separately for information on lignin compositional differences, which include the S/G ratio and degree of acetylation by p-hydroxybenzoate. In the second example, the authors utilized both PCA and orthogonal partial least-squares-discriminant analysis (OPLS-DA) for sample discrimination. The authors concluded that their approach was effective and reliable for revealing structural/compositional differences in all plant cell wall components caused by environmental or genetic modifications. They indicated that the multivariate approach allowed a significant advantage for the interpretation of structural/compositional differences among samples using the 2D NMR loading spectra.

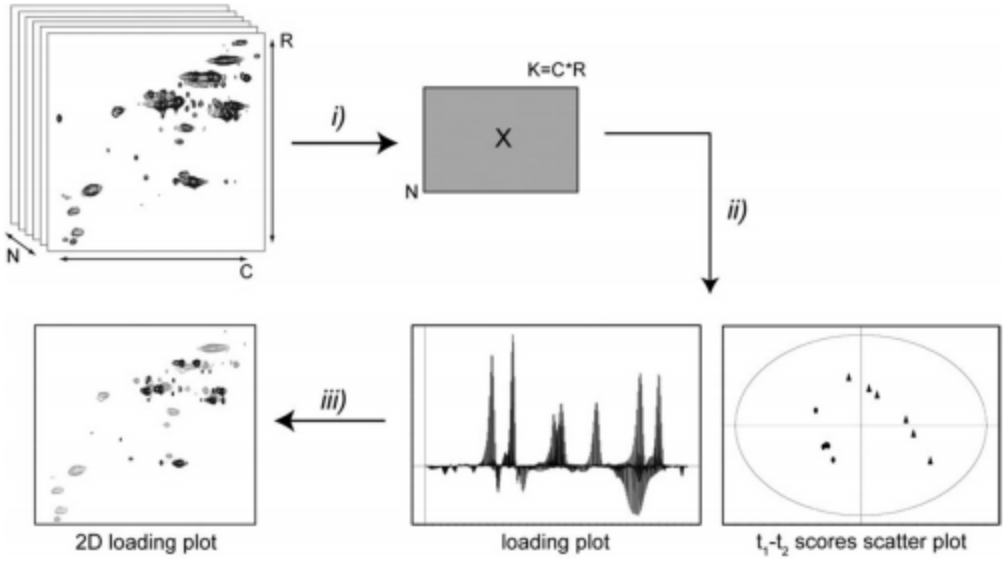


Figure 26. Illustration of the procedure for using multivariate analysis on 2D NMR data where (i) each of N spectrums with R rows and C columns are converted to a row vector and placed in a new data matrix **X**. Data points with intensity below a set threshold were considered noise and removed from the matrix. (ii) Scores and loadings resulting from multivariate analysis of matrix X are analyzed to detect latent structures in the data. The first and second principal components, t_1 and t_2, respectively, are the axis in the score plot. (iii) The loadings, initially represented as line plots of length K (number of columns in X), are converted to 2D loading spectra by reversing the unfolding procedure described in (i). Reproduced from Figure 2 of Hedenström et al. [348].

Hydroxyl groups are key functionalities of technical lignins that influence their properties and reactivity. Lignin hydroxyl group content could be an essential process parameter governing the efficiency of biorefining operations and impacting the properties of the resulting products [164]. Quantitative ^{31}P NMR spectroscopy is a promising technique employed for the specific determination of different lignin hydroxyl groups in lignin samples. The technique involves the phosphitlylation of lignin hydroxyl groups followed by quantitative ^{31}P NMR spectroscopy analysis [145].

Boeriu et al. [349] studied the fractionation of multiple different technical lignins (kraft, soda, organosolv) extracted from softwood, hardwood, and grass feedstocks using a selective extraction in a variety of “green” solvents and multivariate analysis methods. Unfractionated

starting technical lignin as well as the soluble and insoluble portions after selective solvent extraction were characterized and compared in terms of extraction yield, lignin Hildebrand solubility parameter, molecular weight distribution, and functional group comparison. Analytical techniques included quantitative ^{31}P NMR, FT-IR spectroscopy, and alkaline SEC. The ^{31}P NMR data were subjected to PCA in order to determine the difference in hydroxyl group composition among lignin fractions. The PCA results revealed a high heterogeneity of technical lignins and provided insights into changes in functional group content from extraction. The PLS was used to correlate FT-IR spectroscopic data with the chemical composition of lignin fractions determined by ^{31}P NMR. A calibration model was developed to predict the various chemical features of lignin, which included aliphatic and aromatic hydroxyl content (i.e., Alkyl-OH, condensed-OH, S, G, and H), carboxyl content (COOH), total aromatic hydroxyl (Aryl-OH), and S/G ratio, based on significant FT-IR spectral regions. The PLS models accurately predicted S lignin content, aliphatic-OH content, and S/G ratio. The authors concluded that their approach was simple, fast, and accurate for predicting structural and chemical information on lignin using FT-IR and chemometrics.

Aguilera-Saez et al. [350] used multivariate data analysis to predict lignin, cellulose, and hemicellulose content in greenhouse crop residues. The ^{1}H NMR spectra were subjected to PCA in order to visualize the variation in the large data sets. The HCA based on Euclidian distance coupled with Ward's minimum variance method was applied to provide information on the proximity between species, which PCA could not provide. To improve discrimination among species and select the most discriminant variables for regression, partial least square discriminant analysis (PLS-DA) was performed. The PLS-DA model enabled the rapid differentiation of all crop residues in their study. Linear models based on the 59 discriminant spectral buckets (the value of each bucket represents the total area within the respective spectral region) determined from PLS-DA were then developed to predict lignin, cellulose, and hemicellulose contents. The authors concluded that their prediction models were robust, and their method provided a rapid tool for the determination of the cell wall biomass composition of greenhouse crop residues.

Jalali-Heravi et al. [351] investigated the ability of a PCA-ANN method to calculate the ^{13}C chemical shifts of lignin model compounds and predict the chemical shifts of unknown compounds, in order to facilitate the future verification of molecular structures. Theoretical models

were constructed by relating atom-based structural descriptors to ^{13}C NMR chemical shifts. One hundred unique carbon atom types were identified, of which 73 were used as the calibration/training set, 20 for the test set to optimize the model, and 7 for the control set to evaluate the predictive ability of the model. A computer program was used to calculate the geometric, electronic, and topological descriptors of each carbon atom. The PCA method was employed as a feature selection method for determining the best descriptors for the model. An ANN model was then developed to predict ^{13}C chemical shifts using PCA's selected descriptors as inputs. From a data set that included 15 monomeric lignin model compounds, the authors found an accurate relationship between the experimentally determined and calculated ^{13}C NMR chemical shifts. They concluded that based on the accurate simulation of chemical shifts, researchers might be able to learn more about the structure of lignin compounds.

Burger et al. [352] used ^{1}H and DOSY NMR and chemometrics to predict the molecular weight properties (weight and number average molecular weight, and polydispersity) of organosolv lignin. A PLS regression model was developed using reference data on SEC-based molecular weights of lignin and ^{1}H and DOSY NMR spectra. For the DOSY NMR data, MLR was also used to develop a model. They found that while the ^{1}H NMR PLS model outperformed models generated from the DOSY NMR data, the MLR model outperformed the PLS model in DOSY NMR data. The authors also found that their ^{1}H NMR PLS model showed comparable prediction results to the work of Lancefield et al. [309], who predicted the molecular weight of lignins using ATR FT-IR spectroscopy (discussed previously in section 6.1.1). The results ultimately showed that ^{1}H and DOSY NMR combined with chemometrics were suitable for the accurate, high-throughput prediction of the molecular weight of lignins. The authors noted that acquiring a ^{1}H NMR spectrum takes two minutes, which is a significant improvement compared to the fifty minutes required for a SEC chromatogram. The researchers also indicated that they would be investigating multivariate data analysis methods, such as data fusion (e.g., NMR and IR data) to improve their high-throughput molecular weight predictions.

Calibration transfer is essentially the approach used to apply a data set and corresponding calibration model to two or more instruments, while retaining a desired model performance [353]. Chemometric models are often limited to the instrument and sample type that they have been developed with. Slight differences in instrumentation, changes in instrument performance over

time, as well as variation in sample batches are among the concerns regarding the transfer of models [354]. Calibration models provide rapid analysis after they have been generated, but developing a proper calibration model requires the use of the time-consuming analysis that the model itself seeks to help avoid. Additionally, these calibration models often require mindful planning from the design stage, sufficiently large sample quantities, and sufficient sample variation for its application. Thus, calibration transfer is an important topic for chemometricians, as it is desirable for models to retain their performance over time and in response to variation. Calibration transfer and practices have been relatively recently reviewed [353]. Recently, calibration transfer of multivariate calibration models between high- and low-field NMR instruments for predicting the weight average molecular weight of lignin has been demonstrated for the first time [355]. Low-field benchtop NMR instruments lower the barrier of entry for NMR analysis. While they have reduced resolution compared to high-field instruments, they are much less expensive to purchase and operate. This is important because it establishes the plausibility for high-throughput NMR analysis of lignin using low-field bench top NMR instruments. In their work, Lindner et al. [355] have found little difference in calibration errors for the prediction of lignin molecular weight, despite the large difference in resolution between their two NMR instruments (high-field 600MHz, low-field 43 and 60 MHz). Burger et al. [356] have since compared the performance of the high- and low-field NMR instruments for predicting the weight average and number average molecular weight and polydispersity index of lignin using chemometrics without calibration transfer. To their own surprise, the authors found no significant difference in performance between the high- and low-field devices. The approaches described here may facilitate a growth in the applications of chemometrics to NMR analysis of lignin and biomass by demonstrating that NMR approaches are more accessible than previously thought.

5.6.3 Mass Spectrometry

Mass spectrometry (MS) is an analytical technique in which an ion source is used to fragmentize compounds into ions, which are separated based on their mass-to-charge ratio (*m/z*), and detected in a mass analyzer [357]. The relative abundance of fragmented ions with a specific *m/z* are represented on a mass spectrum, which can be qualitatively and quantitatively evaluated to determine information on the original compounds molecular structure [358]. The MS is commonly coupled to other instruments when used for analysis of lignocellulose or its derivatives.

A disadvantage of MS techniques compared to some vibrational spectroscopic or NMR techniques is their destructive nature, which can be a concern if limited sample sizes are available. Some MS techniques such as Py-GC/MS do not require lignin extraction prior to analysis and therefore are sometimes considered to provide in situ structural information such as native lignin S/G ratios. However, the structural mechanisms taking place during biomass/lignin pyrolysis are not fully understood and therefore the assumption that the information from Py-GC/MS is representative of native lignin is an approximation. Recently, Zhang et al. [359] have published a review on characterizing lignin compounds using MS which covers analysis and raw data processing. A list of the papers identified in this review that use MS and chemometrics to study lignocellulosic biomass or lignin along with the preprocessing methods and multivariate techniques employed can be found in Appendix B.

One of the early applications of using an MS based technique and multivariate data analysis for analysing biomass was presented by Kleen et al. [360]. They employed Py-GC/MS to collect compositional data on softwood pulp fibres induced by kraft pulping and PCA to assist in the interpretation of the data set. It was clear from PCA that there were chemical changes occurring in the pulp fibres during the delignification process. A PLS analysis was then used to correlate the information determined from the Py-GC/MS data to the properties determined from more conventional wet chemical methods. Their models had a good prediction of lignin, glucose, xylose, and mannose contents; however, the models were less accurate in predicting the contents of the minor constituents, arabinose and galactose, as compared to the values determined by wet-chemical approaches. They noted that the difference in prediction of the minor contents were still in the same order of magnitude of experimental error for the wet chemical reference method. They concluded that the quantification of lignin and carbohydrates using their method was advantageous based on accurate results for the main constituents and good reproducibility for the minor constituents.

A number of publications from 1994 to 2017, all with connections to the National Renewable Energy Laboratory, USA reported the use of Py-MBMS and multivariate data analysis to analyze lignocellulosic materials [361, 362, 363, 364, 365, 366, 367, 368, 369, 370] [371]. The PCA and PLS were the most commonly used multivariate techniques in these publications which covered a variety of biomass types and properties, including bark source, acid concentration, and

phenolysis temperature [363], weight loss during brown-rot biodegradation of spruce wood, [362] lignin, glucose, xylose, mannose, galactose, arabinose, rhamnose contents [364], and distance from bark to pith [365]. These studies provided useful examples of the use of both PCA for gaining valuable insights into complex data sets, as well as PLS for quantitative and qualitative modeling, and prediction with generally positive conclusions about the use of multivariate data analysis and Py-MBMS for characterizing lignocellulosic materials. Decker et al. [370] reported on the development and deployment of high-throughput sugar release and Py-MBMS pipelines at the National Renewable Energy Laboratory. Additionally, Decker et al. [372] published a review on high-throughput screening techniques for biomass characterization.

Gerber et al. [373] introduced a data processing pipeline (Figure 27) for Py-GC/MS towards the high-throughput characterization of lignocellulosic materials. They highlighted sequential data processing that was common for typical Py-GC/MS data software as inadequate for automatic high-throughput applications and emphasized that the manual curation of peak lists with molecular identifications was a major hold-up for the technique. Their method included MCR by alternate regression and automated peak assignment and was evaluated using aspen and Norway spruce wood samples. The approach allowed for the accurate and fast estimates of relative amounts of lignin monomer units and carbohydrate polymers. The authors concluded that the integration of MCR-AR to the data processing pipeline provided a workaround of sequential data processing and the tedious manual analysis, enabling the application of the technique to larger data sets than previously feasible. They also noted that up to date (published in 2012), FT-IR spectroscopy or Py-MBMS instruments were typically selected for this type of application; however, Py-GC/MS had the advantage of the additional chromatographic separation and exhibits higher sample throughput compared to the aforementioned techniques. It might be noteworthy that based on our observation of literature after this publication, Py-GC/MS seemed to be more prominent for these types of applications.

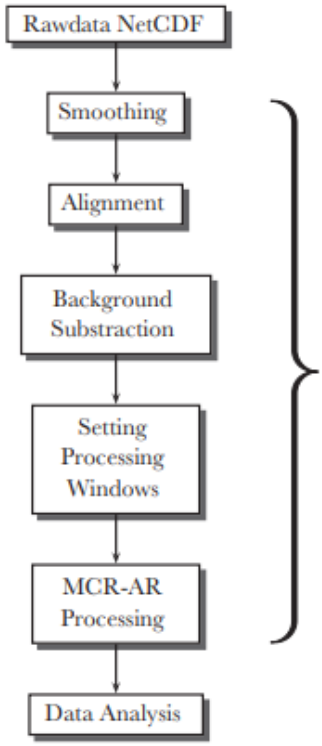


Figure 27. Flow diagram of the data processing pipeline for high-throughput Py-GC/MS analysis. The steps within the bracket were performed with a Matlab script. Data analysis, mainly peak identification and calculation of figures of merit was done in R. NetCDF is Network Common Data Form. Reproduced from Figure 1 of Gerber et al. [373].

Galetta et al. [374] used Py-GC/MS and multivariate data analysis to study the wood composition of Eucalyptus species relevant for pulping in Uruguay. Discriminant analysis with backward variable selection was used to simplify the models and successfully classify the wood groups. The PCA method was used for the unsupervised classification of wood samples and the identification of pyrolysis compounds characteristic to specific samples. The importance of pyrolysis products to predict pulping efficiency (alkali loading) was evaluated using simple, multiple, and PLS regression models. Their simple regression models based on individual compounds showed weak correlation with alkali activity, which suggested that the delignifcation efficacy is a multivariate interdependent property rather than a specific wood property. The PLS method was then used to predict pulping efficiency using 90 pyrolytic variables from the Py-GC/MS data. The PLS models confirmed the previous results from classification using

discriminant analysis with backward variable selection, in that reducing the data set to lignin-derived methoxyphenols with no carbohydrate-derived product improved performance. Multiple regression was employed with the same variable set determined from the previous analysis and confirmed the importance of these selected variables for describing the alkali efficiency. It was concluded that the most successful strategy for the regression models applied to pyrolytic data was data reduction, which facilitated the prediction of technical pulping parameters.

A series of papers from Ghent University in Belgium investigated genetically engineering poplars with Py-GC/MS and multivariate data analysis [375, 376, 377]. Toraman et al. [375] used PCA combined with *k*-means clustering to help understand the variation in the bio-oils produced from wild-type and genetically modified poplars. They demonstrated that significant differences could be achieved through genetic modification of genes involved in the phenylpropanoid and monolignol biosynthetic pathways, and specifically the contents of G and S lignin-derived phenolic compounds present in bio-oils could be modified. In a subsequent study, Toraman et al. [376] could not identify the minor compositional differences of lignin resulting from genetic modification of poplars using PCA. Therefore, they employed a multiway model PARAFAC2 (Figure 28) and PLS-DA. These techniques allowed for the accurate classification of poplars and could identify subtle information in large and complex data sets. SriBala et al. [377] aimed to evaluate the effects of downregulation of three enzymes on bio-oil composition using PCA. The PCA and *k*-means clustering compared with TA experiments showed that the multivariate models were able to interpret G and S lignin composition relatively accurately, which facilitated conclusions on the potential of the genetic modification for favourable compositional changes.

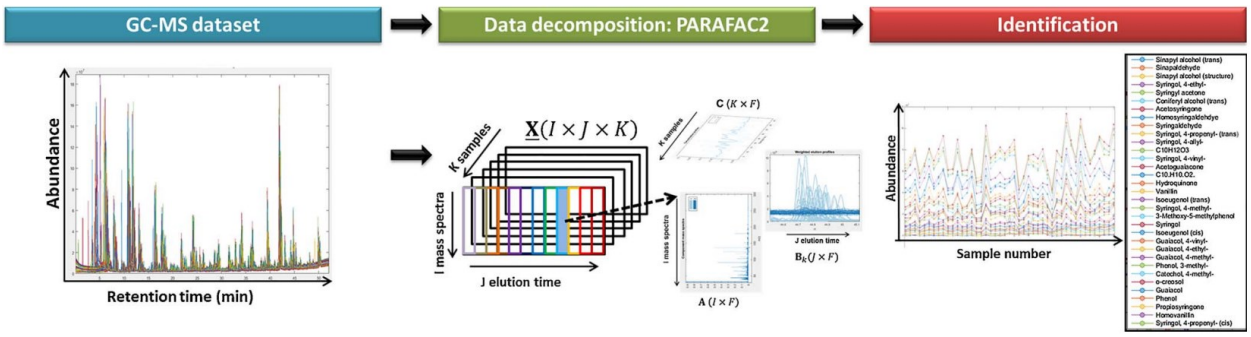


Figure 28. Flow-chart of PARAFAC2 methodology for GC-MS data analysis. **X** represents a three-way array of all overlaid chromatograms including mass channels, **A** is a matrix with the resolved mass spectra of the F analytes, **C** is the sample mode loading matrix with the relative concentration of each chemical compound in the sample, and $\mathbf{B_k}$ is a matrix with the estimated elution profiles for each of the F factors. Reproduced from Figure 3 of Toraman et al. [376].

We want to highlight a useful tool for expediting Py-GC/MS analysis of biomass used by Toraman et al. [376]. Py-GC/MS is a rapid method for identifying compounds and comparing their concentrations among and within samples; however, the complexity of Py-GC/MS data sets makes analysis challenging and time-consuming. Manual analysis can be tedious, error prone, and often requires a heuristics approach to learning. A total ion current (TIC) chromatogram can contain hundreds of peaks a number of which may suffer from co-elution (overlapping/embedded peaks), low signal to noise ratio, shifted retention times between samples/runs, and baselines. Thus, chemometric approaches have been used to automate the analysis process. An automated approach for processing raw GC-MS data within a MATLAB based graphical user interface has been developed and made freely available. The approach is called the PARAllel factor analysis 2 based Deconvolution and Identification System (PARADISe) and allows for the integrated multi-way modelling for processing of raw GC-MS data from several samples simultaneously [378]. GC-MS data can be arranged in a two-way matrix for each sample where the columns representing the length, I, mass spectra for each of the J elution times (rows). When a GC-MS data set contains more than one sample it becomes a three-way structure with K samples. Often when GC-MS is employed for studies there are many samples available and thus multi-way modelling has been proposed as a means for taking advantage of this three-way data structure [379]. The method has been demonstrated to be robust in peak quantification with many advantages compared to similar

commercial GC-MS software's such as AMDIS. More recently, the software was further updated to fully automate the approach, increasing analysis time and reducing user-interaction which can introduce bias and inconsistency [380]. This resource greatly lowers the barrier of entry for Py-GC/MS studies and is particularly suitable for biomass studies.

Reyes-Rivera et al. [381] used Py-GC/MS and chemometrics to characterize the 15 species of Cactaceae, classify the species, and identify compositional differences with taxonomic value. It was determined that Py-GC/MS combined with PCA, HCA, and hierarchical clustering on principal components with *k*-means partition were useful for the analysis of identity and chemical composition of materials (Figure 29). In addition, they noted that identified differences of certain lignin derivatives in different species, organs, or tissues can help observe and understand gene networks that determine the lignin composition of biomass.

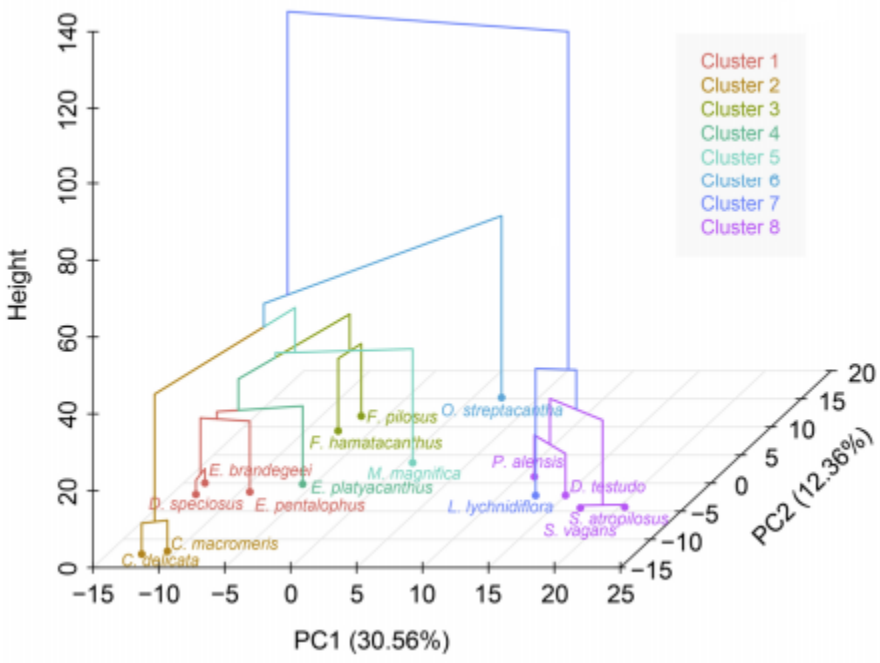


Figure 29. Hierarchical clustering on principal components with k-means partition plots obtained from the PCA reduced data sets considering all the derivatives from Py-GC/MS of the spines for one system. Reproduced from Figure 4 of Reyes-Rivera et al. [381].

5.6.4 Alternative Chemometric Approaches Applied to Lignocellulosic Biomass

This section covers chemometric approaches to study lignocellulosic biomass that do not fit into the main categories outlined so far in this review. This includes less common analytical techniques, molecular modeling applications, or approaches that use models with various input parameters. A list of these papers along with the preprocessing methods and multivariate techniques employed can be found in Appendix B.

Biomass pyrolysis is the thermal deconstruction of biomass without oxygen, leaving a hydrocarbon-rich gas mixture, an oil-like liquid, and a carbon-rich solid residue [382]. Pyrolysis is considered a promising potential treatment for the conversion of lignocellulosic biomass to biofuels or biochemicals. Understanding the reaction mechanisms of biomass pyrolysis is key for the success of this treatment [383]. The heterogeneous nature of biomass and its components makes the mechanisms of biomass pyrolysis extremely complex; therefore, there are major challenges to understanding the thermal deconstruction reaction pathways [384]. The TGA method has become attractive to study the pyrolysis kinetics of lignocellulosic biomass because of its high precision in weight loss recording, which makes it a method of choice the estimation of pyrolysis kinetic parameters and mechanistic investigations [385]. An approach for the kinetic analysis of complex processes with overlapping reactions is implemented with a peak deconvolution methodology using statistical functions that allow for the calculation of kinetic parameters from the separated peaks [386]. In a comparative study looking at deconvolution functions for fitting diverse kinetic pyrolysis models, Perejon et al. [386] have found conventional functions, such as Lorentzian and Gaussian, to be inadequate because of asymmetry. They revealed that the Fraser-Suzuki algorithm can accommodate asymmetric functions and fit kinetic curves for both ideal and non-ideal reactions. Therefore, the authors proposed the deconvolution of complex reactions into its individual processes using the Fraser-Suzuki function, followed by a combined kinetic analysis of the individual processes. The Fraser-Suzuki exponential function was originally developed for the resolution of overlapping bands and is commonly used for signal processing and data analysis in chromatography [387] [388] [389]. Since then, it has been utilized as an approach for separating lignocellulosic biomass pyrolysis into separate parallel reactions that model the decomposition of hemicellulose, cellulose, and lignin. This approach is useful because it has been

suggested that pyrolysis of biomass can be modelled as a weighted sum of the pyrolysis of these three main components [389].

Recently, the Fraser-Suzuki function has been applied in combination with other chemometric techniques and TGA to understand biomass pyrolysis [390] [391]. The high variability in the physical and chemical properties of biomass feedstocks can cause issues with meeting conversion specifications for the production of biofuels; however, the preprocessing of biomass prior to pyrolysis to produce more uniform properties can benefit thermochemical conversion [392]. Castro et al. [390] have compared two chemometric strategies for gauging the influence of chemical pretreatment prior to biomass pyrolysis. They use TGA and derivative thermogravimetric (DTG) analysis, combined with kinetic analysis using the Fraser-Suzuki deconvolution, Friedman isoconversional method [393], and modified Criado method [394] [395] to study the pyrolysis kinetics of pretreated peels of *Nephelium lappaceum* L. The two chemometric strategies are a supervised SVM learning algorithm and an unsupervised artificial neural network Kohonen self-organizing map (SOM). The authors note that the Fraser-Suzuki function is crucial for analysing the influence of chemical pretreatment. The SOM reveals that the similar structures of the pseudo-hemicellulose and pseudo-cellulose components promoted similar orders of thermal decomposition. The SVM model proves to be more effective in the prediction of chemical treatment based on kinetic parameters in their study, but both models show that the wide decomposition temperature range of lignin interferes with the precision of the algorithms. Virgens and Castro [391] evaluate pyrolysis routes for maximum biofuel production from *Syzygium malaccense* seeds using TGA-DTG, Fraser-Suzuki deconvolution, self-organizing maps (SOM), *k*-means clustering, and heat maps. Sixteen pyrolysis routes are investigated, which cover three chemical treatments (sodium hydroxide, sulfuric acid, and phosphoric acid) with no pretreatment and a variation in heating rate to evaluate their effects on biomass pyrolysis for the production of biofuels. The multivariate data analysis tools are employed to scan the data for groups with beneficial properties for biofuel production. A flowchart that illustrates their chemometric approach can be found in Figure 30. Before the multivariate data analysis could be implemented, the Fraser-Suzuki deconvolution has been applied to the DTG curve to separate each biomass component and the resulting data is normalized. Heat map analysis of all the pyrolysis routes allows the authors to identify two routes that show high levels of cellulose, hemicellulose, and

volatile compounds, and low levels of lignin content, which they consider promising features for biofuel production. Further chemometric tools corroborate their initial findings. The authors conclude that their chemometric approach is successful in screening pyrolysis routes for maximum gasification and the production of quality biofuels.

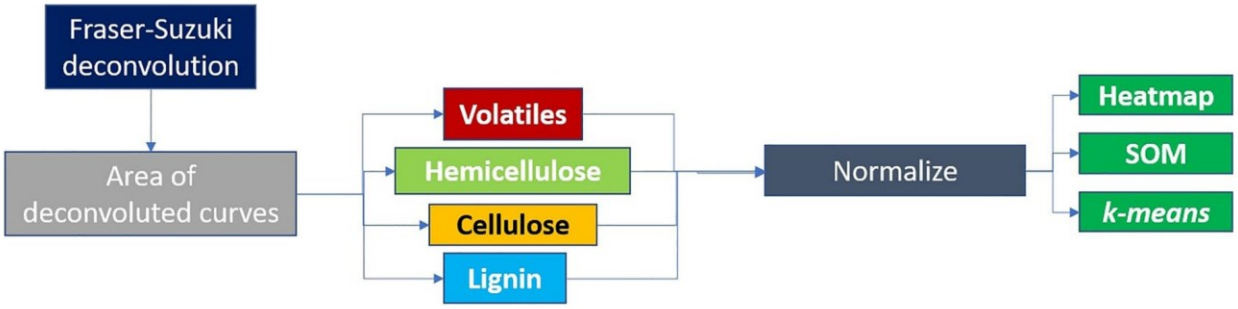


Figure 30. Flowchart for the chemometric approach for evaluating pyrolysis routes for maximum biofuel production. SOM are self-organizing maps. Reproduced from Figure 1 of Virgens and Castro [391].

Khaliliyan et al. [396] introduced a method for quantifying total lignin content by high-performance thin layer chromatography-densitometry combined with multivariate calibration. Their method was motivated by the challenges of the selectivity and robustness of analytical techniques for measuring lignin content. Selectivity refers to the capacity to discern lignin from the other components or impurities present in the sample. Robustness is related to reliable and reproducible results from an analytical instrument. The aim was to study their new method to determine the total lignin content of kraft and lignosulphate industrial liquor samples. Their method, which is based on multivariate calibration using PLS, determined lignin content but with some variation from the "correct" values. The authors noted that their method would be valuable if time was preferred over the accuracy, with analysis times of only a few minutes per sample, significantly shorter compared to the several days required for gravimetric analysis (i.e., Klason method).

Myrvold and Pavlov [397] used PCA to analyze correlations between the parameters of battery expanders and lignin properties. The PCA method allowed for the simultaneous analysis of the chemical composition of the lignin, the physical properties of the expanders, and the battery

performance parameters. It was concluded that the type of lignin was an essential factor, with few correlations occurring in multiple lignin types. Additionally, they found that a decrease in lignin solubility led to an increase in the battery life cycle.

Huang et al. [398] investigated the effects of the physicochemical properties of pretreated lignocellulose on their enzymatic digestibility using a combination of experimental methods and multivariate analysis. They evaluated the lignin content, accessible interior and exterior surface areas, removal of hydrogen bonds, and changes in crystallinity index of corn cobs after various pretreatments using wet chemical analysis, solute exclusion in a packed column, particle size analysis, FT-IR, solid-state ^{13}C NMR, and X-ray diffraction. The authors used PLS to assess the significance of different physicochemical characteristics in enzymatic digestibility. The PLS analysis determined that the extent of cellulose digestion depended primarily on the accessible interior surface area, lignin content, and the destruction of hydrogen bonds. Their results further substantiated the theory that accessible interior surface area and lignin content were important factors for cellulose hydrolysis. Additionally, they highlighted the utility of PLS for estimating the contributions of explanatory variables to target variables.

Lima et al. [48] investigated the relationship of lignin content and composition with embolism resistance and leaf lifespan in 22 tree species in a seasonally dry tropical ecosystem in North-eastern Brazil. They hypothesized that drought resistance would be determined based on lignin content and composition. The authors used PCA to visualize all the relationships between species and the different variables in order to see if species with different leaf lifespan would make distinct groups. They performed two PCAs, one with 22 trees using the variables total lignin content, S lignin content, S/G ratio, wood density, stem water potential in the dry season, and leaf life span, and the other with 15 trees using xylem vulnerability in addition to the previous variables. Their results indicated that dry-season leaf life span and resistance to xylem embolism were related to the monomeric composition rather than the total lignin content. Additionally, lignin content and S/G ratio were not found to be related or wood density and dry-season stem water potential. This is a good example of utilizing PCA to find relationships among large complicated data sets, and sets the stage for future biochemical studies regarding lignin chemistry, plant hydraulic traits and drought tolerance.

To better understand the feedstock variability of agricultural biomass, Ray et al. [399] evaluated the variability of corn stover of *Zea mays* at multiple scales from four counties in a high-yield region in Iowa. They focused on feedstock properties that could influence biomass processing using unsupervised machine learning algorithms to identify patterns. They used NIR spectroscopy to predict compositional characteristics, MLR and multinomial regression of red-green-blue image analysis to predict quality and composition, and *k*-means clustering based on the percentages of glucan, xylan, lignin, extractives, ash, and a variety of inorganic features to identify pattern variations. The organic features were predicted using NIR spectroscopy, demonstrating the applicability of multivariate calibration for supporting researchers with the high-throughput characterization of lignocellulosic materials. However, the NIR model was not able to predict the properties of all samples and highlighted the need to account for a wide range of variability. The multinomial regression of red-green-blue images was able to classify the material at 76.2% accuracy, showing promise but indicating the need for further improvement. The results of the *k*-means clustering of the biomass variables indicated that the major variability was based on region.

A key challenge in lignocellulosic biomass processing is the extraction or isolation of lignin from the other main components. There have been a variety of mechanical, chemical, biological, and thermal strategies for this purpose. Often these approaches are referred to as biomass pretreatment because they fractionate the material for further processing. Biomass pretreatments are complex processes with many variables and potentially inter-related interactions among these variables prompting the use of multivariate analysis. For example, Li et al. [400] have used PCA and PLS to investigate the relationships of key variables in alkaline sulfite pulping for the pretreatment of corn stover. They have conducted 28 lab scale pretreatment experiments on enzymatic saccharification by varying total alkali charge, liquid to solid ratio, temperature, cooking duration at the maximum temperature, and sodium sulfite/sodium hydroxide (Na_2SO_3/NaOH) ratio. The resulting biomass is analyzed for composition and enzymatic hydrolysis sugar yield, and the entire data set is used for PCA modeling. The PLS method is employed to model the relationship between pretreatment process parameters and the compositional variables of pretreated biomass (**X** variables) and the pretreatment effectiveness (**Y** variables). They found multivariate approaches useful for correlating variables in complex pretreatment data sets and provide tools for optimizing pretreatment conditions. More recently, Xu

et al. [401] [402] [403] have used chemometric approaches for investigating various biomass pretreatment strategies. Xu et al. [401] have studied 22 process parameters of sodium hydroxide assisted by mechanical refining pretreatment of corn stover. They use PCA to identify correlations among variables and PLS to study the inter-relations of pretreatment conditions and efficiency. Their results indicate that crystallinity of lignocellulosic biomass is negatively correlated with enzymatic efficiency demonstrating that multivariate analysis is useful for analysing inter-relations among variables in pretreatment and enzymatic hydrolysis.

Ionic liquids are another promising pretreatment solvent for lignocellulosic materials to decrystallise cellulose and dismantle lignin and hemicellulose networks, among other encouraging properties [404]. However, ionic liquids are non-environmentally friendly and have prompted the investigation of deep eutectic solvents as low-cost green alternatives [405]. Xu et al. [402] [403] have continued to investigate important parameters in the pretreatment of lignocellulosic biomass with deep eutectic solvents using chemometrics. Xu et al. [402] have applied chemometrics on 54 important variables of the choline chloride based deep eutectic solvent pretreatment of different lignocellulosic biomass materials. The PCA and PLS methods revealed that pretreatment temperature, polarity, molecular weight, boiling point, mass transfer capacity, hydroxyl group content, acidity, and hydrogen bond strength were the most important variables. These are strongly positively correlated with the removal of lignin and the recovery of glucan. Xu et al. [403] have also used chemometrics to investigate deep eutectic solvent pretreatment, this time with the hydrogen bond donors of different lignocellulosic biomass materials under different reaction conditions. They evaluate 42 key process factors and their relationships using PCA and PLS. The results indicate that hydrophilic ability, polarity, acidity, and ability to form hydrogen bonds with hydrogen bond donors are the most important variables for the removal of lignin in the deep eutectic solvent pretreatment of lignocellulosic biomass.

To compete with petroleum-based biorefineries with years of development and implementation, prospective lignocellulosic biorefineries must aim to optimize all aspects of their operations. To this end, Karlsson et al. have employed a chemometric experimental design strategy to investigate the dependence of lignin properties on the factors of their novel biorefinery strategy previously reported [202] [406]. Lignin properties are determined with HSQC, ^{31}P NMR and ^{13}C NMR spectroscopies, SEC, differential scanning calorimetry, X-ray diffraction, and wet chemical

analysis. They use a DoE approach based on the Box-Behnken design [407] to explore the potential for tuning lignin properties for valorization. The MLR method is used to generate models that include linear, quadratic and interaction effects for the experimental data of the lignin properties (Figure 31). Their approach makes it possible to identify the exact conditions required for targeted lignin properties, allowing for selecting priority inter-unit linkages in the lignin product. This result demonstrates the functionality and power of chemometric experimental design approaches for biorefinery research and development.

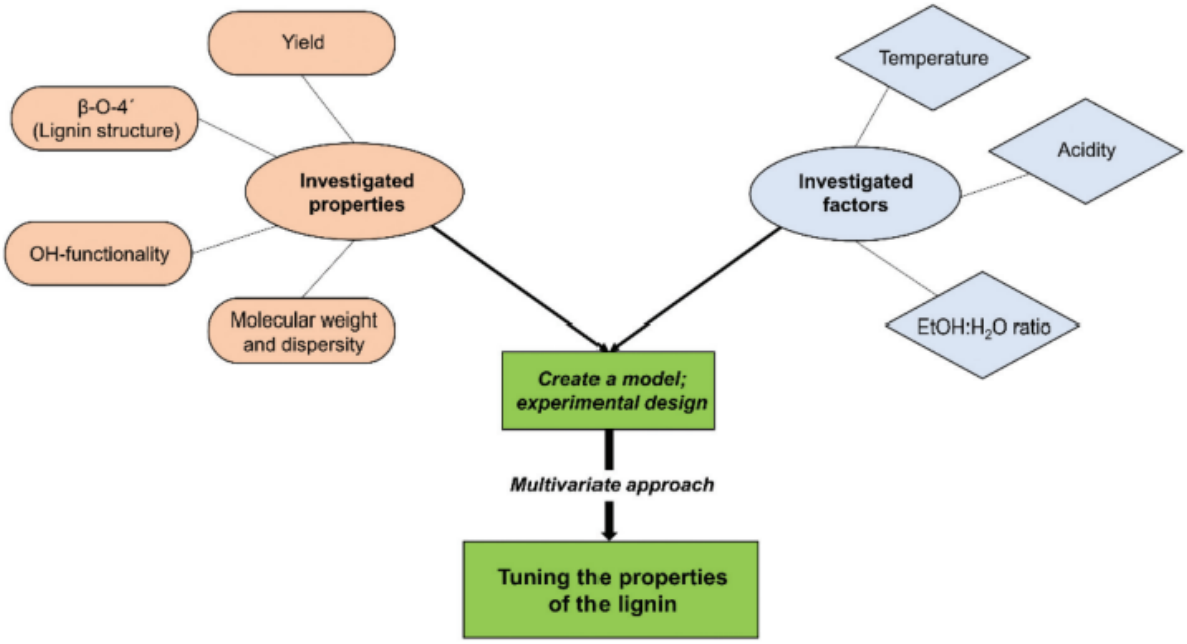


Figure 31. Schematic representation of the chemometric study setup using experimental design to create a multivariate model for a novel biorefinery strategy to identify the optimal conditions for targeted lignin properties. Reproduced from Scheme 2 of Karlsson et al. [202].

Computational studies allow for insights into systems that are out of reach of the current field of analytical chemistry at numerous length and time scales. Computational chemistry can assist in the prediction of the molecular structure and interactions occurring within the plant cell wall, or in the different stages of processing for valorization. It enables the identification and/or prediction of structure-property relationships, potentially providing key insights into the valorization process. It is critical to keep in mind that while computational chemistry and/or molecular modeling studies can provide valuable information on real systems, they are based on

models and the results are heavily dependent on how these models are generated. Basing models on experimentally determined data is important for coming to valid conclusions using a modeling approach. Combining chemometrics, chemoinformatics, and molecular modeling can facilitate the rapid and accurate production of structural models that represent real systems. As these methods and their integration progresses, we are likely to see accelerated advancements in our understanding of complex materials like lignocellulosic biomass.

Li et al. used a chemometric/cheminformatic approach for building a database of complex lignin structures and the prediction of thermochemical properties of lignin pyrolysis. A database of 4100 lignin related molecules and radicals was developed using the automated Rule Input Network Generator (RING) [408] [409]. The simplified molecular-input line-entry system (SMILES) notation was used to decode unique molecular lignin species and then convert them to initial 3D coordinates within OpenBabel and RDKit. Force field optimization and density-functional theory geometry optimizations were performed on specific low-energy molecules. After population of the database, PCA was employed for dimensionality reduction and to transform the groups into independent orthogonal PCA vectors using Scikit-learn, a machine learning toolkit [410]. The thermodynamic properties of the species were calculated using the Python Multiscale Thermochemistry Toolbox (pMuTT) [411]. A group additivity model following the Cohen and Benson's scheme [412] was used, and the group information was analyzed in the Python group additivity (pGrAdd) software developed by their group. They also introduced weak interactions such as hydrogen bonds and local steric effects to the group additivity-principal component analysis model using the SMARTS description. Their work demonstrated a computational framework of multiple chemometric and chemoinformatic techniques, tools for building a database of complex lignin structures, and accurate prediction of thermochemical properties of lignin pyrolysis products (Figure 32). This powerful example of fusing chemometric and chemoinformatic techniques illustrates the broad functionality of combining methods for real-world applications.

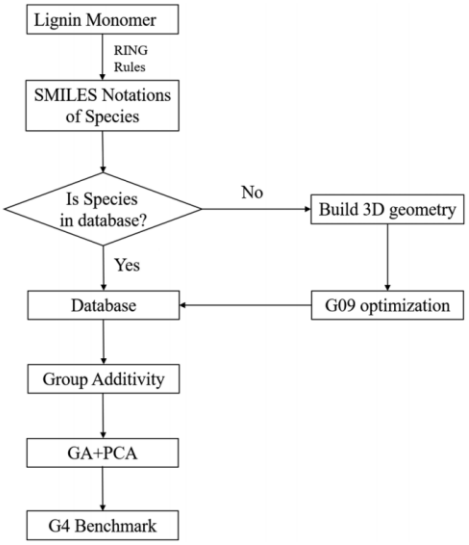


Figure 32. Schematic of the cheminformatic approach of a computational framework for database generation and group additivity (GA) model training. G09 is Gaussian 09 software, G4 refers to the Gaussian-4 theory level, and PCA refers to principal component analysis. Reproduced from Figure 2 of Li et al. [409].

Sattari et al. [413] introduced a chemometric/cheminformatic approach for identifying the reaction networks of complex mixtures with convoluted reaction chemistry using model compounds. The authors have shown the practicality of their approach for the identification of pseudocomponents using self-modeling multivariate curve resolution (a nonnegative matrix factorization method) and Bayesian hierarchical clustering. After pseudocomponent identification, Bayesian networks were used to detect directed pathways between the components and thus produce a proposed reaction mechanism. To validate, the researchers used a hydrous pyrolysis process with cellulose and lignin model compounds as feeds to hypothetic reaction mechanisms that were consistent with the known chemistry of the feedstocks. They employed data preprocessing such as baseline and background correction, smoothing, and PCA, for additional noise removal. A unique aspect of their work was the use of data fusion (Figure 33) between FT-IR and ^{1}H NMR spectra, as the data source for their approach. Additional data preprocessing through normalizing and variable selection was required. This work validated their method for identifying the reaction networks of complex mixtures and showed that the fusion of spectroscopic data can lead to better estimates more consistent with the known chemistry. Their data-driven

approach lowered the barrier of entry for the development of reaction networks because less expert knowledge was required than is commonly the case in analytical characterization. The approach also offered potential for computer process control applications. It would be important to highlight the crossover and smooth synthesis of chemometric and chemoinformatic approaches in this paper, as we believe strategies like this would become more common in the years to come.

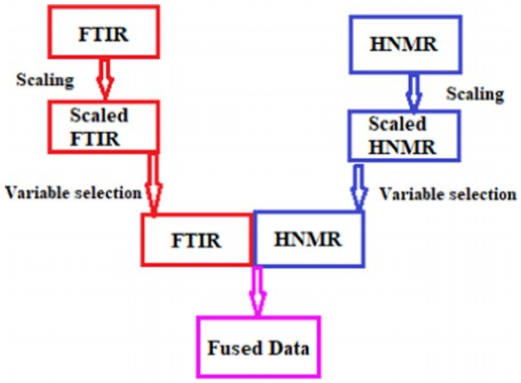


Figure 33. Data fusion scheme of FT-IR and ^{1}H NMR spectroscopic data using scaling and variable selection. Reproduced from Figure 3 of Sattari et al. [413].

In the context of polymer science, the glass transition is a temperature region where the mechanical properties of amorphous and semi-crystalline polymers are observed to change significantly [414]. Clearly, this is an important polymer property, yet this phenomenon is still not completely understood. The collective motion of many repeat units, known as segmental motion, is important for understanding the glass transition of polymers. Vural et al. have combined chemometrics and molecular dynamics simulations to investigate the segmental relaxation of lignin, which is the dynamical process that leads to glass transition [415]. They construct lignin models based on the 2D NMR spectra of vanilla stem lignin and use PCA of the collective motions of the molecules during the simulations to gain insights into the molecular mechanisms behind the glass transition. The results show that below the glass transition lignin has more internal and localized motions. In contrast, above the glass transition segmental motions of 3-5 monomers dominate, increasing chain mobility. Overall, the authors conclude that despite the high heterogeneity of lignin, the temperature dependence of lignin relaxation time is relatable to that of more simple polymers in terms of its Arrhenius behavior above and below glass transition. They

propose that these insights could be valuable when considering approaches for processing biomass at lower temperatures and cost.

5.7 Summary, Perspectives, and Future Outlook

While modern biorefineries appear closer to realization, there is no shortage of challenges to overcome and no firm consensus on the most favorable valorization pathway. Despite this, it is universally agreed that faster means of characterizing and analyzing lignocellulosic biomass will only expedite the shift to a global sustainable bioeconomy. Chemometrics offers a diverse toolbox of approaches that have been comprehensively applied. In this paper, besides the general characteristics of lignocellulosic biomass, we have discussed the challenges surrounding the efficient utilization of biomass resources. In particular, we have focused on lignin, as its structure and interactions within the plant cell wall are considered the key obstacle to the efficient isolation of the three main components. Lignin's complex, variable, and unresolved structure leaves many opportunities for innovation before the resource can be utilized efficiently. Much of our current understanding of lignocellulosic biomass and lignin has been determined through classical laborious wet chemistry approaches. Without diminishing the value that they have provided (and continue to provide), it can be said that these approaches are often costly, time-consuming, destructive and provide only specific/limited information. Many tedious classical techniques can be bypassed using high-throughput approaches combined with chemometrics, rapidly expanding the opportunities for larger scale investigations and screening of lignocellulosic material. Figure 34 shows the percent of publications and patents related to lignin, chemometrics, and their combination, per year, relative to their respective total for last ~20 years. As one can see, chemometrics and lignin are an expanding field despite the minor reduction in lignin specific publications in the last two years. Due to the broad relevance of chemometrics for a variety of problems, it is difficult to generically summarize the field. We will attempt to approach this task by categories below.

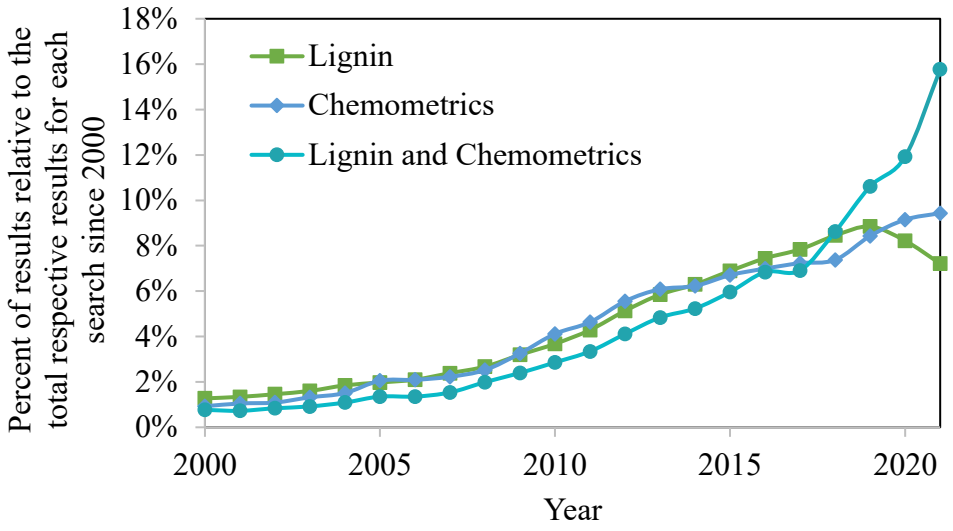


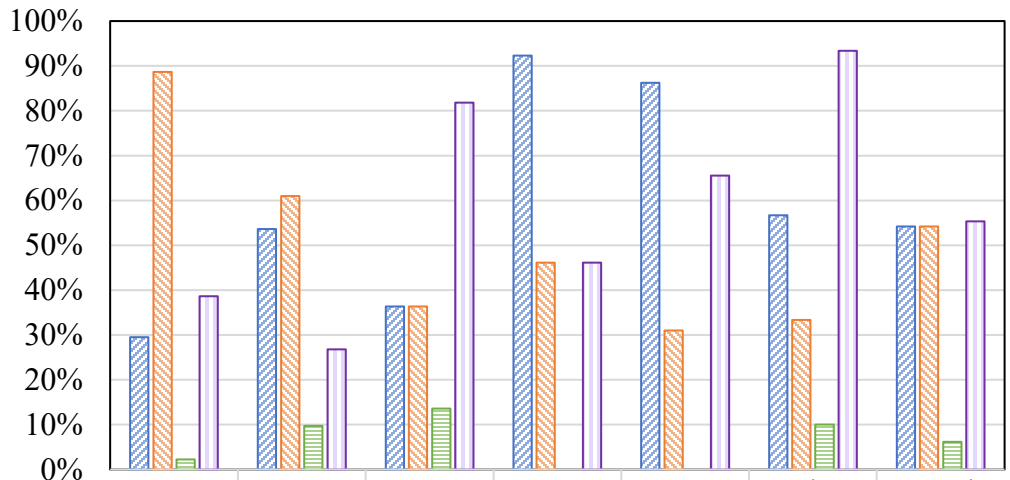
Figure 34. Relative number of scientific publication results on "lignin", "chemometrics" and "lignin, chemometrics" to their total respective publications in the last 22 years according to Google Scholar, searched on August 4th, 2022.

Experimental design is a major pillar of chemometrics, and these techniques should be kept in mind before one sets out to explore a hypothesis that requires large and complex data sets. Formal statistical design approaches can save time and reduce costs and should be especially useful for multivariate calibration approaches where high-quality models are desired. With that said, formal chemometric experimental design approaches have not been commonly observed in the literature explored in this review. While practical constraints such as limited time or sample size can be an obstacle, we encourage researchers to consider implementing these techniques where appropriate. While they add additional steps, they may end up saving time and/or increase the quality of the results, especially as the level of complexity of the application increases.

Data preprocessing strategies can clean, edit, reduce, and transform raw data. These methods allow for a better interpretation of data for analytical chemists and facilitate multivariate data analysis methods, which are sometimes hindered or incompatible with untreated data. Data preprocessing of some form or another is almost always used for data analysis and there are a variety of approaches available. There is no single preprocessing method that will work for every scenario so it is key to evaluate the different options and, where possible, one can consider rules of thumb as guides. It could be tempting to just select a preprocessing step from a dropdown menu

in a chemometric software without consideration of the theory behind the technique; however, we urge that a proper grasp of a technique prior to implementation is essential in order to understand the implications that the preprocessing may have. The selection of these preprocessing methods will directly impact how the data is analysed and the accuracy and performance of any models developed from the pre-processed data. Therefore, these strategies should be carefully considered and evaluated for each application. Since there has been an extensive use and evaluation of preprocessing strategies demonstrated in the literature, as apparent in this review, a good start can be to refer to previous studies with similar systems/objectives to get a general idea of suitable preprocessing strategies. We hope that the summary tables provided in Appendix B offer a straightforward means to narrow down the relevant literature. The area of data preprocessing is fairly mature, and there is not likely going to be a revolution in techniques. However, it appears that data fusion approaches may be a new frontier of data preprocessing. The fusing of complementary data sets from multiple data types and/or processing strategies may yield improved performance.

Multivariate data analysis techniques, which are the core of chemometrics offer both exploratory analysis and regression modeling. Lignocellulosic biomass is by nature information-dense, with a variety of components, bonds, and supramolecular interactions. To make things more difficult, lignocellulosic materials are also highly variable with both abiotic and biotic factors impacting the molecular scale composition and structure of the material. Multivariate data analysis approaches are particularly suitable for this type of problem which could potentially involve enormous quantities of samples and variables. Figure 35 shows the relative frequency of the main chemometric multivariate data analysis approaches for each respective category as used in this review based on the roughly 180 studies identified. One can see that both PCA and PLS were used in roughly 54% of the studies, while all other chemometric approaches together were used in just 61% of the studies, thus, demonstrating the ubiquity of these two fundamental chemometric tools in this field. In stark contrast to PLS, applications of PCR have lagged significantly.



	NIR	MIR	Raman	NMR	MS	Other	Total
PCA	30%	54%	36%	92%	86%	57%	54%
PLS	89%	61%	36%	46%	31%	33%	54%
PCR	2%	10%	14%	0%	0%	10%	6%
Other	39%	27%	82%	46%	66%	93%	55%

Figure 35. Percentage of occurrences of chemometric methods applied in the studies identified in this review, respective to their category of application. Note that multiple chemometric methods may have been used for an individual study and thus the percentages do not equal one hundred. The "Other" techniques include but are not limited to any multivariate data analysis approaches such linear regression, MLR, MCR, cluster analysis, discriminant analysis, alternative algorithms and neural networks. Data preprocessing, experimental design and validation methods were not considered in these calculations.

Exploratory analysis using multivariate techniques can support the identification and recognition of unknown patterns in large, complicated data sets. The PCA can provide an unbiased first look at data by simultaneously assessing all variables, allowing for the detection of unknown patterns, revealing subtle structural differences and reducing instances of confirmation bias. The graphical representations of data in the using PCA offers a human interpretation of the vast amounts of complex information.

Regression modeling allows for the accelerated characterization of biomass by correlating objects and their observed features to object properties. A large portion of the literature regarding chemometrics applied to lignocellulosic biomass involves the calibration of models for making

predictions. This is likely because the tedious traditional analytical approaches, such as wet chemical analysis, can take hours to days for just a few samples, which is not a feasible means of screening such variable feedstocks as lignocellulosic biomass. Thus, building models that can rapidly predict chemical properties will be instrumental for managing the variable nature of lignocellulosic feedstocks. A note of caution here would be that calibration models often face challenges with prediction outside of the samples used for calibration. To account for this, robust models will have to be generated that can encompass the wide variability of these feedstocks. To meet this goal, more advanced regression models might be required. Figure 35 clearly demonstrates the prevalence of PLS for regression in the literature but we also see chemometric regression techniques shifting to more sophisticated machine learning algorithms, and this will likely improve prediction performance for modelling very complex systems. Yet, it is important to understand the objectives of an approach prior to application. For example, the more classical chemometric regression algorithms, such as PLS and PCR, can be favourable for interpretation because machine learning algorithms can be challenging to understand, especially as the complexity of the models increase. Additionally, a cornerstone of chemometric analysis is the concept of parsimonious modelling – keep in mind that the simplest model with the best performance is preferred. Another area for of interest for the rapid screening of feedstocks are portable instruments, such as portable IR spectroscopy. Handheld devices with internal algorithms can be used for the rapid prediction of feedstock properties, which could be a very useful development. On another note, there is a significant value of chemometrics for correlating complex genetic information with biomass properties to support plant breeding programs producing optimal feedstocks for biorefining purposes. Also, desired feedstock traits for efficient processing could be segregated at the forest level prior to their entrance to the biorefinery.

The primary use of chemometrics for the studies of lignocellulosic biomass has been for vibrational spectroscopic data through multivariate modelling of spectral features and biomass properties, such as wet chemical data on lignin, cellulose, and hemicellulose content (Figure 36). The literature indicates these approaches are generally fairly successful in prediction accuracy and performance. Yet, these techniques sometimes have difficulty predicting some of the more minor components of biomass. Future efforts may look to develop models that can predict major and minor components with similar accuracy. When it comes to vibrational spectroscopy and

chemometrics, IR spectroscopy is the frontier of this area, but Raman is becoming increasingly popular. Both have seen success with chemometric approaches. Challenges with fluorescence, overlapping bands, and the misconception that it provides only equivalent information to IR spectra are responsible for the lagging applications of chemometrics and Raman spectroscopy for studying lignocellulosic biomass. Innovations in instrumentation of Raman in addition to improved data analysis approaches have changed this, and recently we have seen more examples of chemometrics techniques being applied. Additionally, vibrational microspectroscopy, which allows spatially resolved *in situ* characterization and imaging of lignocellulosic material over microscopic dimensions, has emerged as an approach well suited to chemometric analysis because of the large data sets produced in the form of hyperspectral datacubes.

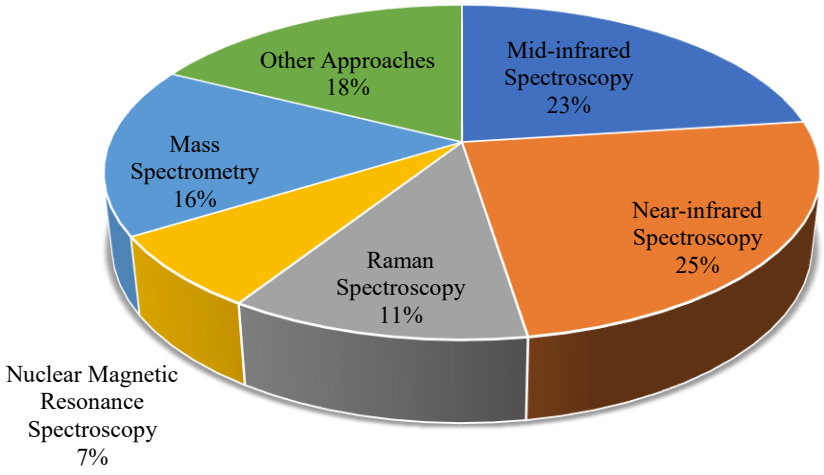


Figure 36. Distribution of methods used in combination with chemometrics in investigations related to lignocellulosic biomass based on the one hundred and eighty-three studies identified in this paper.

NMR spectroscopy may be considered the most comprehensive lignin analytical technique with high resolution and unmatched detail. The detailed information that NMR provides also results in large complex data sets, which are known to be suitable for chemometric approaches. Despite this, our review of lignocellulosic biomass literature found that applications of chemometrics combined with NMR are trailing compared to other analytical techniques such as

vibrational spectroscopy or MS. It could be speculated that reduced observations of chemometric NMR studies is the result of the fact that NMR is more expensive and time consuming. Hence, NMR is not necessarily suitable for rapid characterization of samples in the same way that vibrational spectroscopic techniques are, and this may be why there are less chemometric approaches. Yet, recently approaches that utilize low-field benchtop NMR instruments may lower the barrier of entry for NMR analysis by reducing cost, maintenance, and analysis time. These approaches may change the nature of future NMR investigations and open up possibilities for high-throughput NMR analysis. Other reviews including this one indicate a gap in chemometric applications to NMR analysis of lignocellulosic biomass but speculate that this disparity may be addressed in the near future. Given the excellent detail that NMR can provide on lignocellulosic samples, we aim to encourage more of these applications.

The MS technique has had many successful applications of chemometric approaches to investigate lignocellulosic biomass. The Py-GC/MS is one of the main techniques and has been widely used to characterize biomass because of its rapid and sensitive nature with simple sample preparation and quality chemical fingerprints. When analysing large GC-MS data sets, however, the manual curation of peak lists and settlement of matches makes the data analysis process long and requires expert knowledge or a moderate level training [373]. Therefore, chemometric approaches for data processing pipelines may be key for the successful analysis of large complex lignocellulosic data sets. The PARADISe software for multiway modeling of GC-MS data sets or similar approaches could be quite practical for this type of application. Since Py-GC/MS can provide rapid, sensitive, and reproducible compositional analysis of lignocellulosic biomass, including the relative proportions of S/G/H lignin monomer units, it is particularly suitable for the analysis of biomass feedstocks. We expect to see more research and development of chemometric techniques for Py-GC/MS data analysis in the near future.

Occasionally, chemometrics can be inaccurately described as techniques for analysing spectroscopic data, potentially suggesting that they are limited to this alone; however, they can be applied to a wide variety of data types. Chemometrics have recently been demonstrated to predict molecular weights (conventionally determined using SEC) using vibrational spectroscopy. Chemometrics have been used for optimizing pretreatment conditions for biomass processing, including to select for desired lignin inter-unit linkages in the resulting lignin product, which is of

great significance for producing lignins with targeted properties [202]. They have been combined with the kinetic analysis of thermogravimetric data to understand the impact of chemical pretreatments on the main components of biomass [416] [390]. Correlations between parameters of battery expanders and lignin properties have been analysed for electrical storage [397]. The feedstock variability of agricultural biomass has been investigated using a variety of parameters that influence biomass processing in order to maximize end use application [399]. The relationships among lignin content, composition, and other parameters have been investigated for their role in drought resistance [48]. The variety of examples listed here demonstrate the broad applicability of these techniques, which have endless potential.

Computational chemistry and molecular modeling have become prominent disciplines for investigating the structure-property relationships of systems at multiple length and time scales [417]. Molecular modeling provides opportunities for the investigation of certain systems that are not yet (and may never be) possible under experimental conditions. It can be a rapid and cost-effective technique for the testing of scientific hypotheses and drawing meaningful conclusions about real systems. Molecular modeling of lignocellulosic biomass has already proven its ability to assist in the understanding of lignin's structure and interactions with the other plant cell wall components [418]. As the field grows, multiscale modeling is likely to play a key role in the research and development of lignocellulosic biomass. We would like to highlight the opportunity to see more combinations of computational chemistry approaches with chemometrics to solve real world problems. A few approaches of this nature have been identified in this review, such as the investigation of the glass transition temperature using molecular dynamics simulation and PCA, which reveals that lignin relaxation time is relatable to that of less complex polymers in terms of its Arrhenius behavior above and below glass transition [415]. Another work has investigated the thermochemistry of lignin structures on a molecular level using density function theory, chemometrics, and chemoinformatics [409].

5.8 Conclusions

The transition from conventional energy and materials sources to more sustainable alternatives is considered a top priority. Nonetheless, despite the wide variety of renewable technologies that have been introduced, no single one has dominated markets to this point due to

the unique challenges that are faced by each technology. Lignocellulosic biomass has a distinctive position as a widely accessible feedstock that has been underutilized and undervalued for decades. Recently, lignocellulosic biomass has caught more interest as the attitude towards lignin has transitioned from a low-value waste stream to a valuable material whose chemistry and interactions affect the value of the entire feedstock, as well as its own value as a renewable source of carbon with potential for a wide range of value-added applications.

Like many new technologies, lignocellulosic biomass faces many challenges that have hindered widespread utilization to date. Specifically, its complex chemistry and variability are barriers to its full potential. Research of lignocellulosic biomass and its components has traditionally relied on time-consuming, costly, and information limited laboratory techniques; however, the field is trending towards more advanced approaches. In particular, advancements in spectroscopic techniques have transformed the field with more detailed and higher volumes of information.

Chemometric approaches offer sophisticated approaches for handling, interpreting, analyzing, and identifying patterns in data. These techniques have gained popularity as researchers have begun to identify the utility that they provide. Whether applied to analytical data on the research scale for key insights on a plant's chemistry in relationship to its observed properties, or the industrial scale for rapid online process monitoring of physical/chemical properties of a biomass feedstock, chemometrics are very powerful tools for the future of the field. In addition to analytical approaches, a newer computational space has emerged for the tackling of chemistry problems, so-called chemoinformatics. Chemoinformatics expands the chemical space, provides platforms for knowledge storage, and offers tools and algorithms for a wide variety of chemical problems. Challenging problems require innovative solutions, and our understanding of lignocellulosic biomass will only continue to develop in parallel with chemometric and chemoinformatic methods. We hope that this review will stimulate enthusiasm on this topic leading to more efficient research and efficient utilization of the resource.

Chapter 6: Assessing the impact of drought stress on the composition of Douglas-fir wood and the structure of softwood lignin for valorization

6.1 Introduction

Rising concerns of anthropogenic contributions to climate change accompanied by mounting global demand for secure energy, materials, and chemical sources have placed pressure on researchers to evaluate contemporary alternatives to conventional petroleum-based feedstocks. Lignocellulosic feedstocks are widely available, renewable, and in theory sustainable, and therefore, have garnered interest as a promising fungible substitute.

Lignocellulosic biomass can essentially be considered dry plant material composed primarily of three main biopolymers, cellulose (35-50%), hemicellulose (20-35%), and lignin (14-26%), which are organized together in highly recalcitrant and complex 3D supramolecular structures in the plant cell wall (See Figure 37) [419][420]. Plant cell walls are generally categorized as either primary or secondary cell walls, where the former is established early on and provides basic support and protection while the later is formed later in development and confers strength to the walls [421]. At the microscopic scale, the secondary cell wall components form structures known as macrofibrils. The macrofibrils are ordered structures formed by groups of crystalline cellulose microfibrils coated with amorphous cellulose and hemicelluloses. Some lignin is closely associated with both hemicellulose and some cellulose, likely localized to the surface [422]. Lignin is a complex amorphous polymer composed predominantly of three 4-hydroxyphenylpropanoids distinct in their number of methoxy groups on the 3 and 5 position on their phenyl ring [164]. These are known as *p*-coumaryl (4-hydroxycinnamyl), coniferyl (3-methoxy-4-hydroxycinnamyl), and sinapyl (3,5-dimethoxy-4-hydroxycinnamyl) alcohol, or as *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units, respectively, in the context of the lignin polymer. [164] [165] Lignin provides strength, rigidity, and hydrophobicity to the secondary plant cell walls [423] and is also commonly recognized as *a* main or *the* main contributor to the recalcitrant nature of lignocellulosic biomass to industrial processing [424]. As a result, an entire field of multidisciplinary research surrounding lignin has developed over many years, and in

particular, the field has shown rising momentum in the last 10-20 years as demonstrated by the increased number of lignin related academic research papers and patents [11] [425] [426] [81] [427]. The scope of this field is broad and encompasses multiple length and time scales ranging from understanding the genetic regulatory networks involved in lignin biosynthesis, molecular scale investigations of lignin's polymeric structure and interactions within the plant cell wall, and targeted engineering of lignin-based products with desired properties for specific applications. A common goal of these efforts is to better understand lignin to more efficiently utilize lignocellulosic biomass. In recent years, the term biorefinery has gained some traction although not well defined [162]. Here, we consider a biorefinery to mean any integrated industrial biocomplex that utilizes any biomass feedstock(s) to produce products such as bio-based chemicals, materials, or fuels [9]. The concept of a biorefinery is not especially novel, however, the interpretation has shifted slightly in recent years. While previous definitions may not have included a focus on effectively utilizing lignin process streams, contemporary visions consider lignin valorization as vital for economic viability of the industry [161].

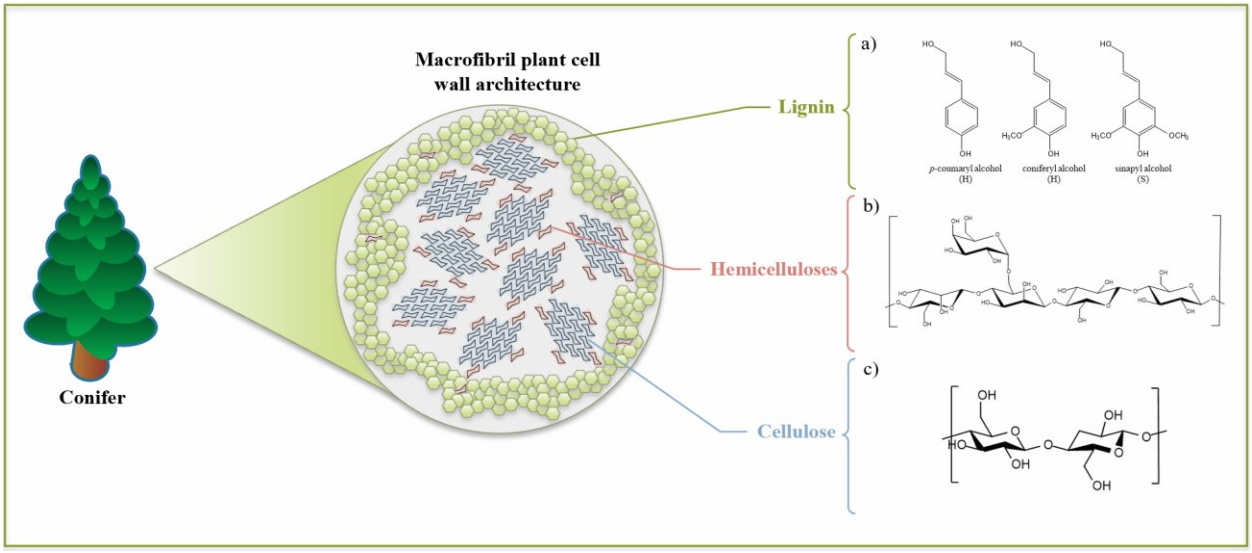


Figure 37. A representation of the molecular architecture of the softwood macrofibrils based on a recently proposed model by Terrett et al. [422]. Where a) Cellulose is a semicrystalline linear polymer solely composed of D-glucopyranose (glucose) units linked via the $\beta-(1\rightarrow 4)-$glycosidic bonds [24] [428]. b) Hemicellulose is an amorphous branched polymer composed of pentoses (D-xylose, D-arabinose), hexoses (D-mannose, D-glucose, D-galactose) and occasionally uronic acids and acetyl moieties as side chain groups [24] [25]. c) Lignin is a complex amorphous phenolic polymer composed predominantly of three 4-hydroxyphenylpropanoids which are distinct in their number of methoxy groups on the 3 and 5 position on their phenyl ring [164].

An inherent challenge of lignin science and engineering relates to its natural variability in composition related to its low degree of order and high heterogeneity of structure [429]. A greater fundamental understanding of the nature of lignin will support the effective valorization of lignin and more efficient utilization of lignocellulosic biomass. The phenylpropanoid biosynthetic pathway responsible for producing lignin and other hydrophobic polymers in the plant cell wall determines the composition and structure of lignin [423]. Differences in lignin content and its composition, such as relative proportions of its three main monomer units, linkage types, or functional groups, may significantly impact the structure and interactions occurring in the plant cell wall. Consequently, the physical properties of plants are also determined by the synthesis and

deposition of lignin in the plant cell wall [423]. Thus, intensive research efforts have been dedicated to its biosynthesis, a detailed description of which can be found elsewhere [430] [431].

A key interdisciplinary overlap occurs here as researchers explore and connect plant biology and its evolution with plant cell wall chemistry and the resulting implications for bioengineering strategies. A successful next generation bioeconomy will be best facilitated by a balance between selective breeding strategies for plants that are viable, productive, and tolerant, while also retaining desired compositional traits for processing into valuable products (See Figure 38). Both selective breeding and genetic modification are currently being utilized and investigated for producing plants with desired traits. Efforts have traditionally aimed at reducing lignin content in order to decrease costly and environmentally harmful delignification for pulp and paper production, and to increase the efficiency of enzymatic conversion of biomass into fermentable sugars in biofuel production [432]. Recent efforts have included targeted modification of lignins structure by altered expression of lignin biosynthesis and transcription factor genes to produce lignins of altered composition, including encouraging the incorporation of alternative lignin monomers units into the polymeric structures [424].

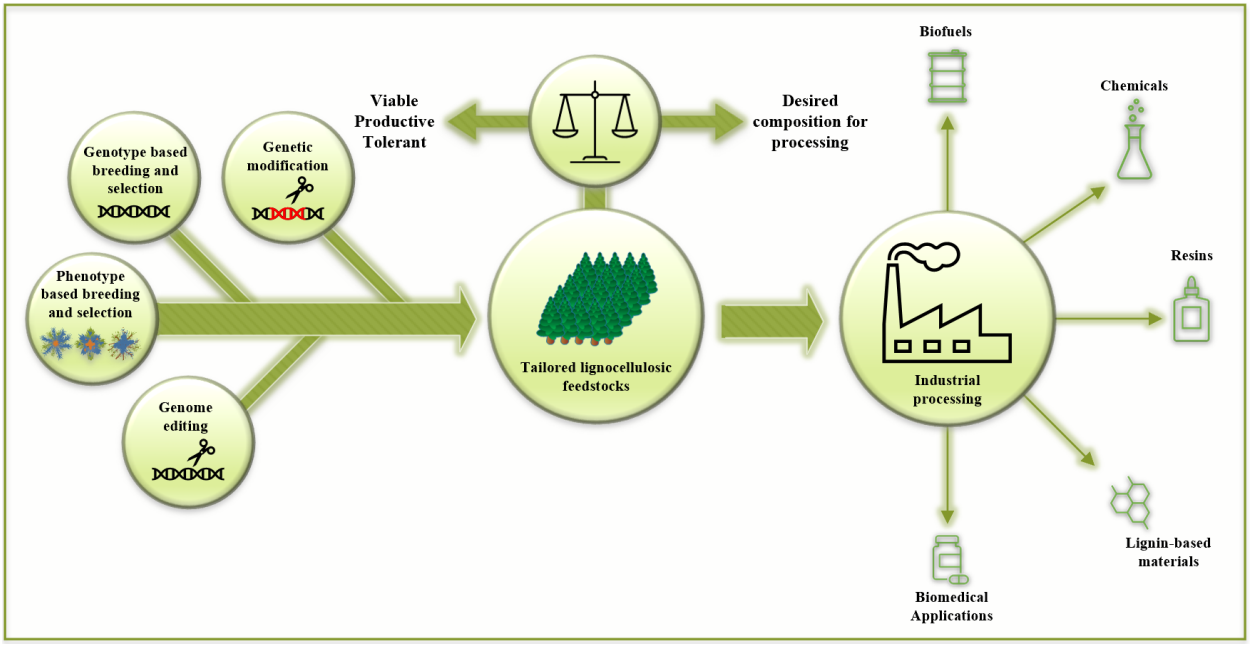


Figure 38. Connecting selective breeding approaches and their value for producing tailored feedstocks for pulp and paper/modern biorefinery processes and some potential applications for biomass-based feedstocks. Figure inspiration from Figure 3 in [433].

Of course, environmental conditions can have major implications for plant-based feedstocks. Drought is one of the most commonly occurring plant stress agents, causing more annual loss in crop yield than all pathogens combined [30]. Plants have developed a variety of complex, variable, and not well understood adaptations in response to drought stress. These adaptations include altering root architecture, leaf rolling, stomata closure, enhanced structural support, and tissue specific lignification. These physical changes are accomplished through complex and intermingling metabolic mechanisms such as gene regulation, transcription factor expression, and hormone and reactive oxygen species (ROS) signaling. Regulation of lignin biosynthesis appears to be associated with (or even a key component of) drought tolerance mechanisms in some plants [52]. Lignin's capacity to provide mechanical strength and pathogen resistance may provide extra protection to vulnerable plants under drought stress, additionally, its hydrophobic properties are thought to reduce the permeability of the xylem, therefore aiding the plant with water and nutrient retention and transport [434] [435]. Further understanding the mechanisms of drought tolerance in lignocellulosic feedstocks can facilitate breeding strategies

that leverage these processes for a variety of objectives. In particular, lignin composition may be a major variable with bidirectional implications for both plant drought tolerance and industrial processing of lignocellulosic biomass.

It is generally understood that lignin accumulation is positively correlated with drought tolerance and many studies covering a variety of plant species have observed this [435] [432]. It must be noted that not all studies find agreement with this correlation, and this may be a consequence of the variety of species, tissues, treatment regimes and study methodologies within the literature [47]. To a lesser extent, lignin *composition* has been evaluated in response to drought stress to determine if there is a correlation between lignin structural features and drought response mechanisms [436] [39] [42]. Moura-Sobczak et al. [39] found variation in S/G ratio in *Eucalyptus globulus* Labill and two hybrids under drought stress with contradictory changes in the apical stem versus the basal regions. dos Santos et al. [42] analysed soluble lignin oligomers in two sugarcane genotypes under water stress and found that the frequency of different oligomers was affected by water deficit. Lima et al. [48] related monomer contents of trees in a tropical semiarid climate to wood traits and embolism and found a positive correlation with S/G ratio and leaf life span. Tu et al. [50] investigated lignin-mediated drought resistance in grapevines using transgenic plants with overexpression of a previously characterized bZIP gene, VlbZIP30, known to enhance osmotic stress resistance during the seedling stages. They found the transgenic plants had increased lignin accumulation, mainly guaiacyl monomer content, compared to wildtype plants, indicating that guaiacyl units may serve a particular function in enhancing drought tolerance as compared to the other monomer units. Zhao et al. [59] isolated the full-length cDNA of a caffeoyl-CoA O-methyltransferase gene (CCoAOMT) from P. *ostii,* an important enzyme for lignin biosynthesis that may play a role in abiotic stress tolerance. They used CCoAOMT overexpressed transgenic tobacco plants to study the drought stress response mechanism associated with this enzyme and found increased lignin content in roots, stem, and leaves compared to wildtype controls. Similar to Tu et al. [50] , the most dramatic change occurred in the content of guaiacyl units, thus, they speculated that the higher guaiacyl lignin was likely associated with drought stress tolerance. Wen et al. [61] isolated a transcriptional factor gene, MsWRKY11, generally considered to be stress-inducible, and investigated its potential mechanism in drought tolerance in alfalfa using transgenic overexpressed MsWRKY11 or dominantly repressed MsWRKY11 plants. They found that the

MsWRKY11 overexpressed plants had enhanced water efficiency and drought tolerance, with increased lignin contents. In contrast with Tu et al. [50] and Zhao et al. [59], Wen et al. found that syringyl monomer units showed the largest increase under drought conditions. While several studies considered lignin composition as it relates to drought response, most limited their scope to the determination of the relative proportion of lignin monomer units and did not consider more detailed structural features such as inter-unit linkages or functional groups, which may be key indicators for industrial processing. Additionally, to the best of our knowledge there are no previous studies that investigated changes lignin content and structure of Douglas-fir in response to drought stress conditions.

Douglas-fir is one of the fastest growing conifers and one of the most frequently planted species in North America with a very broad latitudinal range [437] [438]. Additionally, it is reported to have one of the highest lignin contents in conifer species, which is sometimes credited with its medium to high height and good drought tolerance [439]. Softwoods have higher lignin content than hardwoods [440], and conveniently, structural differences between softwood lignins are reported to be minor as compared to different hardwood lignin's [441]. Therefore, findings regarding altered lignin biosynthesis in response to drought stress for Douglas-fir may also provide a basis for future investigations of other softwood species. Herein, we sought to assess the impact of drought-induced abiotic stress on the composition of lignin of two years old Douglas-fir full siblings grown in drought and control conditions for one year.

The convoluted nature of the plant cell wall components accompanied with the innate complexity and heterogenous composition of lignin make the analysis of lignin challenging. Lignin analysis also must be understood contextually based on two distinct classes, native and technical lignin. This is because in situ lignin, which is lignin in its native state as it resides inside the plant cell wall, can be very different from technical lignin, which is lignin that has been extracted from the plant cell wall. A range of compositional changes can occur in the lignin structure during the extraction processes that depend on the mechanism and severity of the extraction process. Resulting from the challenges of analysing lignin in situ, features of technical lignin are often used to make inferences regarding the native lignin structure. If the extraction process used to isolate the corresponding technical lignin from the plant cell wall is mild enough and the chemical mechanism by which the lignin is released from the plant cell wall is taken into consideration, this

can be a practical approach. A variety of analytical approaches for the analysis of native and technical lignin are available, however, there is no single technique that can provide the complete structural picture of lignin. Thus, a range of techniques are employed to target different structural features or for different levels of resolution. Analytical characterization of lignin includes wet chemical approaches, chromatography, spectrometry and spectroscopy. Wet chemical analysis can be effective for quantifying important structural features but are often too specific, indirect, tedious, time-consuming, degradative, and requiring of larger sample amounts and/or hazardous and environmentally harmful chemicals. Thus, higher through-put techniques such as spectroscopy are becoming more popular. More recently, high throughput techniques have been combined with chemometrics – the field of applying mathematical and statistical tools to multivariate problems in chemistry – to develop models to support the rapid prediction of biomass composition [183]. Here we used a variety of analytical approaches to characterize the wood samples, focusing on lignins content and structure. Relative abundance of cellulose, hemicellulose, and lignin in the wood were determined gravimetrically using a detergent fibre method. Lignin was extracted using a mild modified organic solvent method and further analysed for compositional features. Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy (ATR FT-IR) was used for qualitative analysis of the wood and lignin samples. Pyrolysis Gas Chromatography-Mass Spectrometry (Py-GC/MS) of wood and lignin samples was used to detect compositional differences between samples. Quantitative ^{31}P Nuclear Magnetic Resonance spectroscopy (^{31}P NMR), a hybrid wet chemical and spectroscopy approach, was used for the quantitative determination of different lignin hydroxyl functional groups. Semi-quantitative Solid-state Cross Polarization/Magic Angle Spinning ^{13}C Nuclear Magnetic Resonance spectroscopy (multiCP NMR) was used to estimate both lignin content in wood as well as quantify other structural features of the lignin.

6.2 Methods

6.2.1 Plant materials, growth conditions, and treatments

Douglas-fir seed [*Pseudotsuga menziesii* var. *menziesii* (Mirb.) Franco] was germinated in January 2018. The seed originated from full-sibling families from the BC Ministry of Forests Douglas-fir breeding program. The seed was grown for one year in peat, vermiculite, and pearlite

at 4:1:1 mixture in 112-cell polystyrene foam blocks in a greenhouse with conditions between 0 and 30ºC, and watered with an overhead sprinkler as needed. In May 2019, 24-cell polystyrene foam blocks were set up to assess the amount of water needed and explore directions for wood and lignin analysis. The blocks were placed in a polycarbonate shadehouse with plastic roll up sides, and temperature was kept between 0 and 45°C. The study consisted of a random selection of trees taken from all the families in two water treatments of 30%, and 10% soil moisture. The polystyrene blocks had repeated watering triggered at the target weight corresponding to their 30%, and 10% soil moisture level. The target weight for each block was determined by the weights of dry soil and polystyrene block, plus the weight of the water to reach to 10% and 30% soil moisture targets. When the target weight was reached, water was pumped into each cell using a bottletop dispenser at 75 ml for the control trees, and 50 ml for drought trees. Five trees from both drought and control groups were randomly selected and sampled when the trees became dormant that year. The wood was sampled carefully shaving wood with a scalpel from the outer ring only (2019), and then the wood samples were combined within their respective treatment group to produce enough sample for further analysis. The stem sampling process is shown in Figure 39. Wood from a naturally regenerated western redcedar (*Thuja plicata* Donn) approximately 70 years old, sampled with a 12 mm increment borer at 1.3 meters height, was also used along with the Douglas-fir control and drought samples.

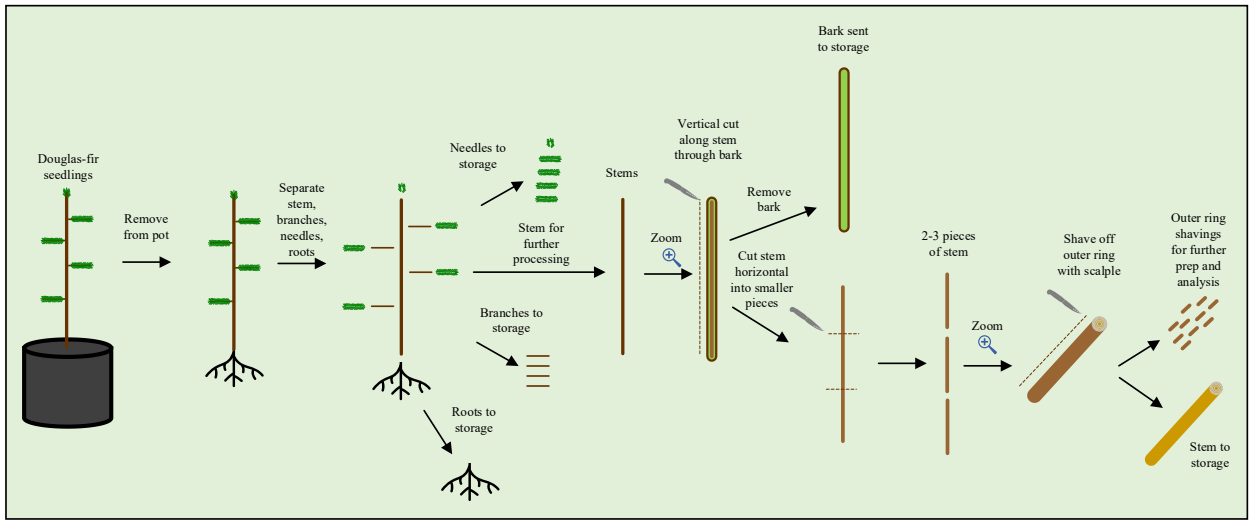


Figure 39. Schematic of the procedure for sampling the outer-rings of Douglas-fir seedling stems.

6.2.2 Determination of cell wall contents

The detergent fibre method is a sequential neutral detergent fibre (NDF) treatment, acid detergent fibre (ADF) treatment, and acid hydrolysis treatment with ashing in a muffle furnace of wood samples that facilitates the gravimetric determination of cellulose, hemicellulose, and lignin. The method was developed via the works of Goering and Van Soest and of which the full details can be found elsewhere [106] [442] [107]. The procedure used in this work are as follows. 0.5 g of milled wood dried to 70℃ for each sample was added to 100 ml glass tubes along with boiling stones. 100.0 ml of neutral detergent solution and 50 µl of α-amylase to the tubes and vortexed. The samples were brought to a boil and allowed to reflux gently for 60 minutes at 95-100℃ on a large block heater. After 30 minutes, another 50 µl of α-amylase was added to the tube, vortexed, and then returned to the block heater for another 30 minutes of boiling. Vacuum filtration was then used to filter the residue from the tubes into crucibles, the boiling stone and all plant material was rinsed from the tube with warm water. The residue was then rinsed with 100 ml of hot tap water, and then 100 ml of Milli-q water. The crucibles with the residue were then dried at 70℃ overnight and weighed. The samples were then transferred quantitatively from the crucible to 50 ml glass tubes using 50 ml of the acid detergent solution, ensuring to scrape all sample residue from the crucible. The samples were then brought to a slow boil to reflux gently for 60 minutes at 95-100℃ on a small block heater (tubes were vortexed at the 30-minute mark). Vacuum filtration was then used to filter the residue from the tubes into crucibles exactly as performed after the first reflux including the rinsing steps, overnight drying at 70℃ and weighing. The crucibles containing the samples were then placed in a disposable beaker and 5 ml of 72% H_2SO_4 was added to the crucible and allowed to drain for approximately 1 hour. This step was performed three times. After the third acid addition drained from the crucible, the crucible was then rinsed with 100 ml of hot tap water, and then 100 ml of Milli-q water. Once rinsed the crucible was dried at 70℃ overnight and weighed. Finally, the samples were ashed at 500℃ for 3 hours in a muffle furnace, allowed to cool in a desiccator, and weighed. Relative weight percent of cellulose, hemicellulose, and lignin were then calculated by difference gravimetrically.

6.2.3 Modified organosolv lignin extraction

Lignin was extracted from the wood samples using a modified organosolv extraction technique under reflux conditions using the techniques described by da Rosa et al. [443] with the objective of obtaining a relatively less altered lignin structure isolated from the plant cell wall. A flowchart for the extraction procedure can be found in Figure 40. Around 5 grams of milled wood sample were weighed and added to a glass beaker followed by a scoopful of Henger boiling chips and 100 mL of hexanes. The beaker was placed in a Velp SER158 Solvent Autoextractor and a non-polar extractables method was run. The sample was refluxed in the hexane for five hours to remove non-polar extractables from the wood. After refluxing with the hexanes and cooling, the sample was vacuum filtered using a #4 filter paper on a Buchner funnel and flask. The filtrate was discarded, and the wood retained for further processing. The dried wood and boiling chips were then transferred to a 500 mL round bottom flask and 200 mL of a 75% ethanol / 3% sulphuric acid solution was added to the flask. The wood is then refluxed in the acid-catalysed solvent solution for 24 hours. After 24 hours, the solution was allowed to cool, and the sample was filtered using vacuum filtration using a #4 filter paper on a Buchner funnel and flask as done previously. A clear dark red lignin rich filtrate solution was retained, and the carbohydrate residue was discarded. The ethanol was then allowed to evaporate using a roto-evaporator until the solution is around one-third its original volume. Chilled Milli-Q water was then added to the solution until it turned cloudy with a pinkish-red suspension (around 200-300 mL). The suspension was then filtered using vacuum filtration with a #42 filter paper to catch all of the solids. The accumulated solid was then washed with Milli-Q water to produce a dark burgundy-pink wet solid with a strong odor. The sample paper was then dried on the filter paper in the oven at 70°C for a minimum if an hour. The dry solid light burgundy-pink solid was then gently scraped off the filter paper onto wax paper where it was weighed prior to storage.

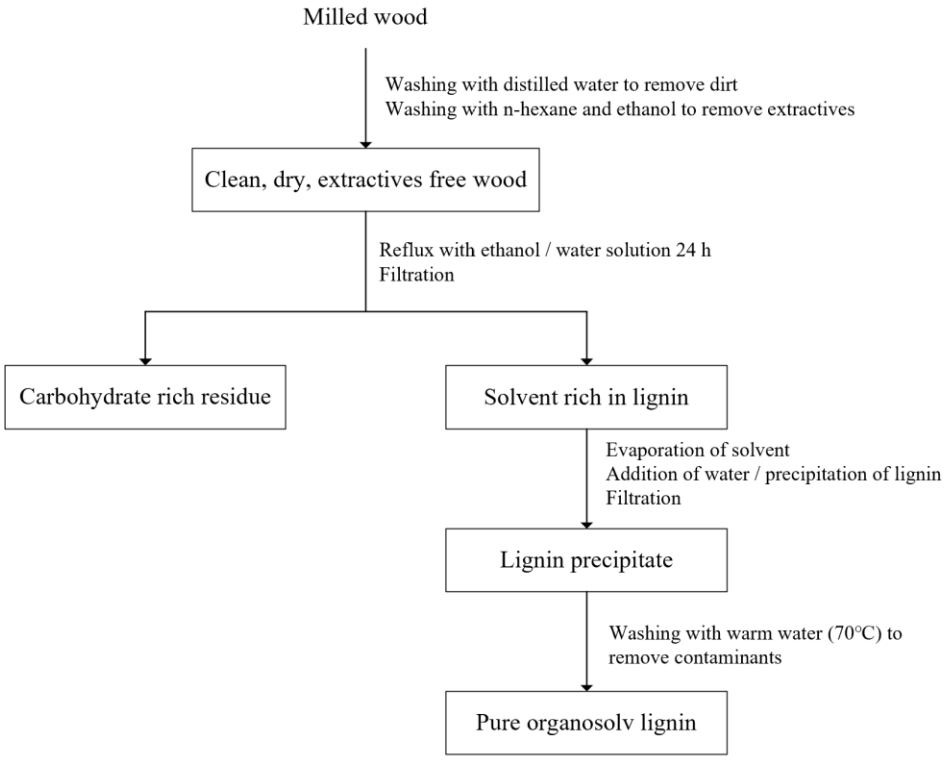


Figure 40. Flowchart for the organic solvent extraction of lignin from wood samples.

6.2.4 Fourier Transform – Infrared Spectroscopy

Attenuated Total Reflectance Fourier Transformed Infrared Spectroscopy (ATR FT-IR) spectra of lignin samples were taken using a Nicolet 6700 spectrometer in triplicates with 32 scans in the mid-infrared range of 4000-650 cm^{-1} at a resolution of 4 cm^{-1}. Samples were dried to 70℃ for 24 hours prior to spectra collection. Background spectrums were collected prior to each sample and subsequently subtracted from the next spectrum recorded. To scan samples, a small amount of sample was placed on the top plate, ensuring full coverage over the crystal. A pressure arm was used to apply uniform force to the samples to ensure better contact with the crystal, the pressure was kept consistent between samples.

6.2.5 Pyrolysis Gas Chromatography – Mass Spectrometry

Py-GC/MS data was obtained with ACEM model 9300 system attached to a GC-MS system. Briefly, 1.5-2 mg of sample (milled wood/organosolv lignin) were inserted into the pyrolysis tubes covered on either end with wool and inserted into the pyrobe. The pyrolysis and GC-MS conditions are reported in Table 3 and Table 4, respectively. Both wood and lignin samples

were run at a final temperature of 600 and 800°C and in duplicates. A blank run was performed prior to the first run and after each sample type (cedar/drought/control).

Table 3. Pyrolysis conditions.

Interface		Pyroprobe		Trap	
Rest temperature	50°C	Initial temperature	200°C	Rest temperature	50°C
Initial temperature	100°C	Initial time	1 s	Desorption temperature	290°C
Initial time	1 min	Ramp rate	20°C /ms	Desorption time	2 min
Ramp rate	0°C /min	Final temperature	600/800°C	Reactant gas	off
Final temperature	200°C	Final time	15 s		
Final time	1 min				
Transfer line temperature	290°C				
Valve oven temperature	290°C				

Table 4. GC-MS conditions.

Equipment/Parameter	Setting
Column	DB-1701 (60m x 0.250mm x 0.25um)
Column flow	He (1.1 mL/min)
Split ratio	100:1
Inlet temperature	300°C
Oven	45°C (1 min) - 290°C (5 min) @ 7°C/min
Electron impact ionization	70 eV
Scanning mass range	40-700 m/z
Transfer line temperature	305°C
Ion source temperature	230°C
Quadrupole temperature	150°C

6.2.6 Quantitative 31P Nuclear Magnetic Resonance Spectroscopy

The ^{31}P NMR spectra of all lignin samples were acquired in duplicates according to the procedures of Meng et al. [145]. Briefly, the procedure involves the solubilization of the substrate in an appropriate solvent, the preparation of an internal standard that does not overlap with the

signals of the lignin, the phosphitylation of hydroxyl groups via the addition of TMDP, followed by the NMR measurement of the sample and some subsequent data processing steps.

6.2.7 Solid-state Cross Polarization Magic Angle Spinning Nuclear Magnetic Resonance Spectroscopy

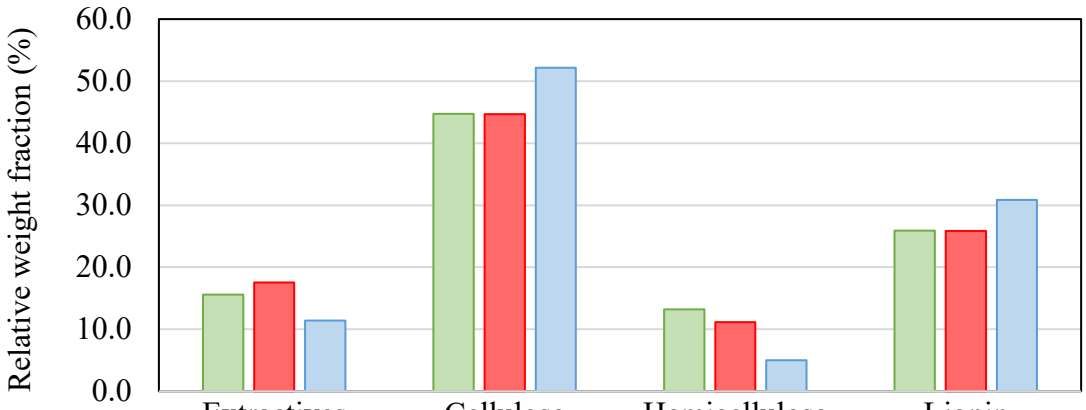
Solid-state ^{13}C NMR spectra [444] [445] were collected on a 9.39 T (n_L (^1H) = 400.1 MHz, n_L (^{13}C) = 100.6 MHz) Bruker AVANCE HD spectrometer equipped with a 4 mm double resonance (H-X) magic-angle spinning (MAS) probe. The $^{13}C\{^1H\}$ multiple-cross polarization (mulitCP) experiments were performed using a 4.13 μs π/2 on ^1H with a ramped Hartman-Hahn match on ^{13}C with TPPM ^1H decoupling ($gB_1/2p$=61 kHz). Each sample was acquired using a 5 s recycle delay, 1024 co-added transients and a contact time of 1 ms. The number of CP loops used in the $^{13}C\{^1H\}$ multiCP experiments were set to a value of 8 loops with a 1 s repolarization period. The π-pulse (echo) performed at the end of the ^{13}C sequence had a pulse length of 8 μs. Samples were placed into 4 mm o.d. zirconia rotors and sealed using Kel-F® drive caps. All ^{13}C NMR spectra were referenced to the high frequency peak of adamantane at 38.56 ppm and a MAS frequency of 12 kHz was employed throughout.

6.3 Results

6.3.1 Wet chemical analysis

The detailed results of the detergent fibre methods can be found in Appendix C. The simplified results of the detergent fibre method are shown in Figure 41. Lignin content determined using the detergent fibre method were consistent with known lignin content from literature (~24 – 29% for Douglas-fir and ~29 – 33% for western red cedar) [446]. The detergent fibre method for determination of wood samples revealed no observable difference in lignin content in drought sample (26%) as compared to control sample (26%), indicating that drought stress may not significantly alter acid-insoluble lignin content in Douglas-fir wood. The lignin and cellulose contents were the most unchanged of the different components evaluated, while the largest difference was that of the extractives and hemicellulose. Extractives are a solubility class encompassing a variety of chemical compounds; thus, extractives content will vary depending on the solvent used. Extractives include nonpolar compounds (e.g., fatty acids and lipids) and polar compounds (sugars and phenolics) and make up 5-25% of Douglas-fir tissues [107] [447].

The results here indicate that at least within the resolution of the detergent fibre method, there are no significant differences observed in the lignin content of the Douglas-fir samples. It should be noted that the detergent fibre method does not account for a minor fraction called the acid-soluble lignin which is dissolved with cellulose in the acid detergent fibre solution. Therefore, it is plausible that other methods could reveal lignin content differences or that the composition of lignin rather than the content of the lignin may be altered in response to drought stress. To explore this prospect further, the wood samples were analysed with quantitative solid-state ^{13}C NMR to estimate lignin content and lignin has been extracted from the wood samples for compositional analysis, the details of which are discussed below.



	Extractives	Cellulose	Hemicellulose	Lignin
Control	15.6	44.7	13.2	25.9
Drought	17.6	44.7	11.1	25.9
Cedar	11.4	52.2	5.0	30.8

Figure 41. Relative content of extractives, cellulose, hemicellulose, and lignin in the wood samples as determined using the detergent fibre method. Not shown here is the ash components (~0.1 – 1%).

6.3.2 Organosolv lignin extraction

The modified organosolv lignin extraction technique was selected as a low severity method for recovering technical lignin more representative of its native structure in the plant cell wall. Yields from the organosolv extraction can be found in Appendix C. The Douglas-fir wood samples recovered yields of only 14.3% and 18.8% relative recovery of the detergent fibre lignin in control and drought, respectively. This is equivalent to 3.7% and 4.9% recovery by total weight of the

wood sample. While the yield of the extraction was quite low as compared to the best yields of da Rosa et al., this is not unexpected as the method was originally designed for rice husk which has major compositional differences compared to softwoods. Additionally, the low yield of lignin may be the result of the relatively mild nature of the technique which may not cleave the more resistant bonds lignin shares within the plant cell wall. While a difference in detergent fibre determined lignin content was not observed between the drought and control samples, there was a difference in relative recovery of lignin from organosolv extraction. The drought samples yielded 4.6% by weight more of the lignin as compared to the control samples. This was well outside of the range of experimental error observed in the cedar test samples (0.9%) and may indicate that structural differences were responsible for the difference in lignin yield between the samples. Thus, structural analysis of the extracted lignin could provide more insights in this observation.

6.3.3 Qualitative Fourier Transform – Infrared spectra of lignin

FT-IR spectra of all wood and lignin samples can be found in the Appendix C. The FT-IR spectra of the wood samples matched the expected features of softwood from literature. The spectra of Douglas-fir samples matched up very closely and the cedar had only minor differences, particularly in intensity rather than peak position. However, the FT-IR spectra of wood was not well resolved, likely due to the compositional complexity of the untreated wood samples.

FT-IR spectroscopy of the lignin samples in the spectral range of 2000 – 650 cm^{-1} are shown in Figure 42. The peaks observed in the spectra reported here matched up with the expected peaks reported in the literature for guaiacyl type lignin and are assigned in Table 5. Consistent with the FT-IR spectra of the wood samples, the lignin spectrums for all three samples were quite similar and lined up very closely throughout the entire spectral range. The cedar test sample exhibited some minor deviation from the Douglas-fir samples around 1650 – 1850 cm^{-1} which corresponds to the conjugated and unconjugated $C = O$ stretching vibrations of ketones, carbonyls, and ester groups. The similarity between the Douglas-fir lignin spectra indicated no dramatic structural differences between the control and drought lignin were observable with the FT-IR technique. Slight variations in intensity of certain peaks here could indicate differences in abundance of certain structures; however, the differences observed are very minor and could be the product of unrelated factors.

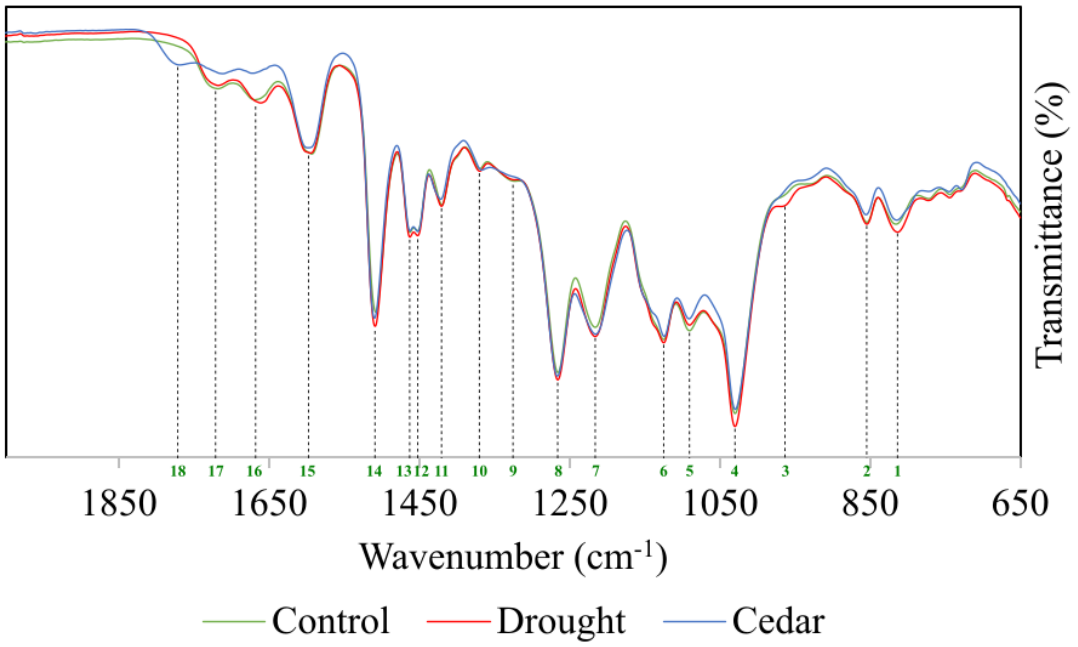


Figure 42. ATR FT-IR of lignin samples and band assignments for the peaks with the largest differences.

Table 5. FT-IR band assignments for lignin spectra

Band ID	Band, cm^{-1}	Vibration	Assignment	Ref.
1	812-817	δ_{op} *Ar* C–H	Positions 2, 5, and 6 of G units	[141]
2	853-856	δ_{op} *Ar* C–H	Positions 2, 5, and 6 of G units	[448]
3	966-990	δ_{op} –HC=CH–	Out of plane deformation (trans)	[449]
4	1029-1030	δ_{ip} Ar C–H, st C–O(H) & st C–O(C), st C=O	Positions 2, 5, and 6 of G units, aliphatic primary alcohols and ether; C=O (unconj.)	[141] [448] [450]
5	1089-1091	st C–O	C–O deformation in secondary alcohols and aliphatic ethers	[451]
6	1123-1125	δ_{ip} *Ar* $C-H$	Positions 2, 5, and 6 of G units	[448]
7	1214-1216	st C–C, st C–O(H), st C–O(Ar), st C=O	Phenolic alcohol and ether, G condensed > G etherified	[141] [448] [450]
8	1265-1266	st C–O, st C=O	G unit aromatic methoxy, G ring breathing C=O	[141] [448] [450]
9	1326	st C–O, st C=O	S unit, G ring condensed (sub. position 5)	[141] [448] [450]
10	1366-1370	δ_{ip} O–H, st C–H	Phenolic OH, aliphatic CH_3 (not methoxy)	[141] [448]
11	1419-1422	st C–C, δ_{ip} C–H	Aromatic skeletal vibrations, asym. C–H deform. in –OCH_3	[141] [448] [450] [451]
12	1451-1453	$\delta_{Asymetric}$ C–H	C–H deformations; asym. in –CH_3 and –CH_2–	[141] [448] [450] [451]
13	1462-1463	$\delta_{Asymetric}$ C–H	C–H deformations; asym. in –CH_3 and –CH_2–	[141] [448] [450] [451]
14	1507-1510	st C–C	C–C aromatic skeletal vibrations; G > S	[141] [448] [450] [451]
15	1592-1599	st C–C, st C=O	C–C aromatic skeletal vibrations, C=O stretch; S > G; G condensed > G etherified	[141] [448] [450] [451]
16	1659-1673	st C=O	Conjugated C=O	[448]
17	1710-1718	st C=O	C=O stretch in unconjugated ketones, carbonyls and ester groups	[141] [448]
18	1767-1770	st C=O	Conjugated C=O ketones	[451]

st: Stretching vibration, δ_{ip}: In-plane deformation vibration, δ_{op}: Out-of-plane deformation vibration.

6.3.4 Pyrolysis Gas – Chromatography / Mass Spectrometry of wood and lignin

Py-GC/MS was performed on both the untreated wood and isolated lignin samples to investigate compositional differences between samples. Py-GC/MS uses pyrolysis to thermally decompose samples at high temperatures in the absence of oxygen, gas chromatography to separate the volatile compounds in time, and a mass spectrometer that ionizes the compounds into fragments that can be used to identify the molecular formula parent compound. The pyrolysis decomposition mechanism of materials such as wood and lignin produce complex chromatograms with a wide variety of molecular product species. Thus, total ion current (TIC) chromatograms

(Figure 43), which plot the summation of signal intensities for all mass to charge ratios at a single retention time, can be populated with hundreds of peaks making manual interpretation and analysis extremely tedious and error prone.

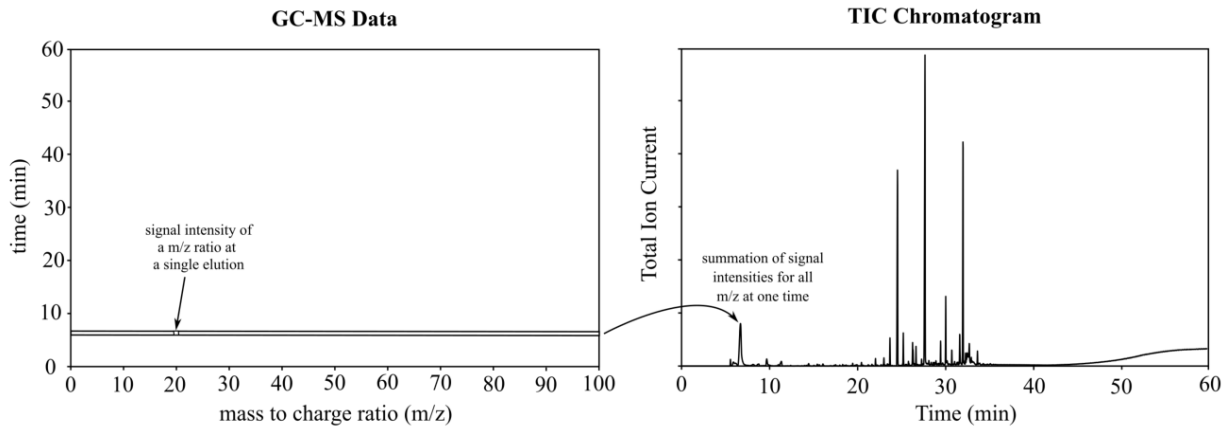


Figure 43. Schematic illustration of GC-MS data in matrix format (left) and a resulting TIC chromatogram (right). Each matrix entry of the GC-MS data contains the measured signal intensity for a specific m/z ratio at the recorded detection time. The TIC chromatogram is constructed via the summation of all signal intensities of all m/z for each time point.

Figure 44 shows an example of the TIC chromatograms for the lignin samples with a variety of compounds identified, each associated with a TIC peak demonstrating the complexity of the analysis process. Thus, automated approaches and multivariate data analysis have been proposed for rapid and more reliable analysis of GC-MS data. In particular, the PARAFAC2 based Deconvolution and Identification System (PARADISe) for processing raw GC-MS data is a promising approach that has been demonstrated to be useful for lignocellulosic biomass samples [378] [376]. The PARADISe software uses a chemometric multi-way modelling approach to resolve peaks, identify and quantify compounds from raw complex GC-MS data of many samples simultaneously. We sought to use the PARADISe software to automate the GC-MS analysis process of the wood and lignin samples in this investigation; however, we encountered a variety of technical challenges with the PARADISe software which is still in its development stages. Due to the technical challenges related to this approach and the time constraints of this project we are not yet able to report meaningful results from the Py-GC/MS analysis of the wood and lignin samples in this thesis. We intend on continuing with this approach either through the continued

application of the PARADISe software or the exploration of alternative GC-MS processing software's. The background subtracted GC-MS of the wood and lignin samples are shown in Figure 45 and Figure 46, respectively.

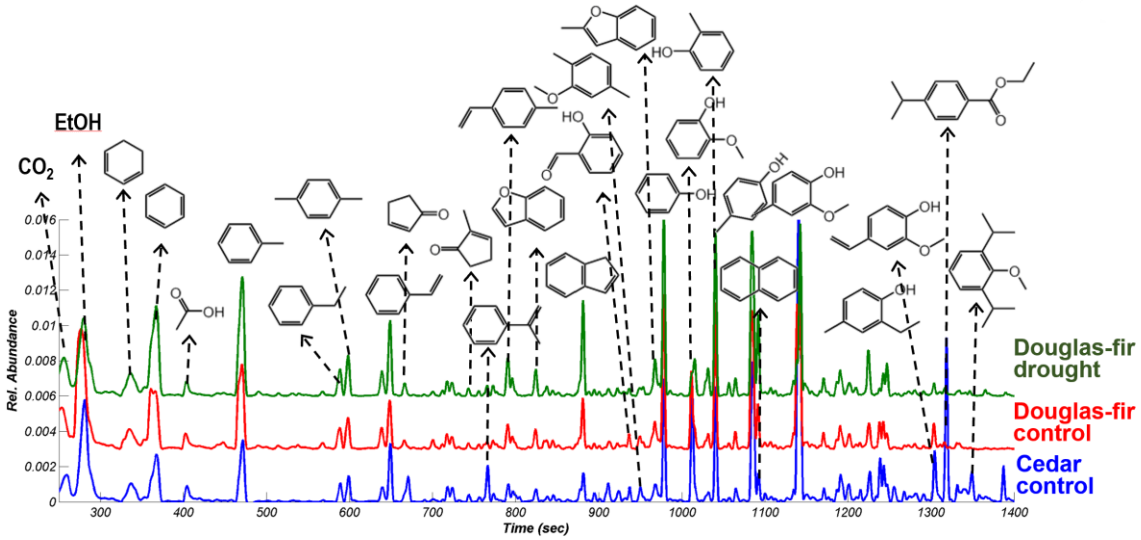


Figure 44. Example TIC chromatogram from Py-GC/MS of Douglas-fir lignin samples and the cedar control. Molecular compounds corresponding to different peaks are illustrated.

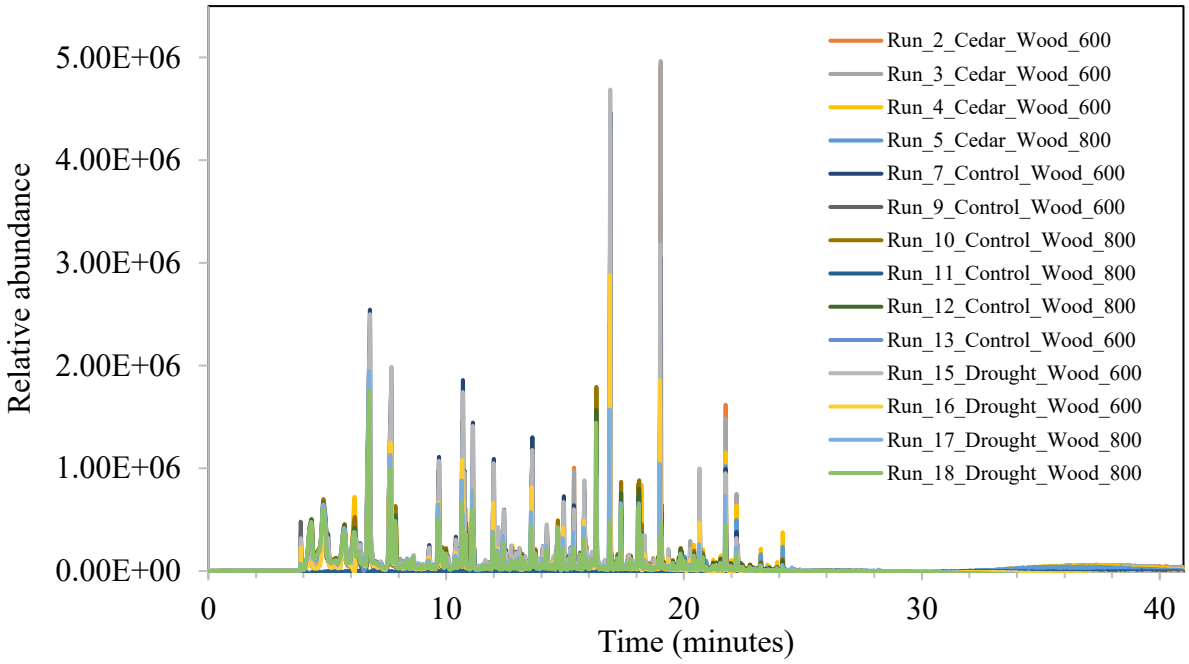


Figure 45. The overlayed TIC chromatogram for the Py-GC/MS analysis of Douglas-fir wood.

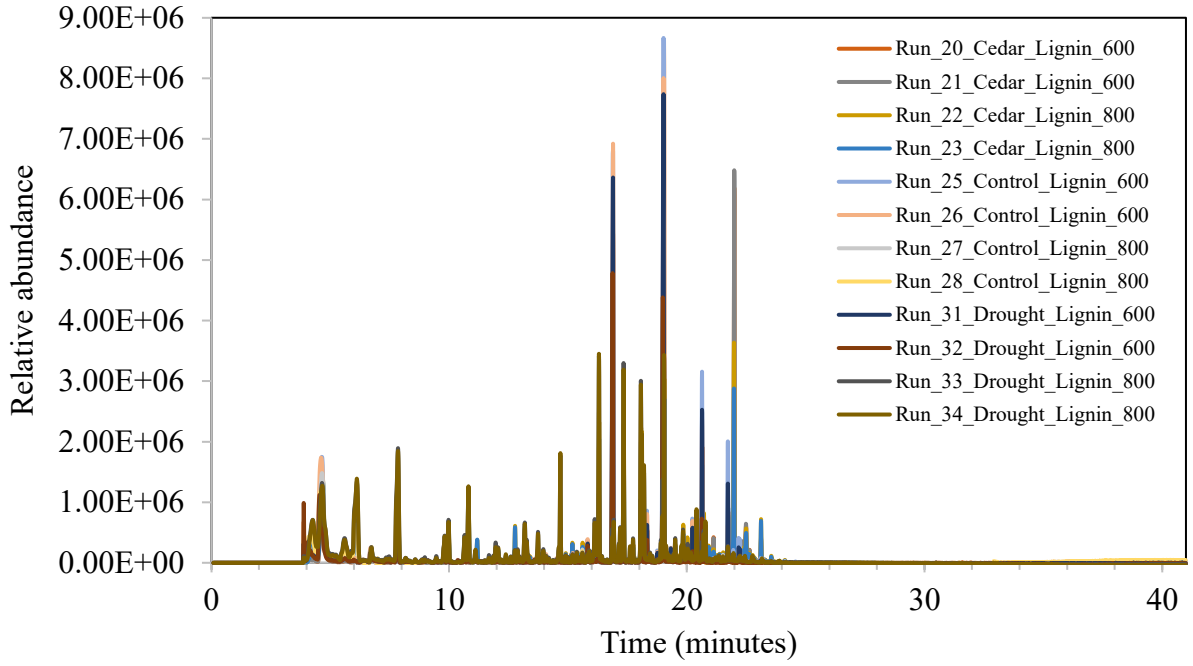


Figure 46. The overlayed TIC chromatogram for the Py-GC/MS analysis of Douglas-fir lignin.

In future work the Py-GC/MS data will be analysed to identify and quantify various structural compounds with a focus on identifying differences between the control and drought samples, potentially supporting the findings of the other analytical approaches. Work has already been underway on a new sample set to further the investigation regarding drought and lignin in Douglas-fir. Over one-hundred trees from ten families in this new drought study have been sampled and are undergoing analysis. Most of these new samples have very limited sample quantities (~1 g). The large sample set, and very small sample masses make Py-GC/MS an important and effective analytical tool because of its minimal sample requirement, rapid analysis time and information-rich data.

6.3.5 Hydroxyl group content of lignin

Figure 47 shows the ^{31}P NMR spectrum along with the corresponding assignments based on Table 6. The ^{31}P NMR spectra were consistent with that of literature with the peaks matching that of the expected functional groups for lignin [145]. As anticipated for softwood, the largest peaks corresponded to the aliphatic ($\sim 145.4 - 150.0$ ppm) and guaiacyl ($\sim 139.0 - 140.2$ ppm) functionalities. There was also a small sharp peak corresponding to syringyl monomer units

(~142.7 ppm) and a small broad peak that corresponded with *p*-hydroxyphenyl monomer units (~137.8 ppm) which indicated that small amounts of these two monomer units may also be present in the extracted lignin. In addition, there were some smaller peaks corresponding to lignin interunit linkages such as $\beta - 5$ (~143.5 ppm), $4 - O - 5$ (~142.3 ppm), $5 - 5'$ (~141.2 ppm), and carboxylic functional groups ($\sim 133.6 - 136.0$ ppm). It is often recommended to combine the syringyl and different condensed 5-substituted units in the range $\sim 140.0 - 144.5$ ppm into the C_5-substituted phenolic OH due to signal overlap in this region, however, since softwoods are composed predominantly of guaiacyl units, there is not as much overlap in this region and the syringyl and 5-substituted units may be quantified.

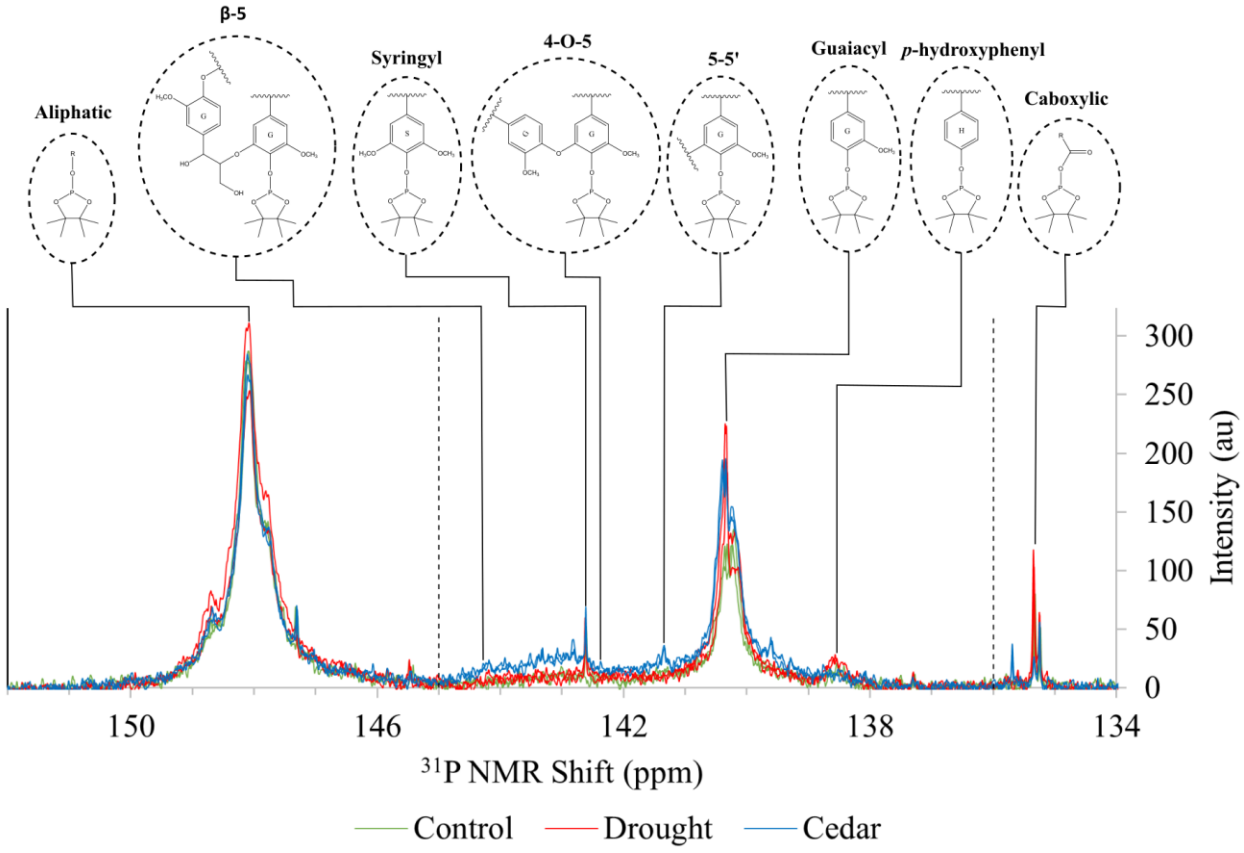


Figure 47. ^{31}P NMR of lignin samples along with the corresponding structures contributing to the peaks.

Table 6. ^{31}P Chemical shifts and corresponding functional groups [145].

Lignin functional group	Chemical shift (p.p.m)
Aliphatic OH	~145.4-150.0
Phenolic OH	~137.6-144.0
C_5 substituted	~140.0-144.5
5-5	~141.2
4-O-5	~142.3
Syringyl	~142.7
β-5	~143.5
Guaiacyl	~139.0-140.2
p-hydroxyphenyl	~137.8
Carboxylic acid OH	~133.6-136.0

Figure 48 shows the quantitative results of the ^{31}P NMR analysis of lignin samples organized by the three main unique hydroxyl groups: aliphatic, phenolic, and carboxylic. The results of the analysis are comparable in magnitude to those reported in the literature for softwood organosolv lignin [145]. Overall, the aliphatic hydroxyl groups were found to have the highest concentration in all samples, with phenolic hydroxyl groups having the next highest concentration, and carboxylic hydroxyl groups having the lowest concentration. Additionally, there were no large differences in concentrations of respective hydroxyl groups, with the one distinct exception being the cedar test sample which had a significantly higher concentration of phenolic hydroxyl groups compared to the Douglas-fir samples.

When comparing the Douglas-fir samples, the results revealed a decrease in aliphatic hydroxyl groups and a comparable increase in phenolic hydroxyl groups in the drought treated sample as compared to the control. For the aliphatic hydroxyl groups, the difference of 0.13 mmol/g between of the drought and control samples of Douglas-fir is comparable to that between Douglas-fir drought and cedar (0.12 mmol/g), which are different species. Moreover, the phenolic hydroxyl group content of the drought sample is larger by 0.15 mmol/g than that of the control. However, this difference due to drought is much smaller than the difference of 0.85 mmol/g between Douglas-fir control and cedar wood. The carboxylic hydroxyl groups remained at considerably lower concentration as compared to the other group types and did not vary much between the different samples. Lignin hydrophobicity, which is considered a major driving force of enzymatic hydrolysis-inhibitory lignin-enzyme interactions, generally increases with aromatic and decreases with aliphatic hydroxyl group content [168]. These results suggest drought stress

may alter the functionality of Douglas-fir lignin, increasing its hydrophobicity which could inhibit cellulose hydrolyzability. These results may suggest structural changes have occurred in the wood samples resulting from the drought treatment leading to the enhancement of lignin phenolic hydroxyl groups over aliphatic.

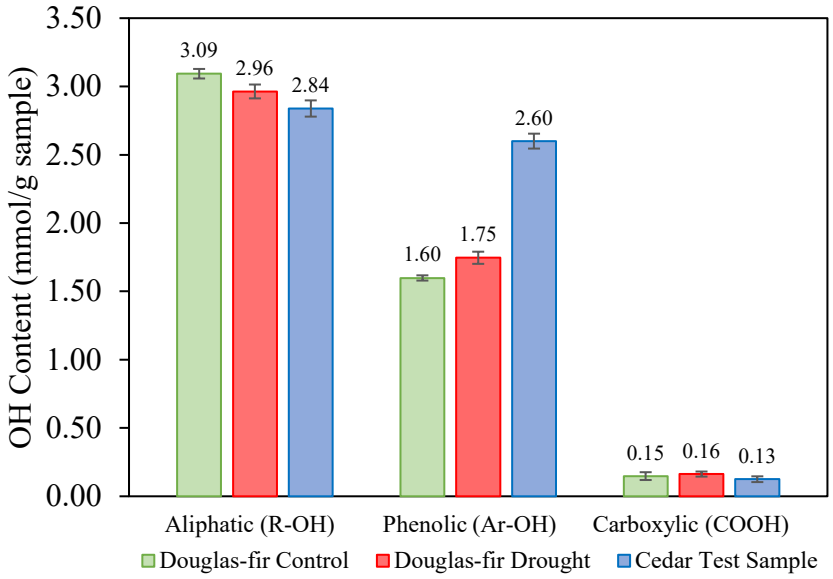


Figure 48. Quantitative results of the ^{31}P NMR analysis of lignin samples organized by the three main different hydroxyl groups, aliphatic, phenolic, and carboxylic.

6.3.6 Characterization of wood and lignin using solid-state ^{13}C Nuclear Magnetic Resonance Spectroscopy

Compared to other methods NMR spectroscopy is an important tool that allows for high-resolution, comprehensive, and quantitative structural characterization of organic material. There are a variety of different NMR techniques for this purpose including multidimensional approaches (1D, 2D, 3D…). The NMR techniques are often most generically categorized as either solution-state or solid-state. Both solution-state and solid-state NMR have been used to study wood and lignin each offering certain advantages and limitations. Solution-state NMR provides high resolution but requires the solubilization of the samples under investigation, thus, limiting the investigations to soluble compounds and restricted sample concentrations, in addition to increasing undesired sample handling, introducing solvent effects, and rendering the sample invalid for future analysis [452].

In contrast, solid-state NMR is a non-destructive technique that can provide detailed compositional and structural information on the components of lignocellulosic materials in situ. Thus, solid-state NMR is a valuable tool for the analysis of lignocellulosic biomass and has been used extensively for this purpose. Applications of solid-state NMR for the analysis of lignocellulosic biomass and its derivatives have lagged compared to solution-state techniques in part due to the belief that solid-state NMR cannot provide high enough resolution for detailed structural analysis [453]. Techniques such as solid-state ^{13}C NMR indeed suffer from broad and convoluted spectra making the peak assignments challenging. Yet, the development of new techniques and advancements in spectral editing approaches have supported more detailed investigations of lignocellulosic biomass [452]. Spectral editing techniques facilitate the selective acquisition of NMR signals, reducing spectral complexity and enabling the assignment of otherwise obscured peaks. In addition to resolving qualitative information from solid-state NMR spectra, producing quantitative spectra is also desired. For quantitative solid-state NMR spectra the ^{13}C direct polarization / magic angle spinning (DP/MAS) is favoured over the more routine ^{13}C cross polarization / magic angle spinning (CP/MAS) but is stifled due to significantly longer acquisition times [452]. More recently, the multiCP approach has been developed for high-throughput quantitative solid-state MAS ^{13}C NMR spectra of organic material [444]. This new method allows for the collection of quantitative spectra with good signal-to-noise ratio at a significantly reduced analysis time compared to other approaches. However, quantitative analysis of NMR spectra is performed by integrating spectral peaks which correspond to distinct chemical structures to determine their relative abundance. The chemical complexity of lignocellulosic material can convolute NMR spectrums; thus, assignment of peaks and determination of their integration boundaries is challenging and often arbitrary. As a result, multiCP spectra of lignocellulosic material may be better described as semi-quantitative.

6.3.7.1 MultiCP NMR of wood

Figure 49 provides an example of the ^{13}C NMR spectrum of wood with the regions of interest delineated. Figure 50 shows the ^{13}C NMR of wood samples from this study. Table 7 shows the chemical shifts and corresponding assignments for ^{13}C NMR of wood from the literature. There are two main chemical shift regions associated with lignin in the NMR spectrum that are distinct from carbon atoms present in carbohydrates. These are the aromatic (~160 – 100 ppm) and

methoxy (~56 ppm) regions. The aromatic region is mostly distinct from the other plant cell wall components and may provide a direct correlation to the number of aromatic rings, and therefore also potentially correlate with the lignin content in the material. However, the region is full of broad, poorly resolved peaks and spectral overlap does occur at the upfield end of the aromatic region (120 – 90 ppm) between alkyl O-C-O and aromatic carbons [454]. Further upfield, the methoxy signals (~56 ppm) are stronger and sharper. Despite the methoxy peak sharing some of its left side with the C-6 carbohydrate atoms, it has been proposed that the area of the methoxy peak can be more accurately estimated by integrating the right side of the peak and multiplying by two [455]. While the methoxy peak is more resolved than the aromatic region, it corresponds to the methoxy groups on the phenyl ring which vary between the different lignin monomers. Thus, there may not be a direct correlation between the methoxy peak and the number of aromatic rings in the material. This adds uncertainty regarding the use of the methoxy peak for quantifying lignin, especially for lignocellulosics with more heterogenous lignin monomer content [455]. Previously, both the aromatic and the methoxy regions have been used to estimate lignin content in organic material. For quantitative solid-state NMR studies of lignin content, often a calibration curve is developed using lignin mixtures of different concentrations [456] [457] [339] [455].

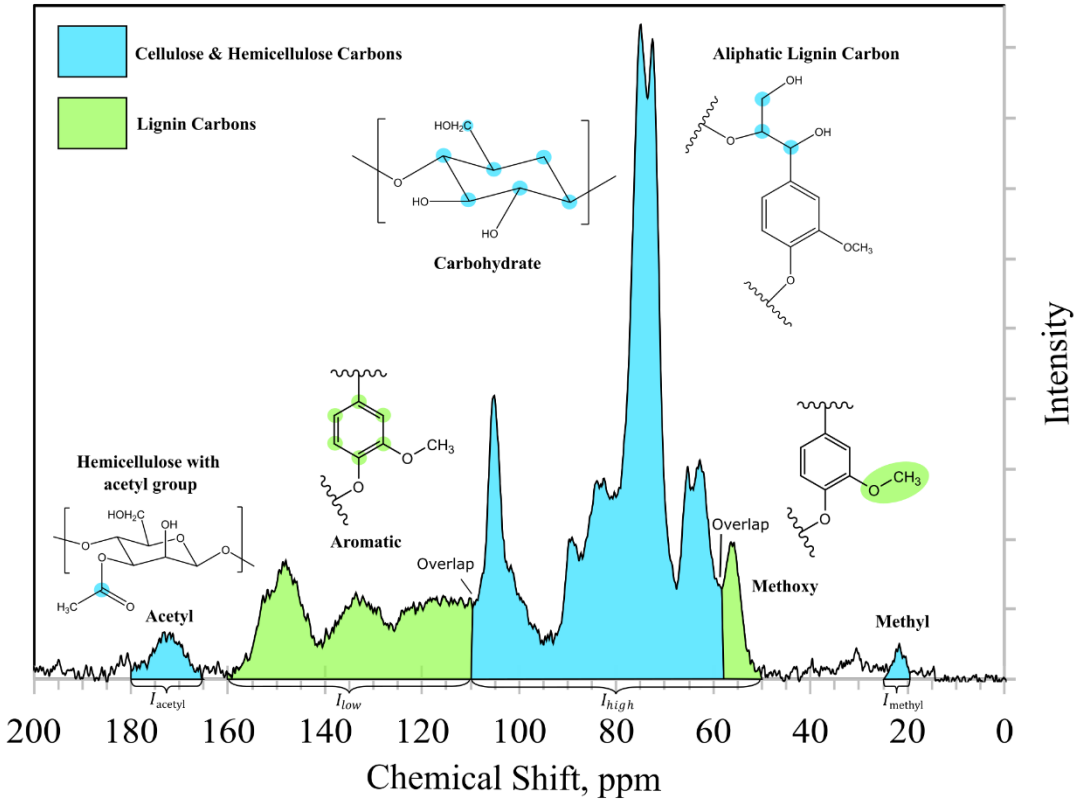


Figure 49. An illustrative example of the solid-state ^{13}C NMR spectra of wood with carbohydrate and lignin regions distinguished. The green and blue fill indicate regions predominantly due lignin and carbohydrate signals, respectively. Key points of spectral overlap between the aromatic and carbohydrate regions are indicated. Relevant examples of different wood compounds and their carbons that contribute to the spectrum are overlayed on the figure. Regions labeled as I_{acetyl}, I_{low}, I_{high}, and I_{methyl} correspond to the integration regions used in the methods described below. Note that the chemical structures within wood are complex and the compounds shown in this figure are examples of reduced complexity used to indicate the types of carbons that are contribute to the spectrum.

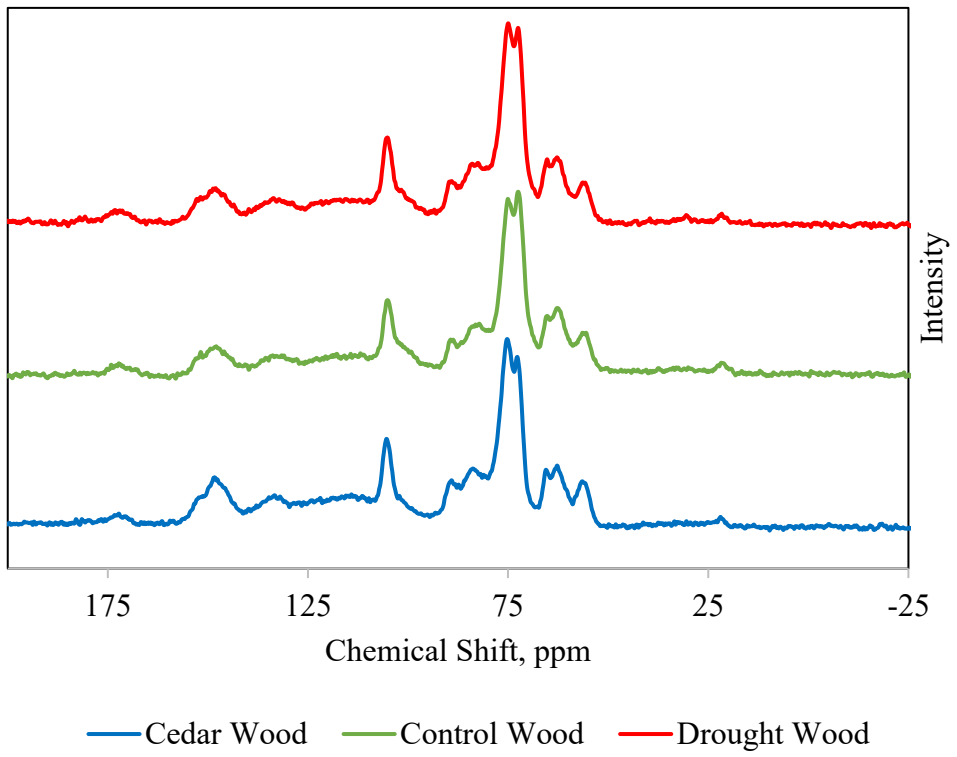


Figure 50. ^{13}C solid-state NMR of the wood samples, where drought and control indicate the Douglas-fir samples and Cedar is the test sample.

Table 7. Chemical shift and corresponding assignments of ^{13}C NMR wood spectrum

Peak ID	^{13}C Chemical Shift, ppm	Assignment	Reference
1	21.6	CH_3 in acetyl groups of hemicelluloses	[458] [459]
2	56.2	Aryl methoxy carbons of lignin	[458] [459]
3	60-64	$OC_\gamma H_2$ carbons of lignin	[458] [459]
4	63	CH_2OH of carbohydrate (amorphous C_6 of cellulose), C_6 carbon of hemicelluloses	[458] [459]
5	65.3	CH_2OH of carbohydrate (crystalline C_6 of cellulose)	[458] [459]
6	72.5	$CHOH$ of carbohydrates (C_2, C_3, C_5 of cellulose)	[458] [459]
7	72-76	$OC_\alpha H_2$ carbons of lignin	[458] [459]
8	75	$CHOH$ of carbohydrates (C_2, C_3, C_5 of cellulose and hemicelluloses)	[458] [459]
9	82-83	$OC_\beta H_2$ of lignin	[458] [459]
10	84	$CHOH$ of carbohydrates (amorphous C_4 of cellulose and hemicelluloses)	[458] [459]
11	87	$CHOH$ of carbohydrates (crystalline C_4 of cellulose)	[458]
12	89	$CHOH$ of carbohydrates (crystalline C_4 of cellulose)	[459]
13	97-102	Shoulder of C_1 carbon of hemicellulose	[458]
14	105	$OCHO$ of carbohydrates (C_1 of cellulose)	[458] [459]
15	112	Aromatic $C-H$ (G_2 of lignin)	[459]
16	114.5	Aromatic $C-H$ (G_5 of lignin)	[459]
17	118	Aromatic $C-H$ (G_6 of lignin)	[459]
18	131.5	Aromatic $C-C$ ($G_{1(ne)}$ of lignin)	[459]
19	134	Aromatic $C-C$ ($G_{1(e)}$ of lignin)	[459]
20	143-147	Aromatic $C-O$ ($G_{4(e)}$ and $G_{4(ne)}$ of lignin)	[459]
21	147-153	Aromatic $C-O$ ($G_{3(e)}$ and $G_{3(ne)}$ of lignin)	[459]
22	173	Acetyl ester $C=O$ carbon hemicellulose	[460]
Clusters			
1	19-24	CH_3 in acetyl groups of hemicelluloses	
2	54-57	Aryl methoxy carbons of lignin (some hemicellulose)	
3	57-110	Carbohydrates (Cellulose, hemicellulose, aliphatic lignin)	
4	102-160	Aromatic carbon (lignin)	
5	167-176	Acetyl ester $C=O$ carbon hemicellulose (Other carboxyl groups)	

A procedure for determining the lignin contents of wood and pulps directly from integrated NMR signal intensities has been described previously [461] [462]. The procedure was originally developed for softwood and uses the integrated area of the low shielding aromatic region (160 – 109 ppm) and the high shielding aliphatic region (109 – 50 ppm) along with some assumptions regarding the chemical formula lignin and carbohydrates to estimate wt% lignin of the sample from the NMR spectra. For this procedure, the following assumptions are required as described by Haw et al. [461]. Wood is composed primarily of cellulose, hemicellulose, and lignin. Softwood lignin is mainly derived from coniferyl alcohol ($C_9H_9O_2(OCH_3)$) monolignol precursors and hence the lignin polymer inside the plant cell wall is mainly composed of guaiacyl monomer units with

approximately 9.92 carbons, 3 oxygens and 10-12 hydrogens. The molecular formula for conifer lignin is $C_9H_{7.15}O_2(H_2O)_{0.4} - (OCH_3)_{0.92}$ ($M_w = 183\ g/mol$) as previously reported [463]. Carbohydrate repeat units have an average empirical formula of $C_6H_{10}O_5$ ($M_w = 162\ g/mol$). Acetate carbonyl and methyl carbon signals are neglected in the above assumption. With these assumptions established one can relate the low shielding integrated area, $I_{160-109}$, which corresponds to the aromatic carbons of lignin to the total integrated ^{13}C intensities due to all lignin carbons, I_{lig}.

$$I_{lig} = \left(\frac{9.92}{6}\right) I_{160-109}$$

Where 9.92 is the average number of carbons in a guaiacyl lignin repeat unit and 6 is the number of carbons in a guaiacyl lignin repeat unit that contribute to the aromatic low shielding region. Likewise, the total integrated ^{13}C intensities due to all carbohydrate carbons, I_{carbs}, can be related to the high shielding integrated area, I_{109-50}, which corresponds to the carbohydrate components as well as the aliphatic ($C-\alpha$, $C-\beta$, $C-\gamma$) and methoxy lignin carbons.

$$I_{carbs} = I_{109-50} - \left(\frac{3.92}{6}\right) I_{160-109}$$

Where 3.92 and 6 are the number of aliphatic and aromatic carbons in a guaiacyl lignin repeat unit. Normalizing the total integrated ^{13}C intensities

$$I'_{lig} = \frac{I_{lig}}{I_{carbs} + I_{lig}}$$

$$I'_{carbs} = \frac{I_{carbs}}{I_{carbs} + I_{lig}}$$

we get the normalized signal intensity due to lignin and carbohydrates as I'_{lig} and I'_{carbs}, respectively. Finally, the dry-weight percent lignin for softwood can be estimated using the following equation.

$$wt\%\ lignin = \frac{\left(\frac{183}{9.92}\right) I'_{lig}}{\left[\left(\frac{183}{9.92}\right) I'_{lig} + \left(\frac{162}{6}\right) I'_{carbs}\right]}$$

Where the coefficients on the normalized signal intensities are the molecular weight divided the number of carbon atoms in the respective average molecular weight formulas of the repeat units of lignin and carbohydrates.

Haw et al. [461] noted that while acetate carbonyl and methyl carbon signals are neglected in their method, they could be accounted for by altering the method to include contribution from esterified sugars in the average formula for the carbohydrate fraction. Here we expand on their approach and include the contribution from these spectral regions in an attempt to reduce the uncertainty in the lignin content estimate. For this we make a few more assumptions: the regions of 24 – 19 ppm and 177 – 165 ppm are purely due to methyl carbon in acetyl groups of hemicellulose and acetyl ester $C = O$ carbon hemicellulose, respectively. We assume softwood hemicellulose as a D-mannopyranose with an O-acetyl group, $M_w = 206\ g/mol$ with 8 carbons [464] [465]. Thus, we define $I_{cellulose}$ and $I_{hemicellulose}$ as the total integrated ^{13}C intensities due to the cellulose and hemicellulose carbons, respectively. Here we can roughly estimate a fraction of the intensity in the region 109 – 50 ppm is due to hemicelluloses with acetyl groups. We can estimate this fraction using the spectral intensity of the acetyl and methyl carbons.

$$I_{hemicellulose} = \left(\frac{8}{2}\right)[I_{177-165} + I_{24-19}]$$

$$I_{cellulose} = I_{109-50} - \left(\frac{3.92}{6}\right)I_{160-109} - \left(\frac{6}{2}\right)[I_{177-165} + I_{24-19}]$$

Just as done previously, we normalize to the total integrated ^{13}C intensities

$$I'_{cellulose} = \frac{I_{cellulose}}{I_{cellulose} + I_{hemicellulose} + I_{lig}}$$

$$I'_{hemicellulose} = \frac{I_{hemicellulose}}{I_{cellulose} + I_{hemicellulose} + I_{lig}}$$

$$I'_{lig} = \frac{I_{lig}}{I_{cellulose} + I_{hemicellulose} + I_{lig}}$$

We can then estimate the lignin content of the samples in weight percent from the ^{13}C spectra of wood as follows.

$$wt\%\ lignin = \frac{\left(\frac{183}{9.92}\right) I'_{lig}}{\left[\left(\frac{162}{6}\right) I'_{cellulose} + \left(\frac{206}{8}\right) I'_{hemicellulose} + \left(\frac{183}{9.92}\right) I'_{lig}\right]}$$

We will term this new modified approach for determining lignin content as the 'extended direct integration' approach so as to differentiate it from the Haw et al. [461] approach.

Table 8 shows the lignin content estimates from the ^{13}C NMR spectra of wood using the Haw et al. [461], the extended direct integration approach, and the wet chemical detergent fibre method. We predicted that the NMR integration methods would predict high lignin content than the wet chemical approach because it can account for the entire lignin content of the samples including the so-called acid-soluble lignin that is unaccounted for in the detergent fibre method. In general, this has been the case as almost all lignin content values were predicted higher using direct integration with the exception of the extended direct integration method applied to the control sample which was slightly lower than the wet chemical determined lignin content (25%). As expected, the extended direct integration approach predicted lower lignin content for all samples compared to the less intricate Haw et al. [461] approach, reducing lignin content estimates by around 1%. Hence, it appears that the extended direct integration approach proposed here may provide slightly improved prediction compared to the Haw et al. [461] approach.

Table 8. Lignin content dry weight estimates from ^{13}C NMR spectrum of wood and comparison to wet chemical determination of lignin content.

	Lignin Content (Dry weight percent)		
Sample	Wet chemical Detergent fibre [106]	Direct integration Haw et al. [461]	Direct integration Extended direct integration
Control	26	26	25
Drought	26	31	30
Cedar	31	38	37

There may be a variety of reasons for the disparity between the detergent fibre and direct integration lignin content prediction. If we focus on the Douglas-fir samples, we can see that the predictions for the control were much closer than that of the drought sample. Direct integration for the control sample predicted ~25 – 26% lignin content which is consistent with the detergent fibre

method (26%). Direct integration for the drought sample predicted ~30 – 31% lignin which is in conflict with the 26% lignin content determined using the detergent fibre method. It is not clear if there is a source of error in the direct integration method which could cause a much higher prediction in lignin content for the drought sample, or if there truly is a greater lignin content than is determined using the detergent fibre method. We know from the detergent fibre method that the drought wood sample had around 2% more extractives content and around 2% less hemicellulose content. Aromatic extractive compounds would not be distinguished from the lignin signals during the integration of the ^{13}C NMR spectra and could be responsible for higher lignin content predictions in the drought sample. To counter this narrative, it should be noted that cedar wood had 3-7% less extractives content compared to Douglas-fir, and still had significantly higher lignin content predicted from integration of the ^{13}C NMR spectra than the detergent fibre method.

There are three important considerations to point out regarding the direct integration approach applied in this work. First, the Haw et al. [461] approach was originally applied to extractives free wood. We have not applied any prior treatment to the wood samples apart from milling. Based on the detergent fibre method, around 11-18% of the wood samples are extractives which include protein, starch, waxes, polar & non-polar extractables and an additional ~1% is ash. Second, the chemical formulas used for the lignin and carbohydrates are not accurate for all compounds in the samples. The molecular formula for the carbohydrates in the high-shield region is less correct for hemicellulose than cellulose and the molecular formula for lignin is based on the Freudenberg's lignin and slightly different formulas have been proposed for conifer lignin. Finally, the boundaries chosen for the integration regions are somewhat arbitrary and there is considerable overlap between the aromatic carbon and the carbohydrates between 120 and 100 ppm. These points must be understood as we consider the reliability of the approach.

Despite the uncertainty in the quantitative results from the direct integration of the NMR spectra, we can still consider the relative differences between the lignin content estimates. While the detergent fibre method does not reveal changes in lignin content, the direct integration results indicate there may be some increased lignification of stems of Douglas-fir resulting from drought stress. Although potentially some other factors such as extractives content could contribute to this difference, it cannot be confirmed based on the data available. For clarity on this, a larger sample size and potentially more techniques that can further support the results would be required.

6.3.7.2 MultiCP NMR of lignin

Figure 51 provides an example of the multiCP ^{13}C NMR spectrum of lignin with the seven main regions of interest delineated while Figure 52 shows the multiCP ^{13}C NMR spectrum of lignin samples from this study. Table 9 shows the chemical shifts and corresponding assignments for ^{13}C NMR of lignin from the literature. It has been pointed out that even just the extreme stereochemical complexity of the lignin polymer significantly broadens spectra from overlapping NMR signals, let alone its structural complexity and heterogeneity [466] [467]. As evident in Figure 51, the spectrum is not well resolved and there is considerable spectral overlap particularly in the aromatic region. Overall, the general features are as expected for a ^{13}C NMR spectrum of softwood lignin, most notably the three broad peaks in the aromatic region ~162 – 102 ppm and the sharp methoxy peak ~56 ppm. Two subtle but clear indicators provide assurance that the spectra is that of typical of softwood primarily composed of G lignin (as opposed to hardwood composed of G and S lignin), namely the reduced signal intensity around 105 ppm and the position of the aromatic oxygenated carbon peak occurring on the upfield side of 150 ppm (~148 ppm) which is distinct in comparison to hardwood lignin of which the peak appears on the downfield side of 150 ppm (~153 ppm) [468]. The lack of signals from 105 – 90 ppm supports the purity of the lignin sample as signals of C1 carbohydrates would be expected to be evident here if there were high levels of carbohydrate contamination [127]. This is consistent with expectations as organosolv lignins are typically of high purity.

Comparing the three spectra in Figure 52, there are no glaring differences in the NMR spectrum of the lignin samples. Thus, to take a closer look at the changes that might be present in the spectra, we focus seven regions of interest corresponding to certain carbons or categories of carbons to quantify and compare their occurrence. As we can see in Figure 51, the seven main regions of interest, from downfield to upfield, include the aromatic oxygenated ($C-O$), aromatic carbon – carbon ($C-C$), aromatic methine ($C-H$), Alkyl-O-Aryl ($\beta-O-4$, $\alpha-O-4$), hydroxyl ($OH_{primary}$, $OH_{secondary}$, $\gamma-O-Alk$), methoxy (OCH_3), and aliphatic (C_α, C_β). These regions have been integrated to give quantitative estimates of these structural features. The integral region from 162 – 102 ppm was set as the reference based on the assumption that it includes six aromatic carbons and 0.12 vinylic carbons. Based on this assumption, one aromatic ring is equivalent to the integrated value for that region divided by 6.12 [469].

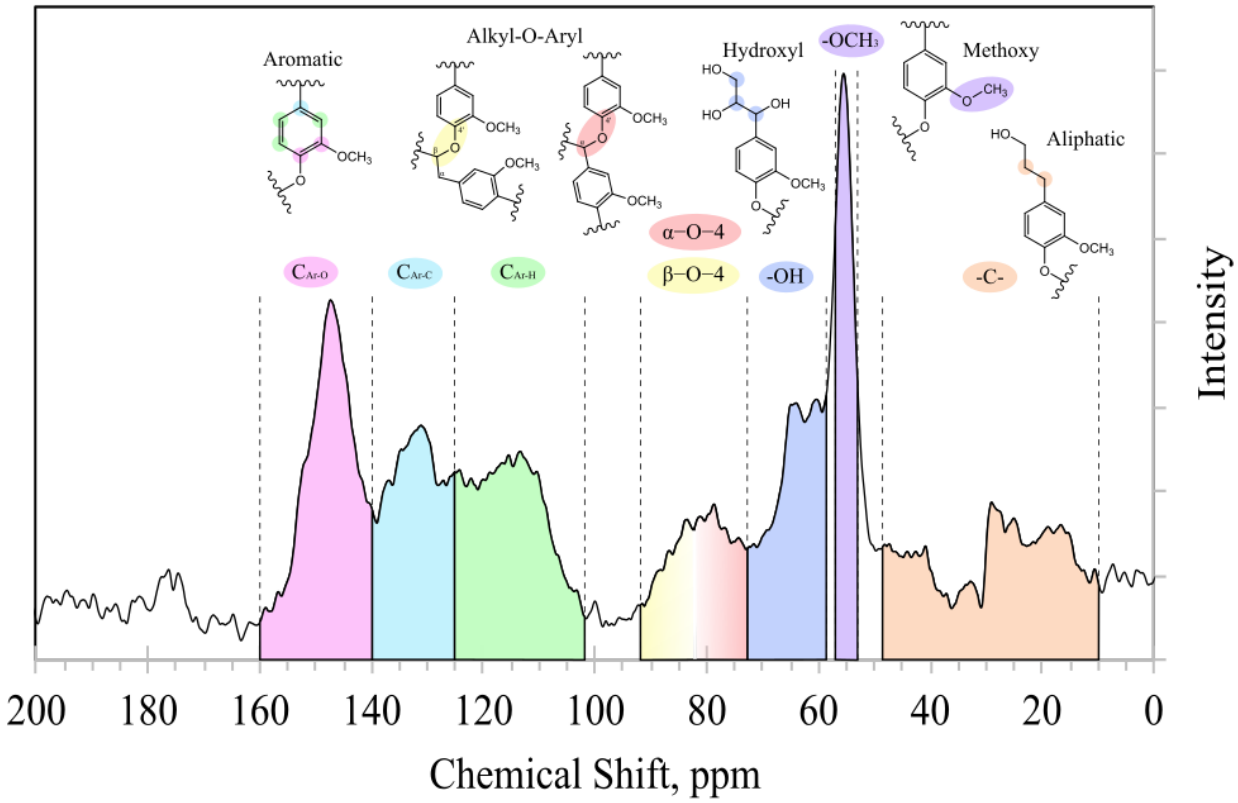


Figure 51. An illustrative example of the ^{13}C solid-state NMR of lignin with seven main regions of interest distinguished, from downfield to upfield: aromatic oxygenated ($C-O$), aromatic carbon – carbon ($C-C$), aromatic methine ($C-H$), Alkyl-O-Aryl ($\beta-O-4,\ \alpha-O-4$), hydroxyl ($OH_{primary},\ OH_{secondary},\ \gamma-O-Alk$), methoxy ($OCH_3$), and aliphatic ($C_\alpha, C_\beta$). The regions filled with colour indicate regions used for integration. Relevant examples of different lignin compounds and their carbons that contribute to the spectrum are overlayed on the figure. Note that the chemical structures within lignin are complex and the compounds shown in this figure are examples of reduced complexity used to indicate the types of carbons that are contribute to the spectrum.

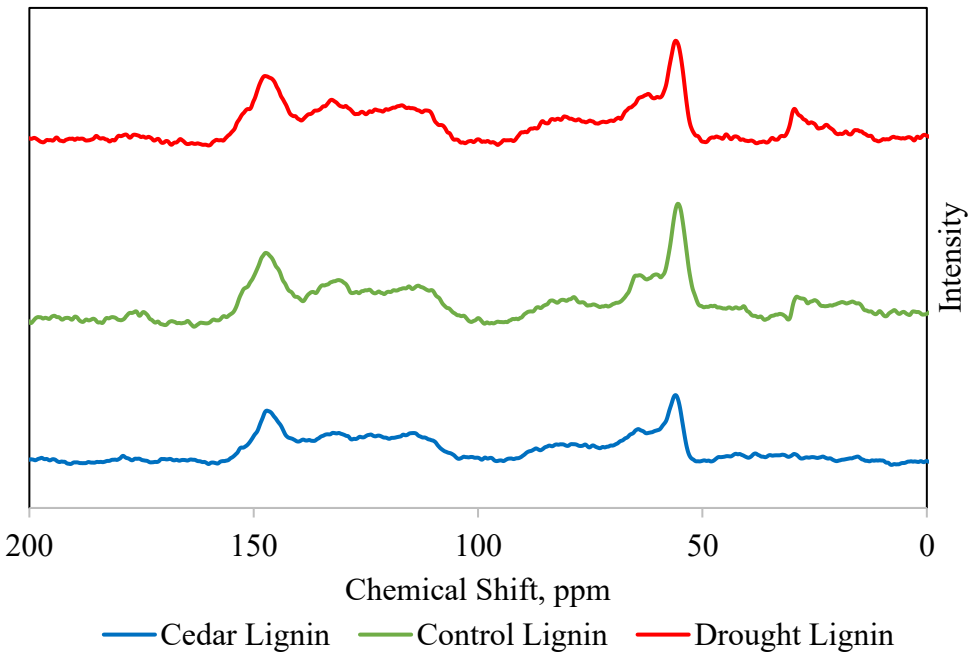


Figure 52. Solid-state multiCP ^{13}C solid-state NMR of the wood samples, where drought and control indicate the Douglas-fir samples and Cedar is the test sample.

Table 9. Chemical shift and corresponding assignments based on of ^{13}C lignin spectrum.

^{13}C Chemical Shift, ppm	Assignment	Reference
13-48	Aliphatic C_α, C_β	[127] [470]
54-57	OCH_3	[127] [470]
58-73	$OH_{primary}$, $OH_{secondary}$, $\gamma - O - Alk$	[127] [470]
73-92	Alkyl-O-Aryl ($\beta - O - 4$, $\alpha - O - 4$)	[127] [470]
102-160	Aromatic carbon	[127] [470]
102-125	Aromatic methine carbon $C - H$ (C_2, C_5, C_6 G units / C_2, C_6 S units / C_2, C_3, C_5, C_6 H units)	[127] [470]
125-142	Aromatic condensed carbon $C - C$ (C_1 $H/G/S$ plus guaiacyl cross linked $5 - 5'$ or $\beta - 5'$)	[127] [470]
142-162	Aromatic oxygenated carbon $C - O$ (C_3, C_4 G units/ C_3, C_4, C_5 S units/ C_4 H units plus diphenyl ether structures)	[127] [470]

Figure 53 shows the results of the integrated NMR spectrums for the three lignin samples over the seven regions of interest previously described including several lignin comparisons from literature. Since the reference value for the aromatic region was set to 6.12 as discussed above, all quantitative values in Figure 53 can be interpreted as per one aromatic ring. Recently, solid-state NMR of different lignin preparations, lignin model compounds, and different polysaccharides was

explored and used to establish a strong linear correlation between alkyl-O-aryl inter-unit bonds and methoxy groups in lignin and the integral of the solid-state ^{13}C NMR spectrum of lignin in the ranges 96 – 68 ppm and 58 – 54 ppm, respectively [471]. These findings provide supporting evidence for the reliability of the alkyl-O-aryl, and methoxy integration regions of the solid-state ^{13}C NMR spectrum of lignin for predicting the content of these structures. It is evident from Figure 53 that the OH and aliphatic regions generally have the greatest disparity with the literature values. These regions may prove difficult to produce reliable information due to several issues which include overlap with carbohydrate impurities, low resolution, baselines, and overlap with the shoulders of the sharp lignin methoxy peak. Thus, although we have also integrated OH and aliphatic regions of the spectrum here, we will mainly consider the aromatic, alkyl-O-aryl, and methoxy regions, which are regarded as more reliable.

Broadly, if we compare the results of our lignin samples to literature values, we see that the different carbon structures quantified are in reasonable agreement with literature reported previously (Figure 53). We expect to see some degree of variation in these structures between studied lignins due to its inherent variability, complexity, isolation procedure, and analytical approaches. However, since they are all softwood lignins, we should see indicators that are consistent with predominantly guaiacyl type lignin. If we assume a lignin structure with no bonding on the two, five, and six position on each phenyl ring, we would expect values of 2, 1, and 3 for the aromatic oxygenated ($C-O$), aromatic condensed ($C-C$), and aromatic methine ($C-H$) carbons, respectively. Of course, lignin has a variety of inter-unit linkages which can occur on these positions. Additionally, although in small amounts, softwood lignin can still also include different lignin monomers such as *p*-hydroxyphenyl or syringyl units. This was evident with signals of S and H units present in the ^{31}P NMR spectrum. Thus, we anticipate deviation from the expected values of the aromatic carbon, which we do observe. In particular, known inter-unit linkages such as ether bonds on the 5 position of the phenyl ring (e.g., $4-O-5$) and condensed bonds (e.g., $5-5'$) would result in greater than 2 aromatic oxygenated carbons and greater than 1 for aromatic condensed per aromatic ring, respectively. The combined effect of these additional bonds on the phenyl ring would necessitate the reduction in aromatic methine carbon to less than 3 per aromatic ring. Our results are consistent with this line of reasoning with the three samples having between 2.07 – 2.38, 1.71 – 1.81, and 2.03 – 2.24 carbons per aromatic ring for the aromatic

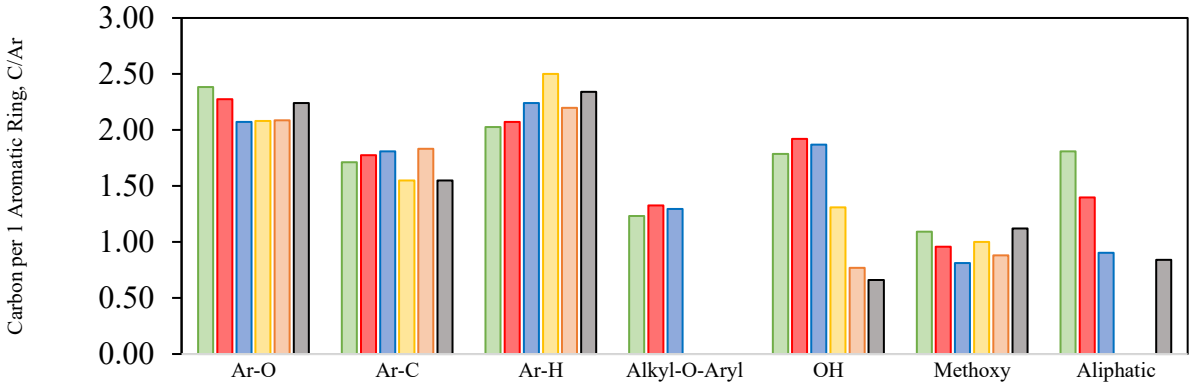
oxygenated, aromatic condensed, and aromatic methine carbons, respectively. These predictions are also observed in the relevant literature (See Figure 53).

In addition to the aromatic carbons, methoxy carbons are also a useful indicator of softwood lignin. Lignin that is predominantly guaiacyl units should have around 1 methoxy group per aromatic carbon and this is corroborated with the literature values in Figure 53. This was roughly the case for the lignin samples in our study which ranged from 0.81 – 1.09 methoxy group per aromatic. We want to point out here that the values for methoxy groups from the solid-state ^{13}C NMR spectrum in this study were very sensitive to the integration regions chosen. Despite being the sharpest peak in the spectrum, shifting the integration region by 1 ppm either upfield or downfield had a substantial impact on the resulting integration value. Given that exact integration regions are considered to some extent arbitrary, the integration region was 'optimized' to more accurately represent methoxy content. In other words, the integration boundaries of the methoxy region were selected to fit the three lignins samples closer to the expected methoxy content of 1 per aromatic ring (57 – 54 ppm). Prior knowledge of the composition of softwood lignin, the ^{31}P NMR spectra, and analytical techniques used in this work support the expected methoxy content of 1 per aromatic ring.

Moving on to the alkyl-O-aryl carbons, which represents the sum of the $\beta - O - 4$ and $\alpha - O - 4$ carbons, we found these to be present at a rate of 1.23 – 1.33 per aromatic ring. Here, there was a lack of comparable literature for comparison. It has been pointed out that this region is susceptible to carbohydrate contaminants which are present in all lignin samples to some degree. Thus, without corrections for these contaminants it should be noted that this rate for alkyl-O-aryl carbons would likely be an overestimate to some degree. It is well known that the $\beta - O - 4$ inter-unit linkage is the most prevalent in the native lignin polymer and thus a key target of valorization strategies. These bonds are also the most readily cleaved during extraction and often creating new phenolic hydroxyl functionalities. Quantification of these units provides practical insights and thus should be a priority for structural studies.

It is quite evident that for most structural features, the difference between the drought and control lignin samples is much less than the difference between the cedar test sample. Thus, this the solid-state ^{13}C NMR analysis does not identify any major structural differences between the

lignin samples. Based on this it does no appear that drought stress has dramatically altered the lignin structure. While we do see some differences in structural features here, the limited sample size does not allow us to make definitive conclusions regarding the significance of the changes.



	Ar-O	Ar-C	Ar-H	Alkyl-O-Aryl	OH	Methoxy	Aliphatic
Control	2.38	1.71	2.03	1.23	1.79	1.09	1.81
Drought	2.27	1.77	2.07	1.33	1.92	0.96	1.40
Cedar	2.07	1.81	2.24	1.30	1.87	0.81	0.90
Capanema et al.	2.08	1.55	2.5		1.31	1	
Nakagame et al.	2.09	1.83	2.20		0.77	0.88	
Holtman et al.	2.24	1.55	2.34		0.66	1.12	0.84

Figure 53. Contents of different carbon structures present in the organosolv lignins. Aromatic C – O (160 – 142 ppm), aromatic C – C (142 – 125 ppm), aromatic C – H (125 – 102 ppm), alkyl – O – aryl (92 – 73 ppm), $\gamma - O - Alk$ and primary and secondary OH (73 – 58 ppm), methoxy (57 – 54 ppm), and aliphatic C_α, C_β (48 – 13 ppm). Comparable lignin from literature includes Capanema et al. [127], Nakagame et al. [472] and Holtman et al. [459]. Note that Nakagame was the average of 3 steam pre-treatment technical lignins.

6.4 Discussion

Overall, this study has investigated the effects of abiotic drought stress on Douglas-fir wood, focusing on lignin content and composition. Wet chemical analysis using the detergent fibre methods demonstrated surprisingly limited differences in wood composition and found no difference in lignin content between control and drought treated samples. Yet, the results of the detergent fibre methods are known to not represent the entire lignin fraction and thus produce incomplete estimates of lignin content. A smaller lignin fraction, typically termed the 'acid-soluble lignin' which is not accounted for in the detergent fibre estimate, could reflect lignin content

differences between samples that was not observed here. Since the detergent fibre method could not account for the entire lignin fraction, we proposed that solid-state ^{13}C NMR of the wood samples may provide a more complete estimate of lignin content. Direct integration of the NMR spectra predicted roughly 5% higher lignin content in drought (30 – 31%) compared to control (25 – 26%) wood.

In addition to assessing lignin content of the wood samples, we were also interested in the lignin composition. We isolated less harshly altered lignin from the wood samples using a mild organosolv extraction technique for further analysis. The approach produced low yield, but high purity lignin as later demonstrated by the solid-state NMR spectra. Drought treated samples yielded slightly higher lignin than control suggesting compositional differences such as more accessible or more cleavable bonds could account for the difference in yield.

While we have encountered hurdles in our approach, we have identified Py-GC/MS as a very useful high-throughput, information rich technique that is particularly suitable for chemometric approaches due to its large and complex data sets. Its ability to rapidly analyse many samples with just micrograms of sample will be especially important for our future investigation where we examine a much larger sample set with very limited sample mass. Work with this technique will be continued for future investigations but has not been successful in this current work.

The ^{31}P NMR spectroscopy technique has enabled the quantitative determination of distinct hydroxyl groups in lignin samples. Differences in lignin hydroxyl group content were observed, most notably a 0.13 mmol/g decrease and a 0.15 mmol/g increase in aliphatic and phenolic hydroxyl groups, respectively, in the drought lignin compared to the control. Hydroxyl groups are very important functionalities for predicting biomass processing efficiency and defining the reactivity of lignin for future processing. Our results suggest drought stress alters the lignin structure producing a more hydrophobic lignin which is considered to promote lignin-enzyme interactions which inhibits the enzymatic hydrolysis of biomass. It remains to be determined if the observed difference is significant in a larger sample set.

Solid-state ^{13}C NMR of the lignin samples was used to quantify detailed structural features of the lignin samples. Distinct aromatic carbons and methoxy groups are readily observed in the

NMR spectrum of lignin sample. Alkyl – O – aryl bonds, primary and secondary hydroxyl groups, and aliphatic carbons are also detected, although much less resolved. The features observed in the lignin NMR spectrums were consistent with softwood lignin and the absence of carbohydrate signals from 105 – 90 ppm supports the high purity of the lignin sample. Quantification of aromatic and methoxy carbon showed reasonable agreement with literature values however, the differences in the three aromatic carbons between the drought and control samples was minor compared to the difference between this study and the literature values. In addition, methoxy content was higher in the control than drought but was very sensitive to integration boundaries, thus, we are skeptical regarding the accuracy of the predicted methoxy content. Taken together, detailed structural analysis via solid-state ^{13}C NMR of the lignin samples has not been able to identify substantial structural differences between the drought and control trees. Slight differences are apparent but cannot be confirmed as significant due to the small sample set. More samples would be required to confidently comment on the significance of these differences.

6.5 Conclusions

The results of this study suggest that drought conditions may increase lignin content and alter lignin composition in the wood of Douglas-fir seedlings. Lignin biosynthesis is suggested to have high dynamic plasticity and increases in lignin content have been frequently associated with drought stress responses. Our results agree with this concept but requires more evidence to confidently settle on this finding. In addition, a mechanistic investigation on the function of lignin content in drought stress tolerance would provide more support to this notion. To support our investigation, we have expanded on a previous approach [461] to predict lignin content directly from the integration of solid-state NMR spectra of wood samples by accounting for more structures in the spectrum. The new approach adds minimal complexity to the calculations but produces a slight improvement in results. However, a more detailed and focused investigation would be required to confirm the accuracy of the approach. Despite this, the results appear promising, and the approach could be further refined in a variety of ways such as removing extractives prior to NMR analysis. In addition to changes in lignin content, differences in lignin composition were observed with a detailed investigation of lignins structure using quantitative ^{31}P NMR spectroscopy and semi-quantitative ^{13}C NMR spectroscopy. These differences, in particular differences in reactive hydroxyl groups are important indicators for processing objectives.

It is well established that drought stress can cause major compositional changes to lignocellulosic feedstocks. Lignin content and composition has major implications for lignocellulosic biomass processing as almost all current industrial approaches require an intensive fractionation/pretreatment step to remove lignin from the other components. This study has identified evidence for an increase in lignin content and altered lignin composition between drought stressed and the control Douglas-fir seedlings. A more detailed investigation is required to further explore the extent of these changes. Future work is already underway with over 100 new samples from 10 Douglas-fir families from another drought experiment. The results of this study will guide future work enabling a more focused investigation. In addition, we intend on applying chemometric approaches for extracting more information from the larger data set in our next investigation.

Chapter 7: Summary, key findings, limitations and future work

7.1 Summary

The objective of this thesis was to assess the impact of drought stress on the composition of Douglas-fir lignin in the context of its implications for feedstock reliability and value for biorefining strategies. Due to the broad and interdisciplinary scope of this investigation, much of the writing in this thesis touches on different sub-fields that support the work that has been undertaken. This chapter summarizes the work in this thesis, presents key findings, limitations, and concludes with a discussion regarding future work.

Chapter 1 provided a broad introduction to the field of lignocellulosic biomass research, defining its significance, the scope of work surrounding the field, and its components. It established the potential for lignocellulosic biomass as a renewable source of energy, materials, and chemicals and introduced contemporary visions for a sustainable bio-based economy established on the principles of green chemistry. Lignin has been identified as both a major obstacle to lignocellulosic biomass processing and as a great opportunity for the renewable production of value-added aromatic compounds. The high variability of lignocellulosic feedstocks combined with the complex chemistry and interactions in the plant cell wall has proven a major challenge for these approaches, but significant efforts are being undertaken to address these challenges. Regardless, it is becoming very clear that efficient utilization of all lignocellulosic components is essential for the economic viability of contemporary bioeconomy concepts.

Chapter 2 reviewed the literature regarding drought stress response and plant cell wall lignification, summarizing the work that has been done and characterizing the current state of our understanding. Increasingly severe and frequent drought may be one of the largest concerns and challenges for a transition to a bio-based economy. High feedstock variability is already a great concern and drought could highly alter both the productivity and composition of lignocellulosic feedstocks. There have been a variety of efforts to improve lignocellulosic feedstocks with selective breeding strategies that reduce or alter lignin content. However, lignin biosynthesis and deposition mechanisms have had around 400 million years of evolution in vascular plants. Thus,

current plants may have developed distinct and highly sophisticated lignin related stress response mechanisms to adapt to various conditions such as drought. Lignin biosynthesis exhibits high plasticity and has been associated with stress response to a variety of biotic and abiotic factors. Lignin has been frequently associated with drought stress response in terms of distribution, content, composition, via interactions with the metabolic pathway, and more. It appears that these responses and mechanisms are likely highly tissue specific demonstrating the complexity of the topic. In general, more studies than not have observed either an increase in lignin content or an increase in the expression of lignin related pathway intermediates in response to drought stress. In addition, modified lignin composition such as altered abundance of certain monomer units has also been identified despite the limited studies that have considered this. Many molecular level response and functional defense mechanisms have been proposed but there is not yet a firm consensus in these related matters.

Chapter 3 provided an introduction to fractionation/pretreatment and an overview the various approaches. Fractionation is one of the first, most expensive, and most important steps in lignocellulosic biomass processing as well as an essential means to study its components such as lignin. The main industrial scale approaches have developed in the pulp and paper industry or for bioethanol production schemes. Fractionation typically operates by either solubilizing the lignin fraction leaving a carbohydrate rich residue or solubilizing the carbohydrate fractions leaving a lignin rich residue. Many emerging approaches are available and more are being developed; however, the most promising approaches are those that effectively isolate the components while preventing undesired reactions such as the breakdown of carbohydrates into toxic compounds and the condensation of lignin structures, all while adhering to the principles of green chemistry. The best approach depends on a variety of factors and will have major implications for downstream processing. Fractionation in the laboratory context was also discussed and the approach varies depending on the application.

Chapter 4 reviewed the main approaches for the characterization of lignin structure and discussed their functions and limitations. There are a variety of wet chemical techniques available, but researchers are shifting to higher-throughput and more comprehensive techniques. Vibrational spectroscopy methods (MIR, NIR, Raman) provide rapid information but are limited by overlapping signals and require multivariate approaches for quantification. GC-MS is becoming a

useful high-throughput tool, but interpretation requires time and experience. Applying chemometric approaches can expedite the analysis and alleviate many hurdles for GC-MS interpretation. NMR provides the most comprehensive analysis but still faces many limitations. New NMR techniques are emerging to address different limitations, such examples include higher dimensional methods, pulse sequences, spectral editing, and wet chemical hybrid methods. Molecular weight is a very important property influencing physical and chemical properties of lignins and SEC is the most common approach for the evaluation of molecular weight distribution of polymers. Complex branched polymers such lignin may not be adequately separated on the basis of hydrodynamic volume and thus SEC determination of lignin molecular weight should be preferably understood in a relative sense. There are still many challenges to lignin characterization due to complexity of plant cell wall and limitations of analytical tools. However, through the application of multiple methods a comprehensive picture of a lignin structure can be developed and validated.

Chapter 5 provided an introduction to the field of chemometrics and subsequently presented an organized review of chemometric approaches applied to the field of lignocellulosic biomass research for lignin valorization. Experimental design was described, which is a fundamental area of chemometrics related to the statistical design of experiments to efficiently acquire high quality data sets. Data preprocessing was introduced and a structured summary of many of the main chemometric preprocessing techniques was developed. Multivariate data analysis was presented including the main approaches for exploratory data analysis supporting classification and clustering and multivariate calibration supporting prediction. A discussion of model optimization, validation, and performance criteria was also included. After introducing the main elements of chemometrics as described above, the application of chemometrics to the field of lignocellulosic biomass research for lignin valorization in categories based on analytical technique was reviewed. This review has established the state of this multidisciplinary sub-space, demonstrating the achievements, discussing trends, limitations, and gaps. Chemometrics is a mature field with a diverse range of effective approaches. While much of our understanding of lignin and lignocellulosic biomass has been established using wet chemical approaches, more efficient high-throughput techniques are becoming increasingly desirable. These rapid techniques require chemometric approaches to accommodate the large and complex data sets that accompany

them. This work highlighted the ubiquity of the most classic chemometric tools PCA and PLS, which have withstood the test of time, while also observing/predicting a shift to more sophisticated machine learning algorithms such as ANN's in the near future. Key accomplishments such as the accurate prediction of molecular weights using FT-IR spectroscopy combined with chemometrics set a strong precedent for future approaches. Altogether, this chapter demonstrated the broad applicability and effectiveness of these tools in the context of their application to lignocellulosic biomass research.

Chapter 6 described an investigation on the effects of drought-induced abiotic stress on the composition of Douglas-fir lignin. Wood was harvested from the outer rings of the stems of two groups of Douglas-fir seedlings, one group grown in severe drought and the other in normal conditions. Lignin was extracted from the wood using a mild organic solvent extraction approach and both the wood and lignin samples were characterized using the following techniques: wet chemical analysis, ATR FT-IR spectroscopy, Py-GC/MS, ^{31}P NMR spectroscopy, solid-state ^{13}C NMR spectroscopy. In addition, an approach for quantifying lignin content in wood via the direct integration of ^{13}C NMR signals was applied and extended, indicating promising results with increased accuracy.

7.2 Key Findings and limitations

This investigation looked at identifying compositional differences between drought stressed and control Douglas-fir seedlings, focusing on the content and composition of lignin, which has been associated with drought response mechanisms in a variety of plant species and tissues. I hypothesised that drought stress would induce chemical changes to the Douglas-fir wood and more specifically I predicted an increase in lignin content and some degree of altered lignin composition. This work has revealed some evidence for an increase in lignin content in Douglas-fir seedling wood of drought-treated plants as compared to the controls. Additionally, compositional differences in lignin hydroxyl content and lignin structural features such as aromatic carbon bonding patterns were observed. Hydroxyl groups are key functionalities of lignin which influence its processing and properties. The drought lignin sample appears to be potentially more hydrophobic than the control based on differences in distinct hydroxyl group contents, which would likely have negative implications for the enzymatic hydrolysis of biomass. These results

suggest that lignin content and composition are impacted by drought stress in Douglas-fir seedlings. However, these differences need further investigation and there are a variety of limitations that should be noted.

Due to the small size of the young Douglas-fir seedling stems (just a few feet tall), harvesting of wood from the outer rings is tedious, susceptible to error, and low yield. Manual shaving of the wood from the stem using a sharp scalpel takes time and patients; moreover, it can be challenging to avoiding inadvertent cutting into the previous years ring. In this drought experiment, the drought treatment was only applied in the last year of growth, so contamination of the sample with the previous years wood will interfere with the results of the experiment. Additionally, each tree yields just around 1 g of sample and thus the wood from 5 trees was combined to produce enough wood for lignin extraction. Since the samples were combined, we must infer the results as a bulk effect and thus tree specific information is lost, potentially diluting the observable differences. To add to the challenge, the organic solvent extraction technique applied to the wood samples was of low yield (~3-5%), which raises two main concerns: the first is that there is not enough wood or lignin for analysis and the second is that the lignin extracted may not be representative of the total lignin. Regarding the first point, even with the combined trees, low yield of lignin greatly limited the analysis that could be performed. For example, the detergent fibre method for lignin content determination of the wood could not be performed in duplicates and limited lignin yield restricted opportunities for many wet chemical approaches. With respect to the second point, the low severity of the extraction technique may indicate that the recovered lignin is more representative of native lignin as it resides in the plant well walls; however, it is also possible, if not probable, that the lignin recovered represents the lignin with bonds more easily cleaved during the extraction process.

While we found evidence for increased lignin content in the drought sample as compared to the control, we have to consider a few important concerns with the methods used to estimate lignin content in the wood samples. Two distinct methods were applied to estimate lignin content: the wet chemical detergent fibre method and a quantitative solid-state ^{13}C NMR approach. The detergent fibre method was not able to identify differences in lignin content between the samples. This is in contradiction with the findings of the solid-state ^{13}C NMR results which predicted higher lignin content in the drought sample. On one hand, while the detergent fibre is known to

underestimate lignin content, it is a robust procedure; on the other hand, solid-state ^{13}C NMR may in theory provide a more representative account of lignin in the sample despite being limited to the natural abundance of the ^{13}C isotope, quantification is hindered by convoluted spectra, and there are not yet well-established procedures for this approach.

Analytical characterization of lignin provided evidence for subtle changes in lignin structure as a result of drought stress; yet there are limitations of the approaches used to assess these changes. FT-IR spectroscopy provides detailed bond-specific information but could not identify differences in the lignin samples. Py-GC/MS could also provide some valuable insight but is challenged by very difficult interpretation. When we consider the results of the quantitative determination of hydroxyl groups using ^{31}P NMR, it is a robust and well-established procedure, performed in duplicates, and thus the results have a higher degree of confidence. While a difference in aliphatic and aromatic hydroxy groups was observed, they are not dramatic differences. A larger sample set would support a stronger conclusion in this regard, but the results are promising and suggest the continued use of the technique for future work. Quantitative solid-state ^{13}C NMR of lignin samples can reveal detailed information regarding distinct lignin carbons and their environment. Despite the high resolution of technique, the extreme heterogeneity and stereochemistry of lignin convolutes the spectrum with broad and overlapping peaks which make assignment challenging and selection of integration bounds somewhat arbitrary. There are five main regions on the ^{13}C NMR spectrum of lignin: aromatic, alkyl-O-aryl, hydroxyl, methoxy, and aliphatic. While the aromatic region is mostly distinct from other carbon types in lignin, internal overlap makes distinction between the types of aromatic carbons (e.g., $Ar-O, Ar-C, Ar-H$) difficult. The methoxy peak is the sharpest in the spectrum and in theory should reflect a roughly 1:1 ratio of methoxy groups per aromatic unit for guaiacyl rich softwood. However, the shoulders (particularly on the downfield side) overlap with the hydroxyl carbons and thus the predictions are highly sensitive to integration boundaries. The alkyl-O-aryl region of solid-state ^{13}C NMR spectra has also recently been suggested to provide accurate estimates of $\beta-O-4$ and $\alpha-O-4$ lignin inter-unit linkages; yet a lack of comparable literature and supporting evidence make validation of these results challenging. The hydroxyl and aliphatic regions are even less reliable. Altogether, while the quantitative solid-state ^{13}C NMR technique provides detailed information and can be

useful for relative comparison, more supporting evidence is required to confirm the reliability of the technique.

7.3 Future work

Despite the many challenges encountered during this project, these obstacles have provided a great opportunity to learn and grow. They support increasingly refined and focused future investigations and identify areas of improvement regarding the approaches applied. Our team is already underway on a follow up investigation with the intent to further explore the hypothesis of this thesis. In order to more comprehensively evaluate our hypothesis, we will be strongly considering implementing many of the following improvements as outlined below.

A key challenge with this investigation was the small sample set and hence the follow up investigation includes over one hundred trees from ten different Douglas-fir families. We intend to evaluate these trees on an individual level in order to identify both tree specific information as well as at the family level. In order to make this possible we will consider approaches for increasing the yield of the organic solvent extraction technique or shift to another approach that can provide research grade lignin with high enough yield to evaluate each tree as an individual sample. We may also consider a MWL extraction procedure as it is more commonly used to make inferences on native lignin and will be more useful for comparison with literature. A larger sample set will also provide a great opportunity for exploring the results using chemometric approaches. We hope that exploratory analysis techniques will be able to identify patterns in our data set or subtle correlations among variables that may go otherwise unnoticed. In addition, we will consider the potential to use multivariate calibration for predicting properties which could support more efficient future investigations. An example would be the prediction of lignin content of wood or the structural features of lignin such as hydroxyl group content using high-throughput approaches like vibrational spectroscopy or Py-GC/MS.

For the determination of lignin content, we may also look to use an alternative approach to the detergent fibre method because of the high sample requirement of the technique. An effective alternative could be the well-known acetyl bromide method which has a significantly reduced sample requirement and relies on the solubilization of lignin and the calibration of absorbance values at distinct wavelength for lignin. This method is gaining popularity and has been shown to

result in higher lignin recovery and thus may provide a more accurate assessment of total lignin content than other approaches. We also intend to more thoroughly evaluate the solid-state ^{13}C NMR direct integration approach for the determination of lignin content as it could be a very promising alternative to the laborious wet chemical procedures. It will be interested to see if our extended method for direct integration will continue to out-perform the previous approach. We will also look to evaluate if the removal of extractives content prior to NMR analysis, calibration approaches, and spectral editing techniques could increase the accuracy of this approach.

While FT-IR was unable to resolve differences between samples, the technique may be more promising with a larger data set. Vibrational spectroscopy has a well-established relationship with chemometric approaches and often chemometric techniques such as PCA have been proven to reveal key variables in large data sets. In particular, score plots that project the data onto principal components to more clearly represent the variation in the data may be able to distinguish differences between drought and control samples. The loadings plot could then demonstrate what variables contribute most to the variation in the data set. For vibrational spectra, these variables would be wavelengths corresponding to specific structural features. Further, ATR FT-IR spectroscopy and multivariate calibration has been shown to accurately predict molecular weight and specific lignin inter-unit linkages, demonstrating the incredible potential for these techniques. We would be very interested in similar approaches that could be applied to our future work which could then be used to facilitate rapid studies with significantly lower sample requirement.

Py-GC/MS is another technique with growing interest in the field of lignocellulosic biomass as a high throughput technique especially suitable for chemometric approaches. Py-GC/MS can provide a wealth of information and the barrier of entry for applying the tool is increasingly being lowered through chemometric tools and software. Open-source, fully automated GC-MS analysis using tensor-based modelling and machine learning tools are among the exciting developments surrounding the approach. Despite there being some challenges with implementing these techniques encountered during this project, it is a very promising approach that we will look to exploit in future work. Arguably one of the most important benefits of the technique is the very modest sample requirement, which I see as a great advantage due to the challenges I encountered with limited sample quantities in this work.

Quantitative solid-state ^{13}C NMR of lignin samples provided detailed structural information including proportions of different aromatic carbons, alkyl-O-aryl bonds and methoxy carbons. The approach is fast and useful, but more validation is required to have more confidence in the findings. Additional NMR techniques with more robust procedures or some wet chemical analysis approaches could be useful for comparison. Furthermore, spectral editing techniques and deconvolution algorithms could also improve the resolution and reduce uncertainty of the results.

Another important aspect of future work that we hope to incorporate is the measurement of the effects of drought stress on the morphology of the trees and the physical and mechanical characteristics of the wood. Evaluation these characteristics will be useful to correlate chemical properties with their functions in drought stress tolerance. This could include factors such as height, diameter, strength of stem, survivability and more. It would also be interesting to evaluate the differences between different tissue or cell types. Many contemporary studies are also considering drought response on the transcriptional level, looking at the regulatory networks that control these responses. We also intend to explore these factors and have already extracted RNA for sequencing from the next sample set.

Altogether, lignocellulosic biomass is a complex matter with incredible potential and a broad range of challenges. Lignin makes up a substantial component of lignocellulosic biomass and is greatly underused despite its abundance and versatility. Thus, regardless of the challenges there are many opportunities in this area. New methods are being proposed and old methods are being improved facilitating more advanced and effective research. The work in this thesis has generated both encouraging results and promising pathways for future work. A final consideration would be the strong encouragement of more interdisciplinary collaboration. The field of lignocellulosic biomass research encompasses a broad range of multiscale and multifaceted challenges that necessitate the communication and interaction between different disciplines.

References

[1] J. Goldemberg, World energy assessment. Energy and the challenge of sustainability., New York: United Nations Development Programme, 2001.

[2] C. Li, X. Zhao, A. Wang, G. W. Huber and T. Zhang, "Catalytic Transformation of Lignin for the Production of Chemicals and Fuels," *Chemical Reviews,* vol. 115, p. 11559–11624, 2015.

[3] H. Lambers, S. F. Chapin III and T. L. Pons, Plant Physiological Ecology, New York: Springer-Verlag, 2008.

[4] Y.-H. P. Zhang, S.-Y. Ding, J. R. Mielenz, J.-B. Cui, R. T. Elander, M. Laser, M. E. Himmel, J. R. McMillan and L. R. Lynd, "Fractionating Recalcitrance Lignocellulose at Modest Reaction Conditions," *Biotechnology and Bioengineering,* vol. 92, no. 2, pp. 214-223, 2007.

[5] S. Akhtari, T. Sowlati and . K. Day, "Economic feasibility of utilizing forest biomass in district energy," *Renewable and Sustainable Energy Reviews,* vol. 33, pp. 117-127, 2014.

[6] M. Patel, X. Zhang and A. Kumar, "Techno-economic and life cycle assessment on lignocellulosic biomass thermochemical conversion technologies: A review," *Renewable and Sustainable Energy Reviews,* vol. 53, pp. 1486-1499, 2016.

[7] "About Renewable Energy," Government of Canada, 13 12 2017. [Online]. Available: https://www.nrcan.gc.ca/our-natural-resources/energy-sources-distribution/renewable-energy/about-renewable-energy/7295. [Accessed 22 12 2020].

[8] "Bio-pathways Project," The Government of Canada, 20 07 2020. [Online]. Available: https://www.nrcan.gc.ca/our-natural-resources/forests-forestry/forest-industry-trade/forest-bioeconomy-bioenergy-biop/bio-pathways-project/13321. [Accessed 23 12 2020].

[9] M. Benali and O. Ajao, "Decision Support Tools for Bioeconomy Transformation Strategies: Introduction of Natural Resources Canada' I-BIOREF Software Platform," Natural Resources Canada, 2018.

[10] F. Cheng and C. E. Brewer, "Producing jet fuel from biomass lignin: Potential pathways to alkyl-benzenes and cycloalkanes," *Renewable and Sustainable Energy Reviews,* vol. 72, pp. 673-722, 2017.

[11] D. S. Bajwa, G. Pourhashem, A. H. Ullah and S. G. Bajwa, "A concise review of current lignin production, applications, products and their environmental impact," *Industrial Crops & Products,* vol. 139, p. 111526, 2019.

[12] J. Zakzeski, P. C. A. Bruijnincx, A. L. Jongerius and . B. M. Weckhuysen, "The Catalytic Valorization of Lignin for the Production of Renewable," *Chemical Reviews,* vol. 110, p. 3552–3599, 2010.

[13] E. R. Lampugnani, G. A. Khan, M. Somssich and S. Persson, "Building a plant cell wall at a glance," *Journal of Cell Science,* vol. 131, 2018.

[14] L. de Vries, S. Guevara-Rozo, M. Cho, L.-Y. Liu, S. Renneckar and S. D. Mansfield, "Tailoring renewable materials via plant biotechnology," *Biotechnology for Biofuels,* vol. 14, p. 167, 2021.

[15] D. H. Northcote, "Chemistry of the plant cell wall," *Annual Review of Plant Physiology,* vol. 23, no. 1, pp. 113-132, 1972.

[16] R. J. Moon, A. Martini, J. Nairn, J. Simonsen and J. Youngblood, "Cellulose nanomaterials review: structure, properties, and nanocomposites," *Chemical Society Reviews,* vol. 40, pp. 3941-3994, 2011.

[17] B. B. Buchanan, W. Gruissem and R. L. Jones, Biochemistry and Molecular Biology of Plants, Wiley, 2015.

[18] L. Vaahtera, J. Schulz and T. Hamann, "Cell wall integrity maintenance during plant development and interation with the environment," *Nature Plants,* vol. 5, pp. 924-932, 2019.

[19] S. P. Chundawat, G. T. Beckham, E. M. Himmel and B. E. Dale, "Deconstruction of Lignocellulosic Biomass to Fuels and Chemicals," *Annual Review of Chemical and Biomolecular Engineering,* vol. 2, pp. 121-145, 2011.

[20] J.-K. Weng and C. Chapple, "The origin and evolution of lignin biosynthesis," *Tansley review,* vol. 187, pp. 273-285, 2010.

[21] B. Zhang, Y. Gao, L. Zhang and Y. Zhou, "The plant cell wall: Biosynthesis, construction, and functions," *Journal of Integrative Plant Biology,* vol. 63, pp. 251-272, 2020.

[22] Y.-C. Lu, Y. Lu and X. Fan, "Structure and Characteristics of Lignin," in *Sharma S., Kumar A. (eds) Lignin. Springer Series on Polymer and Composite Materials*, Springer, Cham, 2020.

[23] M. C. Jarvis, "Structure of native cellulose microfibrils, the starting point for nanocellulose manufacture," *Philosphoical Transactions A,* vol. 376, 2018.

[24] W. Schutyser, T. Renders, S. Van den Bosch, S.-F. Koelewijn, G. T. Beckham and B. F. Sels, "Chemicals from lignin: an interplay of lignocellulose fractionation, depolymerisation, and upgrading," *Chem Soc Rev,* vol. 47, p. 852, 2018.

[25] R. Kumar, S. Singh and O. V. Singh, "Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives," *Journal of Industrial Microbiology and Biotechnology,* vol. 35, pp. 377-391, 2008.

[26] C. Crestini, F. Melone, M. Sette and R. Saladino, "Milled Wood Lignin: A Linear Oligomer," *Biomacromolecules,* vol. 12, p. 3928−3935, 2011.

[27] J. Ralph, C. Lapierre and W. Boerjan, "Lignin structure and its engineering," *Current Opinions in Biotechnology,* vol. 56, pp. 240-249, 2019.

[28] M. Balakshin, E. A. Capanema, X. Zhu, I. Sulaeva, A. Potthast, T. Rosenau and O. J. Rojas, "Spruce milled wood lignin: linear, branched, or cross-linked?," *Green Chemistry,* vol. 22, pp. 3985-4001, 2020.

[29] C. Lapierre, "2 Determining Lignin Structure by Chemical Degredations," in *Lignin and Ligans: Advances in Chemistry*, CRP Press, 2016, pp. 11-48.

[30] A. Gupta, A. Rico-Medina and A. I. Cano-Delgado, "The physiology of plant responses to drought," *Science,* vol. 368, pp. 266-269, 2020.

[31] M. Isaac-Renton, D. Montwe, A. Hamann, H. Spiecker, P. Cherubini and K. Treydte, "Northern forest tree populations are physiologically maladapted to drought," *Nature communications,* vol. 9, p. 5254, 2018.

[32] X. Yang, M. Lu, Y. Wang, Y. Wang, Z. Liu and S. Chen, "Response Mechanism of Plants to Drought Stress," *Horticultuae,* vol. 7, p. 50, 2021.

[33] S. Basu, V. Ramegowda, A. Kumar and A. Pereira, "Plant adaptation to drought stress," *F1000Research,* vol. 5, p. 1554, 2016.

[34] J. C. M. S. Moura, C. A. V. Bonine, J. d. O. F. Viana, M. C. Dornelas and P. Mazzafera, "Abiotic and Biotic Stresses and Changes in the Lignin Content and Composition in Plants," *Journal of Integrated Plant Biology,* vol. 52, no. 4, pp. 360-376, 2010.

[35] L. A. Donaldson, "Abnormal Lignin Distribution in Wood from Severely Drought Stressed Pinus Radiata Trees," *IAWA Journa,* vol. 23, no. 2, pp. 161-178, 2002.

[36] L. Fan, R. Linker, S. Gepstein, E. Tanimoto, R. Yamamoto and P. M. Neumann, "Progressive Inhibition by Water Deficit of Cell Wall Extensibility and Growth along the Elongation Zone of Maize Roots Is Related to Increased Lignin Metabolism and Progressive Stelar Accumulation of Wall Phenolics," *Plant Physiology,* vol. 140, pp. 603-612, 2006.

[37] B.-R. Lee, K.-Y. Kim, W.-J. Jung, J.-C. Avice, A. Ourry and T.-H. Kim, "Peroxidases and lignification in relation to the intensity of water-deficit stress in white clover (Trifolium repens L.)," *Journal of Experimental Botany,* vol. 58, no. 6, pp. 1271-1279, 2007.

[38] Y. Hu, W.-C. Li, Y.-Q. Xu, G.-J. Li, Y. Liao and F.-L. Fu, "Differential expression of candidate genes for lignin biosynthesis under drought stress in maize leaves," *Journal of Applied Genetics,* vol. 50, no. 3, pp. 213-223, 2009.

[39] U. Moura-Sobczak, U. Souza and P. Mazzafera, "Drought stress and changes in the lignin content and composition in Eucalyptus," in *BMC Proceedings*, Arraial d'Ajuda, Bahia, Brazil, 2011.

[40] Z. Li, Y. Peng and X. Ma, "Different response on drought tolerance and post-drought recovery between the small-leafed and the large-leafed white clover (Trifolium repens L.) associated with antioxidative enzyme protection and lignin metabolism," *Acta Physiol Plant,* vol. 35, pp. 213-222, 2013.

[41] R. Terzi, N. S. Guler, N. K. Caliskan and A. Kadioglu, "Lignification response for rolled leaves of Ctenanthe setosa under long-term drought stress," *Turkish Journal of Biology,* vol. 37, pp. 614-619, 2013.

[42] A. B. dos Santos, A. Bottcher, E. Kiyota, J. L. S. Mayer, R. Vincentini, M. d. S. Brito, S. Creste, M. G. Landell and P. Mazzafera, "Water Stress Alters Lignin Content and Related Gene Expression in Two Sugarcane Genotypes," *Journal of Agricultural and Food Chemistry,* vol. 63, pp. 4708-4720, 2015.

[43] S. Srivastava, R. K. Vishwakarma, Y. A. Arafat, S. K. Gupta and B. M. Khan, "Abiotic stress induces change in Cinnamoyl CoA Reductase (CCR) protein abundance and lignin deposition in developing seedlings of Leucaena leucocephala," *Physiology and Molecular Biology of Plants,* vol. 21, no. 2, pp. 197-205, 2015.

[44] C. Xu, X. Fu, R. Liu, L. Guo, L. Ran, C. Li, Q. Tian, B. Jiao, B. Wang and K. Luo, "PtoMYB170 positively regulates lignin deposition during wood formation in poplar and

confers drought tolerance in transgenic Arabidopsis," *Tree Physiology,* vol. 37, pp. 1713-1726, 2017.

[45] T. Van der Weijde, L. M. Huxley, S. Hawkins, E. H. Sembiring, K. Farrar, O. Dolstra, R. G. Visser and L. M. Trindade, "Impact of drought stress on growth and quality of miscanthus for biofuel production," *Global Change Biology Bioenergy,* vol. 9, pp. 770-782, 2017.

[46] H. Wildhagen, S. Paul, M. Allwright, H. K. Smith, M. Malinowska, S. K. Schnabel, M. J. Paulo, F. Cattonaro, V. Vendramin, S. Scalabrin, D. Janz, C. Douthe, O. Brendel, C. Buré, D. Cohen, I. Hummel, D. Le Thiec, F. van Eeuwijk, J. J. Keurentjes, J. Flexas, M. Morgante, P. Robson, M.-B. Bogeat-Triboulot, G. Taylor and A. Polle, "Genes and gene clusters related to genotype and drought-induced variation in saccharification potential, lignin content and wood anatomical traits in Populus nigra," *Tree Physiology,* vol. 38, pp. 320-339, 2017.

[47] J. Yan, A. Aznar, C. Chalvin, D. S. Birdseye, E. E. Baidoo, A. Eudes, P. M. Shih, D. Loque, A. Zhang and H. V. Scheller, "Increased drought tolerance in plants engineered for low lignin and low xylan content," *Biotechnology for Biofuels,* vol. 11, no. 195, 2018.

[48] T. R. Lima, E. C. Carvalho, F. R. Martins, R. S. Oliveira, R. S. Miranda, C. S. Müller, L. Pereira, P. R. Bittencourt, J. C. Sobczak, E. Gomes-Filho, R. C. Costa and F. S. Araújo, "Lignin composition is related to xylem embolism resistance and leaf life span in trees in a tropical semiarid climate," *New Phytologist,* vol. 219, pp. 1252-1262, 2018.

[49] S. W. Bang, D.-K. Lee, H. Jung, P. J. Chung, Y. S. Kim, Y. D. Choi, J.-W. Suh and J.-K. Kim, "Overexpression of OsTF1L, a rice HD-Zip transcriptionfactor, promotes lignin biosynthesis and stomatal closurethat improves drought tolerance," *Plant Biotechnology Journal,* vol. 17, pp. 118-131, 2019.

[50] M. Tu, X. Wang, W. Yin, Y. Wang, Y. Li, G. Zhang, Z. Li, J. Song and X. Wang, "Grapevine VlbZIP30 improves drought resistance by directly activating VvNAC17 and

promoting lignin biosynthesis through the regulation of three peroxidase genes," *Horticulture Research,* vol. 7, no. 150, 2020.

[51] W. Liu, Y. Jiang, C. Wang, L. Zhao, Q. Xing, M. Li, T. Lv and H. Qi, "Lignin synthesized by CmCAD2 and CmCAD3 in oriental melon (Cucumis melo L.) seedlings contributes to drought tolerance," *Plant Molecular Biology,* vol. 103, pp. 689-704, 2020.

[52] W. Xu, W. Tang, C. Wang, L. Ge, J. Sun, X. Qi, Z. He, Y. Zhou, J. Chen, Z. Xu, Y.-Z. Ma and M. Chen, "SiMYB56 Confers Drought Stress Tolerance in Transgenic Rice by Regulating Lignin Biosynthesis and ABA Signaling Pathway," *Frontiers in Plant Science,* vol. 11, no. 785, 2020.

[53] T. Li, Y. Huang, A. Khadr, Y.-H. Wang, Z.-S. Xu and A.-S. Xiong, "DcDREB1A, a DREB-binding transcription factor from Daucus carota, enhances drought tolerance in transgenic Arabidopsis thaliana and modulates lignin levels by regulating lignin-biosynthesis-related genes," *Environmental and Experimental Botany,* vol. 169, p. 103896, 2020.

[54] N. K. Sharma, S. K. Gupta, V. Dwivedi and D. Chattopadhyay, "Lignin deposition in chickpea root xylem under drought," *Plant Signaling & Behavior,* vol. 15, no. 6, p. e1754621, 2020.

[55] H. Gu, Y. Wang, H. Xie, C. Qiu, S. Zhang, J. Xiao, H. Li, L. Chen, X. Li and Z. Ding, "Drought stress triggers proteomic changes involving lignin, favonoids and fatty acids in tea plants," *Nature Scientific Reports,* vol. 10, p. 15504, 2020.

[56] L. Li, K. Yang, S. Wang, Y. Lou, C. Zhu and Z. Gao, "Genome-wide analysis of laccase genes in moso bamboo highlights PeLAC10 involved in lignin biosynthesis and in response to abiotic stresses," *Plant Cell Reports,* vol. 39, pp. 751-763, 2020.

[57] C. Hori, X. Yu, J. C. Mortimer, R. Sano, T. Matsumoto, J. Kikuchi, T. Demura and M. Ohtani, "Impact of abiotic stress on the regulation of cell wall biosynthesis in Populus trichocarpa," *Plant Biotechnology,* vol. 37, pp. 273-283, 2020.

[58] C. Shumin, C. Huang, L. Luo, S. Zheng, Y. Zhong, J. Sun, J. Gui and L. Li, "Cell-Specific Supression of 4-Coumarate-CoA Ligase Gene Reveals Differential Effect of Lignin on Cell Physiological Function in Populus," *Frontiers in Plant Science,* vol. 11, p. 17, 2020.

[59] D. Zhao, Y. Luan, W. Shi, X. Zhang, J. Meng and J. Tao, "A Paeonia ostii caffeoyl-CoA O-methyltransferase confers drought stress tolerance by promoting lignin synthesis and ROS scavenging," *Plant Science,* vol. 303, p. 110765, 2021.

[60] Y. Yan, P. Wang, Y. Lu, Y. Bai, Y. Wei, G. Liu and H. Shi, "MeRAV5 promotes drought stress resistance in cassava by modulating hydrogen peroxide and lignin accumulation," *The Plant Journal,* vol. 107, pp. 847-860, 2021.

[61] W. Wen, R. Wang, L. Su, A. Lv, P. Zhou and Y. An, "MsWRKY11, activated by MsWRKY22, functions in drought tolerance and modulates lignin biosynthesis in alfalfa (Medicago sativa L.)," *Environmental and Experimental Botany,* vol. 184, p. 104373, 2021.

[62] H. Liu, J. Gao, J. Sun, S. Li, B. Zhang, Z. Wang, C. Zhou, D. B. Sulis, J. P. Wang, V. L. Chiang and W. Li, "Dimerization of PtrMYB074 and PtrWRKY19 mediates transcriptional activation of PtrbHLH186 for secondary xylem development in Populus trichocarpa," *New Phytologist,* vol. 234, pp. 918-933, 2022.

[63] A. Casamassimi and A. Ciccodicola, "Transcriptional Regulation: Molecules, Involved Mechanisms, and Misregulation," *International Journal of Molecular Sciences,* vol. 20, p. 1281, 2019.

[64] H. Chen, J. P. Wang, H. Liu, H. Li, Y.-C. J. Lin, R. Shi, C. Yang, J. Gao, C. Zhou, Q. Li, R. R. Sederoff, W. Li and V. L. Chiang, "Hierarchical Transcription Factor and Chromatin Binding Network for Wood Formation in Populus trichocarpa," *The Plant Cell,* vol. 31, pp. 602-626, 2019.

[65] N. Mosier, C. Wyman, B. Dale, R. Elander, Y. Y. Lee, M. Holtzapple and M. Ladisch, "Features of promising technologies for pretreatment of lignocellulosic biomass," *Bioresource Technology,* vol. 96, p. 673–686, 2005.

[66] J. R. Obst and T. K. Kirk, "Isolation of lignin," *Methods in Enzymology,* vol. 161, pp. 3-12, 1988.

[67] P. Bajpai, "Basic Overview of Pulp and Paper Manufacturing Process," in *In: Green Chemistry and Sustainability in Pulp and Paper Industry*, Springer, Cham, 2015, pp. 11-39.

[68] R. A. Young, R. Kundrot and D. A. Tillman, "Pulp and Paper," in *Encyclopedia of Phyiscal Science and Technology*, Academic Press, 249, p. 2001.

[69] N. P. Cheremisinoff and P. E. Rosenfeld, "Chapter 6 - Sources of air emissions from pulp and paper mills," in *Handbook of Pollution Prevention and Cleaner Production*, William Andrew Publishing, 2010, pp. 179-259.

[70] F. S. Chakar and A. J. Ragauskas, "Review of current and future softwood kraft lignin process chemistry," *Industrial Crops and Products,* vol. 20, pp. 131-141, 2004.

[71] C. Xu and F. Ferdosian, "Structure and Properties of Lignin," in *In: Conversion of Lignin into Bio-Based Chemicals and Materials. Green Chemistry and Sustainable Technology*, Springer, Berlin, Heidelberg, 2017, pp. 1-9.

[72] J.-C. Bonhivers and P. R. Stuart, "Applications of Process Integration Methodologies in the Pulp and Paper Industry," in *Handbook of Process Integration (PI) : Minimisation of Energy and Water Use, Waste and Emissions*, Woodhead Publishing Limited, 2013, pp. 765-798.

[73] A. Tribot, G. Amer, M. A. Alio, H. d. Baynast, C. Delattre, A. Pons, J.-D. Mathias, J.-M. Callois, C. Vial, P. Michaud and C.-G. Dussap, "Wood-lignin: Supply, extraction processes and use as bio-based material," *European Polymer Journal,* vol. 112, pp. 228-240, 2019.

[74] P. Bajpai, "Emissions from pulping," in *Biological Odour Treatment*, Springer, Cham, 2014, pp. 9-16.

[75] J. Lora, "Chapter 10 - Industrial Commercial Lignins: Sources, Properties and Applications," in *Monomers, Polymers and Composites from Renewable Resources*, Elsevier, 2008, pp. 225-241.

[76] D. W. K. Chin, S. Lim, Y. L. Pang and M. K. Lam, "Fundamental review of organosolv pretreatment and its challenges in emerging consolidated bioprocessing," *Biofuels, Bioproducts and Biorefining,* vol. 14, no. 4, pp. 808-829, 2020.

[77] H. Luo and M. M. Abu-Omar, "Chemicals from Lignin," *Encyclopedia of Sustainable Technologies,* pp. 573-585, 2017.

[78] G. Gellerstedt, "Chemistry of chemical pulping," in *Pulping chemistry and technology 2*, de Gruyter, 2009, pp. 91-120.

[79] H. Zabed, J. N. Sahu, A. Suely, A. N. Boyce and G. Faruq, "Bioethanol production from renewable sources: Current perspectives and technological progress," *Renewable and Sustainable Energy Reviews,* vol. 71, pp. 475-501, 2017.

[80] H. Zabed, J. N. Sahu, A. N. Boyce and G. Faruq, "Fuel ethanol production from lignocellulosic biomass: An overview on feedstocks and technological approaches," *Renewable and Sustainable Energy Reviews,* vol. 66, pp. 751-774, 2016.

[81] A. Berlin and M. Balakshin, "Industrial Lignins: Analysis, Properties, and Applications," in *Bioenergy Research: Advances and Applications*, Elsevier, 2014, pp. 315-336.

[82] M. P. Pandey and C. S. Kim, "Lignin Depolymerization and Conversion: A Review of Thermochemical Methods," *Chemical Engineering and Technology,* vol. 34, no. 1, pp. 29-41, 2011.

[83] D. M. de Carvalho and L. Colodette, "Comparative Study of Acid Hydrolysis of Lignin and Polysaccharides in biomasses," *Bio Resources,* vol. 12, no. 4, pp. 6907-6923, 2017.

[84] J. J. Bozell, S. E. Chmely, W. Hartwig, R. Key, N. Labbe, P. Venugopal and E. Zuleta, "Lignin Isolation Methodology for Biorefining, Pretreatment and Analysis," in *Lignin Valorization: Emerging Approaches*, Royal Society of Chemistry, 2018, pp. 21-61.

[85] V. B. Agbor, N. Cicek, R. Sparling, A. Berlin and D. B. Levin, "Biomass pretreatment: Fundamentals toward application," *Biotechnology Advances,* vol. 29, pp. 675-685, 2011.

[86] G. Brodeur, E. Yau, K. Badal, J. Collier, K. B. Ramachandran and S. Ramakrishnan, "Chemical and Physicochemical Pretreatment of Lignocellulosic Biomass: A Review," *Enzyme Research,* 2011.

[87] P. Alvira, E. Tomás-Pejó, . M. Ballesteros and M. J. Negro, "Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review," *Bioresource Technology,* vol. 101, p. 4851–4861, 2010.

[88] P. Bajpai, "Pretreatment of Lignocellulosic Biomass," in *Pretreatment of Lignocellulosic Biomass for Biofuel Production*, Singapore, Springer, 2016, pp. 17-70.

[89] Y. Kim, R. Hendrickson, N. S. Mosier and M. R. Ladisch, "Liquid Hot Water Pretreatment of Cellulosic Biomass," in *Biofuels: Methods and Protocols*, Totowa, NJ, Humana Press, 2009, pp. 93-102.

[90] J. Baruah, B. K. Nath, R. Sharma, S. Kumar, R. C. Deka, D. C. Baruah and E. Kalita, "Recent Trends in the Pretreatment of Lignocellulosic Biomass for Value-Added Products," *Frontiers in Energy Research,* vol. 6, no. 141, 2018.

[91] B. E. Dale, "Method for Increasing the Reactivity and Digestibility of Cellulose with Ammonia". US Patent 4600590, 1986.

[92] V. Balan, B. Bals, S. P. Chundawat, D. Marshall and B. E. Dale, "Lignocellulosic Biomass Pretreatment Using AFEX," in *Biofuels: Methods and protocols*, Totowa, NJ, Humana Press, 2009, pp. 61-77.

[93] J. Xu, C. Li, L. Dai, C. Xu, Y. Zhong, F. Yu and C. Si, "Biomass Fractionation and Lignin Fractionation towards Valorization," *ChemSusChem,* vol. 13, pp. 4284-4295, 2020.

[94] Y. Pu, N. Jiang and A. J. Ragauskas, "Ionic Liquid as a Green Solvent for Lignin," *Journal of Wood Chemistry and Technology,* vol. 27, no. 1, pp. 23-33, 2007.

[95] N. Muhammad, Z. Man and M. A. B. Khalil, "Ionic liquid - a future solvent for the enhanced uses of wood biomass," *Eur. J. Wood Prod.,* vol. 70, p. 125–133, 2012.

[96] E. L. Smith, A. P. Abbott and K. S. Ryder, "Deep Euctectic Solvents (DESs) and Their Applications," *Chemical Reviews,* vol. 114, pp. 11060-11082, 2014.

[97] X. Tang, M. Zuo, Z. Li, H. Liu, C. Xiong, X. Zeng, Y. Sun, L. Hu, S. Liu, T. Lei and L. Lin, "Green Processing of Lignocellulosic Biomass and Its Derivatives in Deep Euctectic Solvents," *ChemSusChem,* vol. 10, pp. 2696-2706, 2017.

[98] C. Chio, M. Sain and W. Qin, "Lignin utilization: A review of lignin depolymerization from various aspects," *Renewable and Sustainable Energy Reviews,* vol. 107, pp. 232-249, 2019.

[99] P. Yadav, D. Athanassiadis, I. Antonopoulou, U. Rova, P. Christakopoulos, M. Tysklind and L. Matsakas, "Environmental impact and cost assessment of a novel ligninproduction method," *Journal of Cleaner Production,* vol. 279, p. 123515, 2021.

[100] N. Smolarski, "High-value opportunity for lignin: unlocking its potential," *Frost & Sullivan,* vol. 1, 2012.

[101] J. S. Lupoi, S. Singh, R. Parthasarathi , B. Simmons and R. J. Henry, "Recent innovations in analytical methods for the qualitative and quantitative assessment of lignin," *Renewable and Sustainable Energy Reviews,* vol. 49, pp. 871-906, 2015.

[102] M. N. Collins, M. Nechifor, F. Tanasă, M. Zănoagă, A. McLoughlin, M. A. Stróżyk , M. Culebras and C.-A. Teacă , "Valorization of lignin in polymer and composite systems for advanced engineering applications – A review," *International Journal of Biological Macromolecules,* vol. 131, pp. 828-849, 2019.

[103] K. T. Kirk and J. R. Obst, "Lignin Determination," *Methods in Enzymology,* vol. 161, pp. 87-101, 1988.

[104] C. W. Dence, "The Determination of Lignin," in *In: Lin S.Y., Dence C.W. (eds) Methods in Lignin Chemistry. Springer Series in Wood Science*, Springer, Berlin, Heidelberg, 1992, pp. 33-58.

[105] P. J. Van Soest, "Use of detergents in the analysis of fibrous feeds. II. A rapid method for the determination of fiber and lignin," *Journal of the Association of Official Analytical Chemists,* vol. 46, pp. 829-835, 1963.

[106] H. K. Goering and P. J. Van Soest, "Forage Fiber analysis (apparatus, reagents, procedures, and some applications).," *U.S. Dep. Agric. Agric. Handb.,* vol. 379, 1970.

[107] M. G. Ryan, J. M. Melillo and A. Ricca, "A comparison of methods for determining proximate carbon fractions of forest litter," *Canadian Journal of Forest Research,* vol. 20, no. 2, pp. 166-171, 1990.

[108] M. J. Effland, "Modified procedure to determine acid-insoluble lignin in wood and pulp," *Tappi (United States),* vol. 60, no. 10, 1977.

[109] A. Bjorkman, "Isolation of lignin from finely divided wood with neutral solvents," *Nature,* vol. 174, no. 4440, pp. 1057-1058, 1954.

[110] K. Lundquist, "Wood - Isolation and Purification," in *In: Lin S.Y., Dence C.W. (eds) Methods in Lignin Chemistry. Springer Series in Wood Science*, Springer, Berlin, Heidelberg, 1992, pp. 63-74.

[111] K. M. Holtman, H.-M. Chang and J. F. Kadla, "Solution-State Nuclear Magnetic Resonance Study of the Similarities between Milled Wood Lignin and Cellulolytic Enzyme Lignin," *Journal of Agricultural and Food Chemistry,* vol. 52, pp. 720-726, 2004.

[112] Z. Hu, T. Yeh, H. Chang, Y. Matsumoto and J. F. Kadla, "Elucidation of the structure of cellulolytic enzyme lignin," *Holzforschung,* vol. 60, no. 4, pp. 389-397, 2006.

[113] F. G. Calvo-Flores, J. A. Dobado, J. Isac-Garcia and F. J. Martin-Martinez, Lignin and lignans as renewable raw materials: chemistry, technology and applications, John Wiley & Sons, 2015.

[114] A. Guerra, I. Filpponen, L. A. Lucia, C. Saquing, S. Baumberger and D. S. Argyropoulos, "Toward a Better Understanding of the Lignin Isolation Process from Wood," *Journal of Agricultural and Food Chemistry,* vol. 54, p. 5939−5947, 2006.

[115] G. Brunow, K. Lundquist and G. Gellerstedt, "Lignin," in *Analytical Methods in Wood Chemistry, Pulping, and Papermaking*, Springer-Verlag Berlin Heidelberg, 1999, pp. 77-124.

[116] K. M. Holtman, H.-m. Chang and H. Jameel, "Quantitative 13C NMR Characterization of Milled Wood Lignins Isolated by Different Milling Techniques," *Journal of Wood Chemistry and Technology,* vol. 26, pp. 21-34, 2006.

[117] A. E. Kazzaz and P. Fatehi, "Technical lignin and its potential modification routes: A mini-review," *Industrial Crops & Products,* vol. 154, p. 112732, 2020.

[118] M. Y. Balakshin, E. A. Capanema and H. -M. Chang, "Recent Advances in the Isolation and Analysis of Lignins and Lignin-Carbohydrate Complexes," in *In Characterization of Lignocellulosic Materials, T.Q. Hu (Ed.).*, Blackwell Publishing Ltd., 2008, pp. 148-170.

[119] K. Lundquist, "6.1 Acidolysis," in *Methods in Lignin Chemistry*, Berlin Heidelberg, Springer-Verlag, 1992, pp. 289-300.

[120] J. M. Pepper, P. E. Baylis and E. Alder, "THE ISOLATION AND PROPERTIES OF LIGNINS OBTAINED BY THE ACIDOLYSIS OF SPRUCE AND ASPEN WOODS IN DIOXANE-WATER MEDIUM," *Canadian Journal of Chemistry,* vol. 37, no. 8, pp. 1241-1248, 1959.

[121] K. Lundquist and T. K. Kirk, "Acid Degradation of Lignin," *Act Chemica Scandinavica,* vol. 25, pp. 889-894, 1971.

[122] A. R. Robinson and S. D. Mansfield, "Rapid analysis of poplar lignin monomer composition by a streamlined thioacidolysis procedure and near-infrared reflectance-based preduction modeling," *The Plant Journal,* vol. 58, pp. 706-714, 2009.

[123] J. H. Jung, W. M. Fouad, W. Vermerris, G. Maria and F. Altpeter, "RNAi suppression of lignin biosynthesis in sugarcane reduces recalcitrance for biofuel production from lignocellulosic biomass," *Plant Biotechnology Journal,* vol. 10, pp. 1067-1076, 2012.

[124] C. Fu, . J. R. Mielenz, X. Xiao, . Y. Ge, C. Y. Hamilton, . M. Rodriguez, J. F. Chen, M. Foston, . A. Ragauskas, J. Bouton, R. A. Dixon and Z.-Y. Wang, "Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass," *Proc. Natl. Acad. Sci. U.S.A.,* vol. 108, no. 9, p. 3803−3808, 2011.

[125] X. Li, J.-K. Weng and C. Chapple, "Improvement of biomass through lignin modification," *The Plant Journal,* vol. 54, pp. 569-581, 2008.

[126] S. D. Mansfield, "Solutions for dissolution—engineering cell walls for deconstruction," *Current Opinion in Biotechnology,* vol. 20, pp. 286-294, 2009.

[127] E. A. Capanema, M. Y. Balakshin and J. F. Kadla, "A Comprehensive Approach for Quantitative Lignin Characterization by NMR Spectroscopy," *Journal of Agricultural and Food Chemistry,* vol. 52, pp. 1850-1860, 2004.

[128] T. P. Schultz and M. C. Templeton, "Proposed mechanism for the nitrobenzene oxidation of lignin," *Holzforschung,* vol. 40, no. 2, pp. 93-97, 1986.

[129] C. L. Chen, "6.2 Nitrobenzene and Cupric Oxide Oxidations," in *Methods in Lignin Chemistry*, Berlin Heidelberg, Springer-Verlag, 1992, pp. 301-321.

[130] T. Ohra-aho, F. Gomes, J. L. Colodette and T. Tamminen, "S/G ratio and lignin structure among Eucalyptus hybrids determined by Py-GC/MS and nitrobenzene oxidation," *Journal of Analytical and Applied Pyrolysis,* vol. 101, pp. 166-171, 2013.

[131] B. Jiang, Y. Zhang, H. Zhao, T. Guo, W. Wu and Y. Jin, "Structure-antioxidant relationship of active oxygen catalytic lignin and lignin-carbohydrate complex," *International Journal of Biological Macromolecules,* vol. 139, pp. 21-29, 2019.

[132] E. Billa, M.-T. Tollier and B. Monties, "Characterisation of the Monomeric Composition ofin situWheat Straw Lignins by Alkaline Nitrobenzene Oxidation: Effect of Temperature and Reaction Time," *Journal of the Science of Food and Agriculture,* vol. 72, pp. 250-256, 1996.

[133] G. Gellerstedt, "6.3 Chemical Degredation Methods: Permanganate Oxidation," in *Methods in Lignin Chemistry*, Berlin Heidelberg, Springer-Verlag, 1992, pp. 322-333.

[134] W. G. Glasser, "Classification of Lignin According to Chemical and Molecular Structure," in *Lignin: Historical, Biological, and Materials Perspectives.*, American Chemical Society, 1999, pp. 216-238.

[135] C. Rolando, B. Monties and C. Lapierre, "Thioacidolysis," in *Methods in Lignin Chemistry*, Springer, 1992, pp. 334-349.

[136] F. Lu and J. Ralph, "The DFRC (Derivatization Followed By Reductive Cleavage) Method," in *Lignin: Structrual Analysis, Applications in Biomaterials, and Ecological Significance*, Hauppauge, NY, Nova Science Publishers, Inc., 2014, pp. 27-65.

[137] F. Lu and J. Ralph, "Derivatization Followed by Reductive Cleavage (DFRC Method), a New Method for Lignin Analysis: Protocol for Analysis of DFRC Monomers," *Journal of Agricultural and Food Chemistry,* vol. 45, pp. 2590-2592, 1997.

[138] F. Lu and J. Ralph, "Chapter 6 - Lignin," in *Cereal Straw as a Resource for Sustainable Biomaterials and Biofuels*, Amsterdam, Elsevier, 2010, pp. 169-207.

[139] K. K. Nayak, P. Parkhey and R. Sahu, "Analysis of Lignin Using Qualitative and Quantitative Methods," in *In: Sharma S., Kumar A. (eds) Lignin. Springer Series on Polymer and Composite Materials*, Springer Cham, 2020, pp. 115-138.

[140] A. Lourenco and H. Pereira, "Compositional variabiltiy of lignin in biomass. Lignin - trends and applications," *InTech,* vol. 21, pp. 65-98, 2018.

[141] O. Faix, "Fourier Transform Infrared Spectroscopy," in *Methods in Lignin Chemistry*, Berlin Heidelberg, Springer-Verlag, 1992, pp. 83-109.

[142] U. P. Agarwal, "An overview of Raman spectroscopy as applied to lignocellulosic materials," in *Advances in lignocellulosics characterization*, 1999, pp. 201-225.

[143] P. N. Perera, M. Schmidt, V. L. Chiang, P. J. Schuck and P. D. Adams, "Raman-spectroscopy-based noninvasive microanalysis of native lignin structure," *Analytical and bioanalytical chemistry,* vol. 402, no. 2, pp. 983-987, 2012.

[144] J. Ralph and L. L. Landucci, "Chapter 5. NMR of lignins," in *Lignin and lignans: advances in chemistry*, Boca Raton, FL, Taylor & Francis, 2010, pp. 137-244.

[145] X. Meng, C. Crestini, H. Ben, N. Hao, Y. Pu, A. J. Ragauskas and D. S. Argyropoulos, "Determination of hydroxyl groups in biorefinery resources via quantitative 31P NMR spectroscopy," *Nature Protocols,* vol. 14, pp. 2627-2647, 2019.

[146] M. Balakshin and E. Capanema, "On the Quantification of Lignin Hydroxyl Groups With 31P and 13C NMR Spectroscopy," *Journal of Wood Chemistry and Technology,* vol. 35, pp. 220-237, 2015.

[147] D. S. Argyropoulos, "Quantitative Phosphorus-31 NMR Analysis of Lignins, a New Tool for the Lignin Chemist," *Journal of Wood Chemistry and Technology,* vol. 14, no. 1, pp. 45-63, 1994.

[148] O. D. Sparkman, Z. E. Penton and F. G. Kitson, "Chapter 1 - Introduction and History," in *In: O. David Sparkman, Zelda E. Penton, Fulton G. Kitson. Gas Chromatography and Mass Spectrometry (Second Edition)*, Academic Press, 2011, pp. 2-13.

[149] C. -L. Chen, "Gas Chromatography-Mass Spectrometry (GS-MS)," in *In: Lin S.Y., Dence C.W. (eds) Methods in Lignin Chemistry. Springer Series in Wood Science*, Springer, Berlin, Heidelberg, 1992, pp. 527-548.

[150] O. D. Sparkman, Z. E. Penton and F. G. Kitson, "Chapter 2 - Gas Chromatography," in *In: O. David Sparkman, Zelda E. Penton, Fulton G. Kitson. Gas Chromatography and Mass Spectrometry (Second Edition)*, Academic Press, 2011, pp. 15-83.

[151] O. D. Sparkman, Z. E. Penton and F. G. Kitson, "Chapter 4 - Mass Spectrometry Instrumentation," in *In: O. David Sparkman, Zelda E. Penton, Fulton G. Kitson. Gas Chromatography and Mass Spectrometry (Second Edition)*, Academic Press, 2011, pp. 89-148.

[152] O. D. Sparkman, Z. E. Penton and F. G. Kitson, "Chapter 5 - Mass Spectral Data Interpretation," in *In: O. David Sparkman, Zelda E. Penton, Fulton G. Kitson. Gas Chromatography and Mass Spectrometry (Second Edition)*, Academic Press, 2011, pp. 149-205.

[153] A. Alves, M. Schwanninger, H. Pereira and J. Rodrigues, "Analytical pyrolysis as a direct method to determine the lignin content in wood: Part 1: Comparison of pyrolysis lignin with Klason lignin," *Journal of Analytical and Applied Pyrolysis,* vol. 76, no. 1-2, pp. 209-213, 2006.

[154] Y. Lu, X.-Y. Wei, J.-P. Cao, P. Li, F.-J. Liu, Y.-P. Zhao, X. Fan, W. Zhao, L.-C. Rong, Y.-B. Wei, S.-Z. Wang, J. Zhou and Z.-M. Zong, "Characterization of a bio-oil from pyrolysis of rice husk by detailed compositional analysis and structural investigation of lignin," *Bioresource Technology,* vol. 116, pp. 114-119, 2012.

[155] F. W. Karasek and R. E. Clement, Basic gas chromatography-mass spectrometry: principles and techniques, Elsevier, 2012.

[156] Z. Zeng, J. Li, H. M. Hugel, G. Xu and P. J. Marriott, "Interpretation of comprehensive two-dimensional gas chromatography data using advanced chemometrics," *Trends in Analytical Chemistry,* vol. 53, pp. 150-166, 2014.

[157] N. Heshka, R. Gieleciak and C. Lay, "Characterization of lignin samples by pyrolysis and gas chromatography-mass spectrometry," Natural Resources Canada, CDEV-2015-0011-RT.

[158] D. Meier and O. Faix, "Pyrolysis-Gas Chromatography-Mass Spectrometry," in *In: Lin S.Y., Dence C.W. (eds) Methods in Lignin Chemistry*, Springer Series in Wood Science. Springer, Berlin, Heidelberg, 1992, pp. 177-199.

[159] B. Scholze and D. Meier, "Characterization of the water-insoluble fraction from pyrolysis oil (pyrolytic lignin) Part I. PY-GC/MS, FTIR, and functional groups," *Journal of Analytical and Applied Pyrolysis,* vol. 60, no. 1, pp. 41-54, 2001.

[160] C. Lay, R. Gieleciak and N. Heshka, "Catalytic pyrolysis of biomass-derived products," Natural Resources Canada, CanmetENERGY - CDEV-2016-0054-RE.

[161] M. M. Abu-Omar, K. Barta, G. T. Beckham, J. S. Luterbacher, J. Ralph, R. Rinaldi, Y. Roma´n-Leshkov, J. S. M. Samec, B. F. Sels and F. Wang, "Guidelines for performing lignin-first biorefining," *Energy & Environmental Science,* vol. 14, p. 262, 2021.

[162] M. Y. Balakshin, E. A. Capanema, I. Sulaeva, P. Schlee, Z. Huang, M. Feng, M. Borghei, O. J. Rojas, A. Potthast and T. Rosenau, "New Opportunities in the Valorization of Technical Lignins," *ChemSusChem,* vol. 14, pp. 1-22, 2021.

[163] J. D. McMillan, "Enzymatic Conversion of Biomass for Fuels Production," in *Pretreatment of Lignocellulosic Biomass*, American Chemical Society, 1994, pp. 292-324.

[164] R.-C. Sun, "Lignin Source and Structural Characterization," *ChemSusChem,* vol. 13, pp. 4385-4393, 2020.

[165] S. Gillet, M. Aguedo, L. Petitjean, A. R. Morais, A. M. da Costa Lopes, R. M. Lukasik and P. T. Anastas, "Lignin transformations for high value applications: towards targeted modifications using green chemistry," *Green Chemistry,* vol. 19, pp. 4200-4233, 2017.

[166] K. E. Achyuthan, A. M. Achyuthan, P. D. Adams, S. M. Dirk, J. C. Harper, B. A. Simmons and A. K. Singh, "Supramolecular Self-Assembled Chaos: Polyphenolic Lignin's Barrier to Cost-Effective Lignocellulosic Biofuels," *Molecules,* vol. 15, pp. 8641-8688, 2010.

[167] S.-Y. Ding, Y.-S. Liu, Y. Zeng, M. E. Himmel, J. O. Baker and E. A. Bayer, "How Does Plant Cell Wall Nanoscale Architecture Correlate with Enzymatic Digestibility?," *Science,* vol. 338, no. 6110, pp. 1055-1060, 2012.

[168] M. Li, Y. Pu and A. J. Ragauskas, "Current Understanding of the Correlation of Lignin Structure with Biomass Recalcitrance," *Frontiers in Chemistry,* vol. 4, p. 45, 2016.

[169] L. Catherine, "Application of new methods for the investigation of lignin structure," in *In Forage Cell Wall Structure and Digestibility (eds H.G. Jung, D.R. Buxton, R.D. Hatfield and J. Ralph)*, 1993, pp. 133-166.

[170] R. Vanholme, K. Morreel, J. Ralph and W. Boerjan, "Lignin engineering," *Currnet Opinion in Plant Biology,* vol. 11, pp. 278-285, 2008.

[171] R. Rinaldi, R. Jastrzebski, M. T. Clough, J. Ralph, M. Kennema, P. C. A. Bruijnincx and B. M. Weckhuysen, "Paving the Way for Lignin Valorisation: Recent Advances in Bioengineering, Biorefining and Catalysis," *The Journal of the German Chemical Society,* vol. 55, pp. 8164-8215, 2016.

[172] O. A. T. Dias, D. R. Negrão, D. F. C. Gonçalves, I. Cesarino and A. L. Leão, "Recent approaches and future trends for ligninbased materials," *Molecular Crystals and Liquid Crystals,* vol. 655, pp. 204-223, 2017.

[173] R. Liu, L. Dai, C. Xu, K. Wang, C. Zheng and C. Si, "Lignin-Based Micro- and Nanomaterials and their Composites in Biomedical Applications," *ChemSusChem,* vol. 13, no. 17, pp. 1864-5631, 2020.

[174] M. P. Vinardell and M. Mitjans, "Lignins and Their Derivatives with Beneficial Effects on Human Health," *International Journal of Molecular Sciences,* vol. 18, p. 1219, 2017.

[175] L. Dai, R. Liu, L.-Q. Hu, Z.-F. Zou and C.-L. Si, "Lignin Nanoparticle as a Novel Green Carrier for the Efficient Delivery of Resveratrol," *ACS Sustainable Chemistry & Engineering,* vol. 5, pp. 8241-8249, 2017.

[176] M. Witzler, A. Alzagameem, M. Bergs, B. E. Khaldi-Hansen, S. E. Klein, D. Hielscher, B. Kamm, J. Kreyenschmidt, E. Tobiasch and M. Schulze, "Lignin-Derived Biomaterials for Drug Release and Tissue Engineering," *Molecules,* vol. 23, p. 1885, 2018.

[177] C. Moretti, B. Corona, R. Hoefnagels, I. Vural-Gürsel, R. Gosselink and M. Junginger, "Review of life cycle assessments of lignin and derived products: Lessons learned," *Science of the Total Environment,* vol. 770, p. 144656, 2021.

[178] J. K. Kurian, G. R. Nair, A. Hussain and V. Raghavan, "Feedstocks, logistics and pre-treatment processes for sustainable lignocellulosic biorefineries: A comprehensive review," *Renewable and Sustainable Energy Reviews,* vol. 25, pp. 205-219, 2013.

[179] C. G. Yoo, X. Meng, Y. Pu and A. J. Ragauskas, "The critical role of lignin in lignocellulosic biomass conversion and recent pretreatment strategies: A comprehensive review," *Bioresource Technology,* vol. 301, p. 122784, 2020.

[180] H. Pereira, J. Graca and J. C. Rodrigues, "Wood chemistry in relation to quality.," in *Wood quality and its biological basis*, Boca Raton, Forida, CRC Press, 2003, pp. 53-86.

[181] C.-M. Popescu, C. Vasile, M.-C. Popescu, G. Singurel, V. I. Popa and B. S. Munteanu, "Analytical Methods for Lignin Characterization. II. Spectroscopic Studies," *Cellulose Chemsitry and Technology,* vol. 40, no. 8, pp. 597-622, 2006.

[182] Y. Lu, Y.-C. Lu, H.-Q. Hu, F.-J. Xie, X.-Y. Wei and X. Fan, "Structural Characterization of Lignin and Its Degradation Products with Spectroscopic Methods," *Journal of Spectroscopy,* vol. 2017, p. 8951658, 2017.

[183] B. R. Kowalski, "Chemometrics: Views and Propositions," *The Journal of Chemical Information and Computer Sciences,* vol. 15, no. 4, pp. 201-203, 1975.

[184] P. Geladi and K. Esbensen, "The Start and Early History of Chemometrics: Selected Interviews. Part 1," *Journal of Chemometrics,* vol. 4, pp. 337-354, 1990.

[185] K. Esbensen and P. Geladi, "The Start and Early History of Chemometrics: Selected Interviews. Part 2," *Journal of Chemometrics,* vol. 4, pp. 389-412, 1990.

[186] M. M. Sena, J. W. B. Braga, M. C. Breitkreitz, M. F. Ferrao and C. S. Silva, "Chemometrics in Bioanalytical Chemistry," in *Tools and Trends in Bioanalytical Chemistry*, Springer, 2021, pp. 497-541.

[187] J. Gasteiger, "The Scope of Chemoinformatics," in *Handbook of Chemoinformatics*, 2003, pp. 3-5.

[188] J. Gasteiger, "Introduction," in *Chemoinformatics: A Textbook*, 2003, pp. 1-13.

[189] K. Varmuza and P. Filzmoser, Introduction to Multivariate Statistical Analysis in Chemometrics, CRC Press, 2009.

[190] J. Gasteiger, "Chemoinformatics: Achievements and Challenges, a Personal View," *Molecules,* vol. 21, p. 151, 2016.

[191] J. Gasteiger, "Frontmatter," in *Handbook of Chemoinformatics*, 2003, pp. 1-44.

[192] S. Chonde and S. Kumara, "Cheminformatics: An Introductory Review," in *IIE Annual Conference.Proceedings 2316-2325*, 2014.

[193] B. Lavine and J. J. Workman, "Chemometrics: Past, Present, and Future," in *Chemometrics and Cheminformatics ACS Symposium Series*, American Chemical Society, 2005, pp. 1-13.

[194] A. Iwaniak, P. Minkiewicz, M. Darewicz, M. Protasiewicz and D. Mogut, "Chemometrics and cheminformatics in the analysis of biologically active peptides from food sources," *Journal of Functional Foods,* vol. 16, pp. 334-351, 2015.

[195] B. K. Lavine and J. W. Jr, "Chemometrics," *Analytical Chemistry,* vol. 85, pp. 705-714, 2013.

[196] K. Kjeldahl and R. Bro, "Some common misunderstandings in chemometrics," *Journal of Chemometrics,* vol. 24, pp. 558-564, 2012.

[197] S. N. Deming, J. A. Palasota and J. M. Palasota, "Experimental Design in Chemometrics," *Journal of Chemometrics,* vol. 5, pp. 181-192, 1991.

[198] R. G. Brereton, Applied Chemometrics for Scientists, Chichester, England: John Wiley & Sons, 2007.

[199] R. Phan-Tan-Luu and R. Cela, "1.15 - Experimental Designs: Conclusions, Terminology, and Symbols," in *Comprehensive Chemometrics, Editor(s): Steven D. Brown, Romá Tauler, Beata Walczak,*, Elsevier, 2009, pp. 501-505.

[200] S. Roussel, S. Preys, F. Chauchard and J. Lallemand, "Chapter 2 Multivariate Data Analysis (Chemometrics)," in *In: O'Donnell C., Fagan C., Cullen P. (eds) Process Analytical Technology for the Food Industry. Food Engineering Series.*, New York, Springer, 2014, pp. 7-60.

[201] R. Leardi, "Experimental design in chemistry: A tutorial," *Analytical Chimica Acta,* vol. 652, pp. 161-172, 2009.

[202] M. Karlsson, V. L. Vegunta, R. Deshpande and M. Lawoko, "Protected lignin biorefining through cyclic extraction: gaining fundamental insights into the tuneable properties of lignin by chemometrics," *Green Chemistry,* vol. 24, pp. 1211-1223, 2022.

[203] R. G. Brereton, J. Jansen, J. Lopes, F. Marini, A. Pomerantsev, O. Rodionova, J. M. Roger, B. Walczak and R. Tauler, "Chemometrics in analytical chemistry—part I: history,

experimental design and data analysis tools," *Analytical and Bioanalytical Chemistry,* vol. 409, p. 5891–5899, 2017.

[204] J. Duckworth, "Mathematical Data Preprocessing," in *Near-Infrared Spectroscopy in Agriculture, Volume 44*, 2004, pp. 113-132.

[205] J. Gerretzen, E. Szymanska, J. J. Jansen, J. Bart, H.-J. v. Manen, E. R. van den Heuvel and L. M. Buydens, "Simple and Effective Way for Data Preprocessing Selection Based on Design of Experiments," *Analytical Chemistry,* vol. 87, pp. 12096-12103, 2015.

[206] K. Varmuza, "Multivariate Data Analysis in Chemistry," in *Handbook of Chemoinformatics: From Data to Knowledge in 4 volumes*, Wiley, 2003, pp. 1098-1133.

[207] H. Jonsson and J. Gabrielsson, "2.11 - Evaluation of Preprocessing Methods," in *Comprehensive Chemometrics*, Elsevier, 2009, pp. 199-206.

[208] P. R. Mobley, B. R. Kowalski, J. J. Workman and R. Bro, "Review of Chemometrics Applied to Spectroscopy: 1985-95, Part 2," *Applied Spectroscopy Reviews,* vol. 31, no. 4, pp. 347-368, 1996.

[209] P. Mishra, A. Biancolillo, J. M. Roger, F. Marini and D. N. Rutledge, "New data preprocessing trends based on ensemble of multiple preprocessing techniques," *Trends in Analytical Chemistry,* vol. 132, p. 116045, 2020.

[210] A. Craig, O. Cloarec, E. Holmes, J. K. Nicholson and J. C. Lindon, "Scaling and Normalization Effects in NMR Spectroscopic Metabonomic Data Sets," *Analytical Chemistry,* vol. 78, pp. 2262-2267, 2006.

[211] R. G. Brereton, Chemometrics for Pattern Recognition, John Wiley & Sons, Ltd, 2009.

[212] K. Pearson, "On lines and planes of closest fit to systems of points in space," *The London, Edinburgh Philosophical Magazine and Journal of Science,* vol. 2, no. 11, pp. 559-572, 1901.

[213] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of educational psychology,* vol. 24, no. 6, p. 417, 1933.

[214] I. T. Jolliffe, Principle Component Analysis, New York: Springer-Verlag, 2002.

[215] J.-C. Jong and S. Kotz, "On a Relation Between Principal Components and Regression Analysis," *The American Statistician,* vol. 53, no. 4, pp. 349-351, 1999.

[216] K. H. Esbensen and P. Geladi, "2.13 - Principle Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice," in *Comprehensive Chemometrics*, Elsevier, 2009, pp. 211-226.

[217] S. Wold, K. Esbensen and P. Geladi, "Principle Component Analysis," *Chemometrics and Intelligent Laboratory Systems,* vol. 2, pp. 37-52, 1987.

[218] S. Wold, C. Albano, J. W. Dunn III, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg and M. Sjostrom, "Multivariate Calibration," in *In: Kowalski B.R. (eds) Chemometrics Mathematics and Statistics in Chemistry*, Dordrecht, Springer, 1984, pp. 147-156.

[219] A. de Juan and R. Tauler, "Linear Soft-Modeling: Introduction," in *Comprehensive Chemometrics*, Elsevier B.V., 2009, pp. 207-210.

[220] F. Westad and M. Kermit, "Independent Component Analysis," in *Comprehensive Chemometrics*, Elsevier B.V., 2009, pp. 227-248.

[221] W. H. Lawton and E. A. Sylvestre, "Self Modeling Curve Resolution," *Technometrics,* vol. 13, no. 3, pp. 617-633, 1971.

[222] P. Ebrahimi, N. Vierkeck, R. Bro and S. B. Engelsen, "Chemometric Analysis of NMR Spectra," in *In: Webb G. (eds) Modern Magnetic Resonance*, Springer Cham, 2017.

[223] S. C. Rutan, A. de Juan and R. Tauler, "Introduction to Multivariate Curve Resolution," in *Comprehensive Chemometrics*, Elsevier B.V., 2009, pp. 249-258.

[224] A. de Juan, J. Jaumot and R. Tauler, "Multivariate Curve Resolution (MCR). Solving the mixture analysis problem," *Analytical Methods,* vol. 6, pp. 4964-4976, 2014.

[225] R. Tauler and A. de Juan, "Multivariate Curve Resolution for Quantitative Analysis," in *Data Handling in Science and Technology*, Elsevier, 2015, pp. 247-292.

[226] P. Xanthopoulos, P. M. Pardalos and T. B. Trafalis, "Linear Discriminant Analysis," in *Robust Data Mining*, vol. 7, Springer, 2013, pp. 27-33.

[227] J. M. Keller, M. R. Gray and J. A. Givens, Jr, "A Fuzzy K-Nearest Neighbor Algorithm," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 15, no. 4, pp. 580-585, 1985.

[228] S. Wold and M. Sjostrom, "SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy," in *Chemometrics: Theory and Application*, Washington, DC, American Chemical Society, 1977, pp. 243-282.

[229] F. Marini, R. Bucci, A. L. Magri and A. D. Magri, "Artificial neural networks in chemometrics: History, examples and perspectives," *Microchemical Journal,* vol. 88, pp. 178-185, 2008.

[230] P. Xanthopolous, P. M. Pardalos and T. B. Trafalis, "Support Vector Machines," in *Robust Data Mining*, Springer, 2013, pp. 35-48.

[231] B. Lavine and J. Workman, "Chemometrics," *Analytical Chemistry,* vol. 80, no. 12, pp. 4519-4531, 2008.

[232] S. D. Brown, T. B. Blank, S. T. Sum and L. G. Weyer, "Chemometrics," *Analytical Chemistry,* vol. 66, pp. 315R-359R, 1994.

[233] J. H. Kalivas, "3.01 - Calibration Methodologies," in *Comprehensive Chemometrics*, Elsevier, 2009, pp. 1-32.

[234] H. Martens and T. Naes, Multivariate Calibration, John Wiley & Sons, 1992.

[235] H. Martens and T. Naes, "Multivariate Calibration," in *In: Kowalski B.R. (eds) Chemometrics Mathematics and Statistics in Chemistry*, Dordrecht, Springer, 1984, pp. 147-156.

[236] T. P. Ryan, "Chapter 4 Introduction to Multiple Linear Regression," in *Modern Regression Methods, 2nd Edition*, John Wiley & Sons, Inc., 2009, pp. 146-189.

[237] N. Kumar, A. Bansal, G. S. Sarma and R. K. Rawal, "Chemometrics tools used in analytical chemistry: An overview," *Talanta,* vol. 123, pp. 186-199, 2014.

[238] W. Massy, "Principal Components Regression in Exploratory Statistical Research," *Journal of the American Statistical Association,* vol. 60, no. 309, pp. 234-256, 1965.

[239] M. G. Kendall, A Course in Multivariate Analysis, London, England: Charles Griffen and Company, Ltd., 1957.

[240] I. T. Jolliffe, "A Note on the Use of Principal Components in Regression," *Applied Statistics,* vol. 31, no. 3, pp. 300-303, 1982.

[241] H. Wold, "Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach," *Journal of Applied Probability,* vol. 12, no. S1, pp. 117-142, 1975.

[242] S. Wold, M. Sjostrom and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and intelligent laboratory systems,* vol. 58, no. 2, pp. 109-130, 2001.

[243] A. C. Olivieri, "Non-linearity and Artificial Neural Networks," in *Introduction to multivariate calibration*, Springer, Cham, 2018, pp. 207-226.

[244] G. Dreyfus, Neural netowrks: methodology and applications, Springer Science & Business Media, 2005.

[245] L.-N. Li, X.-F. Liu, F. Yang, W.-M. Xu, J.-Y. Wang and R. Shu, "A review of artificial neural network based chemometrics applied in laser-induced breakdown spectroscopy analysis," *Spectrochimica Acta Part B: Atomic Spectroscopy,* vol. 180, p. 106183, 2021.

[246] X. Bian, P. Diwu, Y. Liu, P. Liu, Q. Li and X. Tan, "Ensemble calibration for the spectral quantitative analysis of complex samples," *Journal of Chemometrics,* vol. 32, p. e2940, 2018.

[247] C. L. Morais, M. C. Santos, K. M. Lima and F. L. Martin, "Improving data splitting for classification applications in spectrochemical analyses employing a random-mutation Kennard-Stone algorithm approach," *Bioinformatics,* vol. 35, no. 24, pp. 5257-5263, 2019.

[248] R. G. Brereton, "Validation and Optimization," in *Chemometrics for Pattern Recognition*, John Wiley & Sons, Ltd, 2009, pp. 311-391.

[249] M. B. Seasholtz and B. Kowalski, "The parsimony principle applied to multivariate calibration," *Analytica Chimica Acta,* vol. 277, pp. 165-177, 1993.

[250] F. Xu, J. Yu, T. Tesso, F. Dowell and D. Wang, "Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review," *Applied Energy,* vol. 104, pp. 801-809, 2013.

[251] R. Bro, K. Kjeldahl, A. K. Smilde and H. A. Kiers, "Cross-validation of component models: A critical look at current methods," *Analytical and Bioanalytical Chemistry,* vol. 390, pp. 1241-1251, 2018.

[252] K. H. Esbensen and P. Geladi, "Principles of Proper Validation: use and abuse of re-sampling for validation," *Journal of Chemometrics,* vol. 24, pp. 168-187, 2010.

[253] F. Westad and F. Marini, "Validation of chemometric models e A tutorial," *Analytica Chimica Acta,* vol. 893, pp. 14-24, 2015.

[254] D. J. Krasznai, R. C. Hartley, H. M. Roy, P. Champagne and M. F. Cunningham, "Compositional analysis of lignocellulosic biomass: conventional methodologies and future outlook," *Critical Reviews in Biotechnology,* vol. 38, no. 2, pp. 199-217, 2018.

[255] A. Alzagameem, M. Bergs, X. T. Do, S. E. Klein, J. Rumpf, M. Larkins, Y. Monakhova, R. Pude and M. Schulze, "Low-Input Crops as Lignocellulosic Feedstock for Second-

Generation Biorefineries and the Potential of Chemometrics in Biomass Quality Control," *Applied Sciences,* vol. 9, p. 2252, 2019.

[256] J. S. Lupoi, S. Singh, B. A. Simmons and R. J. Henry, "Assessment of Lignocellulosic Biomass Using Analytical Spectroscopy: an Evolution to High-Throughput Techniques," *Bioenergy Research,* vol. 7, pp. 1-23, 2014.

[257] L. Xiao, H. Wei, M. E. Himmel, H. Jameel and S. S. Kelley, "NIR and Py-mbms coupled with multivariate data analysis as a high-throughput biomass characterization technique: a review," *Frontiers in Plant Science,* vol. 5, no. 388, 2014.

[258] J. Rumpf, X. T. Do, R. Burger, Y. Monakhova and M. Schulze, "Types of lignin, properties, and structural characterization techniques," in *Lignin-based Materials for Biomedical Applications*, Elsevier Inc., 2021, pp. 105-158.

[259] K. B. Bec, J. Grabska and C. W. Huck, "Biomolecular and bioanalytical applications of infrared spectroscopy - A review," *Analytica Chimica Acta,* vol. 1133, pp. 150-177, 2020.

[260] J. S. Lupoi, S. Singh, M. Davis, D. J. Lee, M. Shepherd, B. A. Simmons and R. J. Henry, "High-throughput prediction of eucalypt lignin syringyl/guaiacyl content using multivariate analysis: a comparison between mid-infrared, near-infrared, and Raman spectroscopies for model development," *Biotechnology for Biofuels,* vol. 7, p. 93, 2014.

[261] C. L. Morais, K. M. Lima, M. Singh and F. L. Martin, "Tutorial: multivariate classification for vibrational spectroscopy in biological samples," *Nature Protocols,* vol. 15, pp. 2143-2162, 2020.

[262] M. Hof and R. Machan, "Basics of Optical Spectroscopy," in *Handbook of Spectroscopy: Second, Enlarged Edition*, Wiley-VCH, 2014, pp. 31-38.

[263] A. E. Segneanu, I. Gozescu, A. Dabici, P. Sfirloaga and Z. Szabadai, "Organic Compounds FT-IR Spectroscopy," in *Macro to Nano Spectroscopy*, InTech, 2012, pp. 145-164.

[264] F. L. Martin, J. G. Kelly, V. Llabjani, P. L. Martin-Hirsch, I. I. Patel, J. Trevisan, N. J. Fullwood and M. J. Walsh, "Distinguishing cell types or populations based on the computational analysis of their infrared spectra," *Nature Protocol,* vol. 5, no. 11, p. 1748, 2010.

[265] C. L. Morais, M. Paraskevaidi, L. Cui, N. J. Fullwood, M. Isabelle, K. M. Lima, P. L. Martin-Hirsch, H. Sreedhar, J. Trevisan, M. J. Walsh, D. Zhang, Y.-G. Zhu and F. L. Martin, "Standardization of complex biologically derived spectrochemical datasets," *Nature Protocols,* vol. 14, pp. 1546-1577, 2019.

[266] M. J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S. W. Fogarty, N. J. Fullwood, K. A. Heys, C. Hughes, P. Lasch, P. L. Martin-Hirsch, B. Obinaju, G. D. Sockalingum, J. Sulé-Suso, R. J. Strong, M. J. Walsh, B. R. Wood, P. Gardner and F. L. Martin, "Using Fourier transform IR spectroscopy to analyze biological materials," *Nature Protocols,* vol. 9, no. 8, p. 1771, 2014.

[267] M. Diem, "Raman Spectroscopy," in *Modern Vibrational Spectroscopy and Micro-Spectroscopy: Theory, Instrumentation and Biomedical Applications*, John Wiley & Sons, Ltd, 2015, pp. 103-122.

[268] K. Hashimoto, V. R. Badarla, A. Kawai and T. Ideguchi, "Complementary vibrational spectroscopy," *nature communications,* vol. 10, p. 4411, 2019.

[269] M. Diem, "Vibrational Microspectroscopy (MSP)," in *Modern Vibrational Spectroscopy and Micro-Spectroscopy*, John Wiley & Sons, Ltd, 2015, pp. 235-250.

[270] M. Diem, "Data Preprocessing and Data Processing in Microspectral Analysis," in *Modern Vibrational Spectroscopy and Micro-Spectroscopy: Theory, Instrumentation, and Biomedical Applications*, John Wiley & Sons, Ltd, 2015, pp. 251-282.

[271] P. Geladi, H. Grahn and M. Manley, "Data Analysis and Chemometrics for Hyperspectral Imaging," in *Raman, Infrared, and Near-Infrared Chemical Imaging*, John Wiley & Sons, Inc., 2010, pp. 93-107.

[272] N. Gierlinger, "New insights into plant cell walls by vibrational microspectroscopy," *Applied Spectroscopy Reviews,* vol. 53, no. 7, pp. 517-551, 2018.

[273] J. Zhu, H. Wang, F. Guo, L. Salmen and Y. Yu, "Cell wall polymer distribution in bamboo visualized with in situ imaging FTIR," *Carbohydrate Polymers,* vol. 274, p. 118653, 2021.

[274] M. Schwanninger, J. C. Rodrigues and K. Fackler, "A review of band assignments in near infrared spectra of wood and wood components," *Journal of Near Infrared Spectroscopy,* vol. 19, pp. 287-308, 2011.

[275] T. Theophile, Infrared spectroscopy: Materials science, engineering and technology, IntechOpen, 2012.

[276] S.-Y. Yang, Y. Han, Y.-S. Chang, J.-H. Park, Y. Park, H. Chung and H. Yeo, "Classification of the hot air heat treatment degree of larch wood using multivariate analysis of near-infrared spectroscopy," *Journal of Wood Science,* vol. 64, pp. 220-225, 2018.

[277] C.-L. So, B. K. Via, L. H. Groom, L. R. Schimleck, T. F. Shupe, S. S. Kelley and T. G. Rials, "Near infrared spectroscopy in the forest products industry," *Forest Products Journal,* vol. 54, no. 3, pp. 6-16, 2004.

[278] S. Tsuchikawa and H. Kobori, "A review of recent application of near infrared spectroscopy to wood science and technology," *Journal of Wood Science,* vol. 61, pp. 213-220, 2015.

[279] J. Sandak, A. Sandak and R. Meder, "Assessing trees, wood and derived products with near infrared spectroscopy: hints and tips," *Journal of Near Infrared Spectroscopy,* vol. 24, pp. 485-505, 2016.

[280] L. Wallbacks, U. Edlund and B. Norden, "Multivariate characterization of pulp using solid-state 13C NMR, FTIR, and NIR," *Tappi Journal,* pp. 201-206, 1991.

[281] F. S. Poke, J. K. Wright and C. A. Raymond, "Predicting Extractives and Lignin Contents in Eucalyptus globulus Using Near Infrared Reflectance Analysis," *Journal of Wood Chemistry and Technology,* vol. 2004, no. 1, pp. 55-67, 2004.

[282] D. J. Stackpole, R. E. Vaillancourt, A. Alves, J. Rodrigues and B. M. Potts, "Genetic Variation in the Chemical Components of Eucalyptus globulus Wood," *Genes | Genomes | Genetics,* vol. 1, no. 2, pp. 151-159, 2011.

[283] A. Alves, M. Schwanninger, H. Pereira and J. Rodrigues, "Calibration of NIR to assess lignin composition (H/G ratio) in maritime pine wood using analytical pyrolysis as the reference method," *Holzforschung,* vol. 60, pp. 29-31, 2006.

[284] A. Alves, A. Santos, D. d. S. Perez, J. Rodrigues, H. Pereira, R. Simoes and M. Schwanninger, "NIR PLSR model selection for Kappa number prediction of maritime pine Kraft pulps," *Wood science and technology,* vol. 41, pp. 491-499, 2007.

[285] A. Alves, A. Santos, P. Rozenberg, L. E. Paques, J.-P. Charpentier, M. Schwanninger and J. Rodrigues, "A common near infrared—based partial least squares regression model for the prediction of wood density of Pinus pinaster and Larix x eurolepis," *Wood Science and Technology,* vol. 46, pp. 157-175, 2012.

[286] A. Alves, R. Simoes, D. J. Stackpole, R. E. Vaillancourt, B. M. Potts, M. Schwanninger and J. Rodrigues, "Determination of the syringyl/guaiacyl ratio of Eucalyptus globulus wood lignin by near infrared-based partial least squares regression models using analytical pyrolysis as the reference method," *Journal of Near Infrared Spectroscopy,* vol. 19, pp. 343-348, 2011.

[287] A. M. Alves, R. F. Simoes, C. A. Santos, B. M. Potts, J. Rodrigues and M. Schwanninger, "Determination of Eucalyptus globulus wood extractives content by near infrared-based partial least squares regression models: comparison between extraction procedures," *Journal of Near Infrared Spectroscopy,* vol. 20, pp. 275-285, 2012.

[288] M. Schwanninger, J. C. Rodrigues, N. Gierlinger and B. Hinterstoisser, "Determination of lignin content in Norway spruce wood by Fourier transformed near infrared spectroscopy and partial least infrared spectroscopy and partial least squares regression. Part 1:

Wavenumber selection and evaluation of the selected range," *Journal of Near Infrared Spectroscopy,* vol. 19, pp. 319-329, 2011.

[289] M. Schwanninger, J. C. Rodrigues, N. Gierlinger and B. Hinterstoisser, "Determination of lignin content in Norway spruce wood by Fourier transformed near infrared spectroscopy and partial least squares regression analysis. Part 2: Development and evaluation of the final model," *Journal of Near Infrared Spectroscopy,* vol. 19, pp. 331-341, 2011.

[290] Y. Horikawa, T. Imai, R. Takada, T. Watanabe, K. Takabe, Y. Kobayashi and J. Sugiyama, "Near-Infrared Chemometric Approach to Exhaustive Analysis of Rice Straw Pretreated for Bioethanol Conversion," *Applied Biochemistry and Biotechnology,* vol. 164, pp. 194-203, 2011.

[291] Y. Horikawa, T. Imai, R. Takada, T. Watanabe, K. Takabe, Y. Kobayashi and J. Sugiyama, "Chemometric Analysis with Near-Infrared Spectroscopy for Chemically Pretreated Erianthus toward Efficient Bioethanol Production," *Applied Biochemistry and Biotechnology,* vol. 166, pp. 711-721, 2012.

[292] B. K. Via, C. Zhou, G. Acquah, W. Jiang and L. Eckhardt, "Near Infrared Spectroscopy Calibration for Wood Chemistry: Which Chemometric Technique Is Best for Prediction and Interpretation?," *Sensors,* vol. 14, no. 8, pp. 13532-13547, 2014.

[293] N. J. Kohan, B. K. Via and S. E. Taylor, "Prediction of Strand Feedstock Mechanical Properties with Near Infrared Spectroscopy," *BioResources,* vol. 7, no. 3, pp. 2996-3007, 2012.

[294] S. Tsuchikawa, Y. Hirashima, Y. Sasaki and K. Ando, "Near-Infrared Spectroscopic Study of the Physical and Mechanical Properties of Wood with Meso- and Micro-Scale Anatomical Observation," *Applied Spectroscopy,* vol. 59, no. 1, pp. 86-93, 2005.

[295] P. Hoffmeyer and J. G. Pedersen, "Evaluation of density and strength of Norway spruce by near infrared reflectance spectroscopy," *Holz als Roh-und werkstoff,* vol. 53, no. 3, pp. 165-170, 1995.

[296] C. Krongtaew, K. Messner, T. Ters and K. Fackler, "Characterization of Key Parameters for Biotechnological Lignocellulose Conversion Assessed by FT-NIR Spectroscopy. Part I: Qualitative Analysis of Pretreated Straw," *BioResources,* vol. 5, no. 4, pp. 2063-2080, 2010.

[297] L. Liang, L. Wei, G. Fang, F. Xu, Y. Deng, K. Shen, Q. Tian, T. Wu and B. Zhu, "Prediction of holocellulose and lignin content of pulp wood feedstock using near infrared spectroscopy and variable selection," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* vol. 225, p. 117515, 2020.

[298] H. Yang, Y. Liu, Z. Xiong and L. Liang, "Rapid Determination of Holocellulose and Lignin in Wood by Near Infrared Spectroscopy and Kernel Extreme Learning Machine," *Analytical Letters,* vol. 53, no. 7, pp. 1140-1154, 2020.

[299] E. J. Wolfrum, C. Payne, A. Schwartz, J. Jacobs and R. W. Kressin, "A Performance Comparison of Low-Cost Near-Infrared (NIR) Spectrometers to a Conventional Laboratory Spectrometer for Rapid Biomass Compositional Analysis," *BioEnergy Research,* vol. 13, pp. 1121-1129, 2020.

[300] X. Sheng, Z. Li, Z. Li, J. Dong, J. Wang and J. Yin, "Nondestructive determination of lignin content in Korla fragrant pear based on near-infrared spectroscopy," *Spectroscopy Letters,* vol. 53, no. 4, pp. 306-314, 2020.

[301] X. Wu, G. Li, X. Liu and F. He, "Rapid non-destructive analysis of lignin using NIR spectroscopy and chemo-metrics," *Food and Energy Security,* p. e289, 2021.

[302] O. Elle, R. Richter, M. Vohland and A. Weigelt, "Fine root lignin content is well predictable with near-infrared spectroscopy," *Scientific Reports,* vol. 9, no. 6396, 2019.

[303] Y. Tamaki and G. Mazza, "Rapid Determination of Lignin Content of Straw Using Forier Transform Mid-Infrared Spectroscopy," *Journal of agricultural and food chemistry,* vol. 59, no. 2, pp. 504-512, 2011.

[304] G. Bekiaris, C. Peltre, S. T. Barsberg, S. Bruun, K. M. Sorensen, S. B. Engelsen, J. Magid, M. Hansen and L. S. Jensen, "Three different Fourier-transform mid-infrared sampling techniques to characterize bio-organic samples," *Journal of Environmental Quality,* vol. 49, pp. 1310-1321, 2020.

[305] T. P. Schultz, M. C. Templeton and G. D. McGinnis, "Rapid Determination of Lignocellulose by Diffuse Reflectance Fourier Transform Infrared Spectrometry," *Analytical Chemistry,* vol. 57, pp. 2867-2869, 1985.

[306] H. Chen, C. Ferrari, M. Angiuli, J. Yao, C. Raspi and E. Bramanti, "Qualitative and quantitative analysis of wood samples by Fourier transform infrared spectroscopy and multivariate analysis," *Carbohydrate Polymers,* vol. 82, pp. 772-778, 2010.

[307] P. J. Van Soest and R. H. Wine, "Use of detergents in the analysis of fibrous feeds. IV. Determination of plant cell-wall constituents," *Journal of the Association of Official Analytical Chemists,* vol. 50, pp. 50-55, 1967.

[308] C. G. Boeriu, D. Bravo, R. J. Gosselink and J. E. van Dam, "Characterisation of structure-dependent functional properties of lignin with infrared spectroscopy," *Industrial Crops and Products an International Journal,* vol. 20, pp. 205-218, 2004.

[309] C. S. Lancefield, S. Constant, P. d. Peinder and P. C. Bruijnincx, "Linkage Abundance and Molecular Weight Characteristics of Technical Lignins by Attenuated Total Reflection-FTIR Spectroscopy Combined with Multivariate Analysis," *ChemSusChem,* vol. 12, no. 6, pp. 1139-1146, 2019.

[310] K. N. Khalili, P. de Peinder, J. Donkers, R. J. Gosselink, P. C. Bruinjnincx and B. M. Weckhuysen, "Monitoring Molecular Weight Changes during Technical Lignin Depolymerization by Operando Attenuated Total Reflectance Infrared Spectroscopy and Chemometrics," *ChemSusChem,* vol. 14, pp. 5517-5524, 2021.

[311] K. N. Khalili, P. de Peinder, P. C. Bruijnincx and B. M. Weckhuysen, "Monitoring Aqueous Phase Reactions by Operando ATR-IR Spectroscopy at High Temperature and Pressure: A Biomass Conversion Showcase," *Chemistry—Methods,* vol. 1, pp. 468-476, 2021.

[312] O. Derkacheva and D. Sukhov, "Investigation of Lignins by FTIR spectroscopy," *Macromolecular Symposia,* vol. 265, pp. 61-68, 2008.

[313] L. M. Kline, D. G. Hayes, A. R. Womac and N. Labbe, "Simplified determination of lignin content in hard and soft woods via UV-spectrophotometric analysis of biomass dissolved in ionic liquids," *Bio Resources,* vol. 5, no. 3, pp. 1366-1383, 2010.

[314] R. Herrera-Diaz, L. Bodi-Paul, O. Gordobil and J. Labidi, "Fast methods for the identification of suitable chemo-enzymatic treatments of Kraft lignin to obtain aromatic compounds," *Biofuels, Bioproducts & Biorefining,* vol. 14, pp. 521-532, 2020.

[315] U. P. Agarwal, "Analysis of Cellulose and Lignocellulose Materials by Raman Spectroscopy: A Review of the Current Status," *Molecules,* vol. 24, p. 1659, 2019.

[316] U. P. Agarwal and R. H. Atalla, "Raman Spectroscopy," in *Surface Analysis of Paper*, CRC Press, 2020, pp. 152-181.

[317] U. P. Agarwal, "1064 nm FT-Raman spectroscopy for investigations of plant cell walls and other biomass materials," *Frontiers in Plant Science,* vol. 5, no. 490, 2014.

[318] B. Prats-Mateu, M. Felhofer, A. de Juan and N. Gierlinger, "Multivariate unmixing approaches on Raman images of plant cell walls: new insights or overinterpretation of results?," *Plant Methods,* vol. 14, no. 52, 2018.

[319] U. P. Agarwal, "Raman imaging to investigate ultrastructure and composition of plant cell walls: distribution of lignin and cellulose in black spruce wood (Picea mariana)," *Planta,* vol. 224, pp. 1141-1153, 2006.

[320] H. J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, M. J. Walsh, M. R. McAinsh, N. Stone and F.

L. Martin, "Using Raman spectroscopy to characterize biological materials," *Nature Protocols,* vol. 11, pp. 664-687, 2016.

[321] T. Ona, T. Sonoda, K. Ito, M. Shibata, T. Katayama, T. Kato and Y. Ootake, "Non-Destructive Determination of Lignin Syringyl Quaiacyl Monomeric Composition in Native Wood by Fourier Transform Raman Spectroscopy," *Journal of Wood Chemistry and Technology,* vol. 18, no. 1, pp. 43-51, 1998.

[322] T. Ona, T. Sonoda, J. Ohshima, S. Yokota and N. Yoshizawa, "A Rapid Quantitative Method to Assess Eucalyptus Wood Properties for Kraft Pulp Production by FT-Raman Spectroscopy," *Journal of Pulp and Paper Science,* vol. 29, no. 1, pp. 6-10, 2003.

[323] A.-M. Saariaho, A.-S. Jaaskelainen, M. Nuopponen and T. Vuorinen, "Ultra Violet Resonance Raman Spectroscopy in Lignin Analysis: Determination of Characteristic Vibrations of p-Hydroxyphenyl, Guaiacyl, and Syringyl Lignin Structures," *Applied Spectroscopy,* vol. 57, pp. 58-66, 2003.

[324] A.-M. Saariaho, D. S. Argyropoulos, A.-S. Jaaskelainen and T. Vuorinen, "Development of the partial least squares models for the interpretation of the UV resonance Raman spectra of lignin model compounds," *Vibrational Spectroscopy,* vol. 37, pp. 111-121, 2005.

[325] M. W. Meyer, J. S. Lupoi and E. A. Smith, "1064 nm dispersive multichannel Raman spectroscopy for the analysis of plant lignin," *Analytica Chimica Acta,* vol. 706, pp. 164-170, 2011.

[326] J. S. Lupoi and E. A. Smith, "Characterization of Woody and Herbaceous Biomasses Lignin Composition with 1064 nm Dispersive Multichannel Raman Spectroscopy," *Applied Spectroscopy,* vol. 66, no. 8, pp. 903-910, 2012.

[327] J. S. Lupoi, A. Healey, S. Singh, R. Sykes, M. Davis, D. J. Lee, M. Shepherd, B. A. Simmons and R. J. Henry, "High-Throughput Prediction of Acacia and Eucalypt Lignin Syringyl/Guaiacyl Content Using FT-Raman Spectroscopy and Partial Least Squares Modeling," *Bioenergy Resources,* vol. 8, pp. 953-963, 2015.

[328] P. N. Perera, M. Schmidt, V. L. Chiang, P. J. Schuck and P. D. Adams, "Raman-spectroscopy-based noninvasive microanalysis of native lignin structure," *Analytical and Bioanalytical Chemistry,* vol. 402, pp. 983-987, 2012.

[329] N. Gierlinger, "Revealing changes in molecular composition of plant cell walls on the micron-level by Raman mapping and vertex component analysis (VCA)," *Frontiers in Plant Science,* vol. 5, no. 306, 2014.

[330] C. J. Colares, T. C. Pastore, V. T. Coradin, J. A. Camargos, A. C. Moreira, J. C. Rubim and J. W. Braga, "Exploratory Analysis of the Distribution of Lignin and Cellulose in Woods by Raman Imaging and Chemometrics," *Journal of the Brazilian Chemical Society,* vol. 26, no. 6, pp. 1297-1305, 2015.

[331] B. Liu, P. Wang, J. I. Kim, D. Zhang, Y. Xia, C. Chapple and J.-X. Cheng, "Vibrational Fingerprint Mapping Reveals Spatial Distribution of Functional Groups of Lignin in Plant Cell Wall," *Analytical Chemistry,* vol. 87, no. 18, pp. 9436-9442, 2015.

[332] B. Prats-Mateu, M. T. Hauser, A. Heredia and N. Gierlinger, "Waterproofing in Arabidopsis: Following Phenolics and Lipids In situ by Confocal Raman Microscopy," *Frontiers in Chemistry,* vol. 4, no. 10, 2016.

[333] X. Zhang, "Visualising lignin quantitatively in plant cell walls by micro-Raman spectroscopy," *RSC Advances,* vol. 11, p. 131124, 2021.

[334] E. Widjaja and M. Garland, "Use of Raman Microscopy and Band-Target Entropy Minimization Analysis To Identify Dyes in a Commercial Stamp. Implications for Authentication and Counterfeit Detection," *Analytical Chemistry,* vol. 80, pp. 729-733, 2008.

[335] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, "Optimization by Simulated Annealing," *Science,* vol. 220, no. 4598, pp. 671-680, 1983.

[336] W. Gao, T. Shu, Q. Liu, S. Ling, Y. Guan, S. Liu and L. Zhou, "Predictive Modeling of Lignin Content for the Screening of Suitable Poplar Genotypes Based on Fourier Transform−Raman Spectrometry," *ACS Omega,* vol. 6, pp. 8578-8587, 2021.

[337] W. Gao, T. Shu, Y. Guan, S. Ling, S. Liu and L. Zhou, "Novel strategy for establishment of an FT-Raman spectroscopy based quantitative model for poplar holocellulose content determination," *Carbohydrate Polymers,* vol. 277, p. 118793, 2022.

[338] W. Gao, L. Zhou, Y. Guan, H. Gao and S. Liu, "Monitoring the kappa number of bleached pulps based on FT-Raman spectroscopy," *Cellulose,* vol. 29, pp. 1069-1080, 2022.

[339] D. F. Cipriano, L. Chinelatto Jr., S. A. Nascimento, C. A. Rezende, S. M. de Menezes and J. C. Freitas, "Potential and limitations of 13C CP/MAS NMR spectroscopy to determine the lignin content of lignocellulosic feedstock," *Biomass and Bioenergy,* vol. 142, p. 105792, 2020.

[340] A. V. Faleva, A. Y. Kozhevnikov, S. A. Pokryshkin, D. I. Falev, S. L. Shestakov and J. A. Popova, "Structural characteristics of different softwood lignins according to 1D and 2D NMR spectroscopy," *Journal of Wood Chemistry and Technology,* vol. 40, no. 3, pp. 178-189, 2020.

[341] J.-L. Wen, S.-L. Sun, B.-L. Xue and R.-C. Sun, "Recent Advances in Characterization of Lignin Polymer by Solution-State Nuclear Magnetic Resonance (NMR) Methodology," *Materials,* vol. 6, pp. 359-391, 2013.

[342] S. Y. Lin and C. W. Dence, Methods in Lignin Chemistry, Springer Science & Business Media, 2012.

[343] R. M. Happs, B. Addison, C. Doeppke, B. S. Donohoe, M. F. Davis and A. E. Harman-Ware, "Comparison of methodologies used to determine aromatic lignin unit ratios in lignocellulosic biomass," *Biotechnology for Biofuels,* vol. 14, no. 58, 2021.

[344] H. Winning, F. H. Larsen, R. Bro and S. B. Engelsen, "Quantitative analysis of NMR spectra with chemometrics," *Journal of Magnetic Resonance,* vol. 190, pp. 26-32, 2008.

[345] T. M. Alam and M. K. Alam, "Chemometric Analysis of NMR Spectroscopy Data: A Review," *Annual Reports on NMR Spectroscopy,* vol. 54, pp. 42-77, 2005.

[346] S. B. Engelsen, F. Savorani and M. A. Rasmussen, "Chemometric Exploration of Quantitative NMR Data," *eMagRes,* vol. 2, pp. 267-278, 2013.

[347] D. Stanisic, L. G. Martins and L. Tasic, "Nuclear Magnetic Resonance Spectroscopy in Analyses of Biological Samples," in *Tools and Trends in Bioanalytical Chemistry*, Springer, 2021, pp. 203-221.

[348] M. Hedenstrom, S. Wiklund-Lindstrom, T. Oman, F. Lu, L. Gerber, P. Schatz, B. Sundberg and J. Ralph, "Identification of Lignin and Polysaccharide Modifications in Populus Wood by Chemometric Analysis of 2D NMR Spectra from Dissolved Cell Walls," *Molecular Plant,* vol. 2, no. 5, pp. 933-942, 2009.

[349] C. G. Boeriu, F. I. Fitigau, R. J. Gosselink, A. E. Frissen, J. Stoutesdijk and F. Peter, "Fractionation of five technical lignins by selective extraction in green solvents and characterisation of isolated fractions," *Industrial Crops and Products,* vol. 62, pp. 481-490, 2014.

[350] L. M. Aguilera-Saez, F. M. Arrabal-Campos, A. J. Callejon-Ferre, M. D. S. Medina and I. Fernandez, "Use of multivariate NMR analysis in the content prediction of hemicellulose, cellulose and lignin in greenhouse crop residues," *Phytochemistry,* vol. 158, pp. 110-119, 2019.

[351] M. Jalali-Hevari, S. Masoum and P. Shahbazikhah, "Simulation of 13C nuclear magnetic resonance spectra of lignin compounds using principal component analysis and artificial neural networks," *Journal of Magnetic Resonance,* vol. 171, no. 1, pp. 176-185, 2004.

[352] R. Burger, J. Rumpf, X. T. Do, Y. B. Monakhova, B. W. Diehl, M. Rehahn and M. Schuzle, "Is NMR Combined with Multivariate Regression Applicable for the Molecular Weight

Determination of Randomly Cross-Linked Polymers Such as Lignin?," *ACS Omega,* vol. 6, pp. 29516-29524, 2021.

[353] J. J. Workman Jr, "A Review of Calibration Transfer Practices and Instrument Differences in Spectroscopy," *Applied Spectroscopy,* vol. 72, no. 3, pp. 340-365, 2018.

[354] Y. Wang, D. J. Vetkamp and B. R. Kowalski, "Multivariate Instrument Standardization," *Analytical Chemistry,* vol. 63, pp. 2750-2756, 1991.

[355] S. Lindner, R. Burger, D. N. Rutledge, X. T. Do, J. Rumpf, W. K. Diehl, M. Schulze and Y. B. Monakhova, "Is the Calibration Transfer of Multivariate Calibration Models between High- and Low-Field NMR Instruments Possible? A Case Study of Lignin Molecular Weight," *Analytical Chemistry,* vol. 94, pp. 3997-4004, 2022.

[356] R. Burger, S. Lindner, J. Rumpf, X. T. Do, B. W. Diehl, M. Rehahn, Y. B. Monakhova and M. Schulze, "Benchtop versus high field NMR: Comparable performance found for the molecular weight determination of lignin," *Journal of Pharmaceutical and Biomedical Analysis,* vol. 212, p. 114649, 2022.

[357] E. De Hoffmann, Mass Spectrometry, Kirk-Othmer Encyclopedia of Chemical Technology, 2000.

[358] J. H. Gross, "Introduction," in *Mass Spectrometry: A textbook*, Springer Science & Business Media, 2006, pp. 1-10.

[359] R. Zhang, Y. Qi, C. Ma, J. Ge, Q. Hu, F.-J. Yue, S.-L. Li and D. A. Volmer, "Characterization of Lignin Compounds at the Molecular Level: Mass Spectrometry Analysis and Raw Data Processing," *Molecules,* vol. 26, no. 178, 2021.

[360] M. Kleen and G. Lindblad, "Quantification of lignin and carbohydrates in kraft pulps using analytical pyrolysis and multivariate data analysis," *Journal of Analytical and Applied Pyrolysis,* vol. 25, pp. 209-227, 1993.

[361] F. A. Agblevor, R. J. Evans and K. D. Johnson, "Molecular-beam mass-spectrometric analysis of lignocellulosic materials I. Herbaceous biomass," *Journal of Analytical and Applied Pyrolysis,* vol. 30, pp. 125-144, 1994.

[362] S. S. Kelley, J. Jellison and B. Goodell, "Use of NIR and pyrolysis-MBMS coupled with multivariate analysis for detecting the chemical changes associated with brown-rot biodegradation of spruce wood," *FEMS Microbiology Letters,* vol. 209, pp. 107-111, 2002.

[363] M. H. Alma and S. S. Kelley, "The application of pyrolysis-molecular beam mass spectrometry for characterization of barkphenolysis products," *Biomass and Bioenergy,* vol. 22, pp. 411-419, 2002.

[364] S. S. Kelley, R. M. Rowell, M. Davis, C. K. Jurich and R. Ibach, "Rapid analysis of the chemical composition of agricultural fibers using near infrared spectroscopy and pyrolysis molecular beam mass spectrometry," *Biomass and Bioenergy,* vol. 27, pp. 77-88, 2004.

[365] N. Labbe, T. G. Rials, S. S. Kelley, Z.-M. Cheng, J.-Y. Kim and Y. Li, "FT-IR imaging and pyrolysis-molecular beam mass spectrometry: new tools to investigate wood tissues," *Wood Sci Technol,* vol. 39, pp. 61-77, 2005.

[366] M. F. Davis, G. A. Tuskan, P. Payne, T. J. Tschaplinski and R. Meilan, "Assessment of Populus wood chemistry following the introduction of a Bt toxin gene," *Tree Physiology,* vol. 26, pp. 557-564, 2006.

[367] R. Sykes, B. Kodrzycki, G. Tuskan, K. Foutz and M. Davis, "Within tree variability of lignin composition in Populus," *Wood Science Technology,* vol. 42, pp. 649-661, 2008.

[368] R. Sykes, M. Yung, E. Novaes, M. Kirst, G. Peter and M. Davis, "High-Throughput Screening of Plant Cell-Wall Composition Using Pyrolysis Molecular Beam Mass Spectroscopy," in *Jonathan R. Mielenz (ed.), Biofuels: Methods and Protocols, Methods in Molecular Biology, vol. 581*, Humana Press, a part of Springer Science+Business Media, 2009, pp. 169-183.

[369] D. G. Mann, N. Labbe, R. W. Sykes, K. Gracom, L. Kline, I. M. Swamidoss, J. N. Burris, M. Davis and C. N. Stewart Jr., "Rapid Assessment of Lignin Content and Structure in Switchgrass (Panicum virgatum L.) Grown Under Different Environmental Conditions," *Bioenergy Resources,* vol. 2, pp. 246-256, 2009.

[370] S. R. Decker, R. W. Sykes, G. B. Turner, J. S. Lupoi, C. Doepkke, M. P. Tucker, L. A. Schuster, K. Massa, M. E. Himmel, M. F. Davis and E. Gjersing, "High-throughput Screening of Recalcitrance Variations in Lignocellulosic Biomass: Total Lignin, Lignin Monomers, and Enzymatic Sugar Release," *Journal of Visualized Experiments,* vol. 103, 2015.

[371] C. W. Edmunds, P. Peralta, S. S. Kelley, V. L. Chiang, R. R. Sharma-Shivappa, M. F. Davis, A. E. Harman-Ware, R. W. Sykes, E. Gjersing, M. W. Cunningham, W. Rottmann, Z. D. Miller and I. Peszlen, "Characterization and enzymatic hydrolysis of wood from transgenic Pinus taeda engineered with syringyl lignin or reduced lignin content," *Cellulose,* vol. 24, pp. 1901-1914, 2017.

[372] S. R. Decker, A. Harman-Ware, R. M. Happs, E. J. Wolfrum, G. A. Tuskan, D. Kainer, G. B. Oguntimein, M. Rodriguez, D. Weighill, P. Jones and D. Jacobson, "High Throughput Screening Technologies in Biomass Characterization," *Frontiers in Energy Research,* vol. 6, no. 120, 2018.

[373] L. Gerber, M. Eliasson, J. Trygg, T. Moritz and B. Sundberg, "Multivariate curve resolution provides a high-throughput data processing pipeline for pyrolysis-gas chromatography/mass spectrometry," *Journal of Analytical and Applied Pyrolysis,* vol. 95, pp. 95-100, 2012.

[374] M. A. Galetta, L. Reina, F. Resquin, C. Mantero, J. A. González-Pérez, G. Almendros, M. P. Menéndez and F. J. González-Vila, "Chemometric appraisal of lignin pyrolytic assemblages from Eucalyptus woods relevant for pulping in Uruguay," *Journal of Analytical and Applied Pyrolysis,* vol. 109, pp. 296-303, 2014.

[375] H. E. Toraman, R. Vanholme, E. Borén, Y. Vanwonterghem, M. R. Djokic, G. Yildiz, F. Ronsse, W. Prins, W. Boerjan, K. M. Van Geem and G. B. Marin, "Potential of genetically engineered hybrid poplar for pyrolytic production of bio-based phenolic compounds," *Bioresource Technology,* vol. 207, pp. 229-236, 2016.

[376] H. E. Toraman, V. Abrahamsson, R. Vanholme, R. Van Acker, F. Ronsse, G. Pilate, W. Boerjan, K. M. Van Geem and G. B. Marin, "Application of Py-GC/MS coupled with PARAFAC2 and PLS-DA to study fast pyrolysis of genetically engineered poplars," *Journal of Analytical and Applied Pyrolysis,* vol. 129, pp. 101-111, 2018.

[377] G. SriBala, H. E. Toraman, S. Symoens, A. Dejardin, G. Pilate, W. Boerjan, F. Ronsse, K. M. Van Geem and G. B. Marin, "Analytical Py-GC/MS of Genetically Modified Poplar for the Increased Production of Bio-aromatics," *Computational and Structural Biotechnology Journal,* vol. 17, pp. 599-610, 2019.

[378] L. G. Johnsen, P. B. Skou, B. Khakimov and R. Bro, "Gas chromatography - mass spectrometry data processing made easy," *Journal of Chromatography A,* vol. 1503, pp. 57-64, 2017.

[379] J. M. Amigo, T. Skov, J. Coello, S. Maspoch and R. Bro, "Solving GC-MS problems with PARAFAC2," *Trends in Analytical Chemistry,* vol. 27, no. 8, pp. 714-725, 2008.

[380] G. Baccolo, B. Quintanilla-Casas, S. Vichi, D. Augustijn and B. Rasmus, "From untargeted chemical profiling to peak tableseA fully automatedAI driven approach to untargeted GC-MS," *Trends in Analytical Chemistry,* vol. 145, p. 116451, 2021.

[381] J. Reyes-Rivera, E. Solano, T. Terrazas, M. Soto-Hernandez, S. Arias, Y. C. Almanza-Arjona and L. A. Polindara-Garcia, "Classification of lignocellulosic matrix of spines in Cactaceae by Py-GC/MS combined with omic tools and multivariate analysis: A chemotaxonomic approach," *Journal of Analytical and Applied Pyrolysis,* vol. 148, p. 1047956, 2020.

[382] A. Demirbas, "Combustion characteristics of different biomass fuels," *Progress in Energy and Combustion Science,* vol. 30, pp. 219-230, 2004.

[383] A. A. Boateng, D. E. Daugaard, N. M. Goldberg and K. B. Hicks, "Bench-Scale Fluidized-Bed Pyrolysis of Switchgrass for Bio-Oil Production," *Industrial & Engineering Chemistry Research,* vol. 46, pp. 1891-1897, 2007.

[384] S. Wang, G. Dai, H. Yang and Z. Luo, "Progress in Energy and Combustion Science," *Lignocellulosic biomass pyrolysis mechanism: A state-of-the-art review,* vol. 62, pp. 33-86, 2017.

[385] M. Hu, Z. Chen, S. Wang, D. Guo, C. Ma, Y. Zhou, J. Chen, M. Laghari, S. Fazal, B. Xiao, B. Zhang and S. Ma, "Thermogravimetric kinetics of lignocellulosic biomass slow pyrolysis using distributed activation energy model, Fraser–Suzuki deconvolution and iso-conversional method," *Energy Conversion and Management,* vol. 118, pp. 1-11, 2016.

[386] A. Perejon, P. E. Sanchez-Jimenez, J. M. Criado and L. A. Perez-Maqueda, "Kinetic Analysis of Complex Solid-State Reactions. Kinetic Analysis of Complex Solid-State Reactions.," *The Journal of Physical Chemistry B,* vol. 115, pp. 1780-1791, 2011.

[387] R. D. Fraser and E. Suzuki, "Resolution of Overlapping Absorption Bands by Least Squares Procedures," *Analytical Chemistry,* vol. 38, no. 12, pp. 1770-1773, 1966.

[388] R. D. Fraser and E. Suzuki, "Resolution of Overlapping Bands: Functions for Simulating Band Shapes," *Analytical Chemistry,* vol. 41, no. 1, pp. 37-39, 1969.

[389] Z. Cheng, W. Wu, P. Ji, X. Zhou, R. Liu and J. Cai, "Applicability of Fraser–Suzuki function in kinetic analysis of DAEM processes and lignocellulosic biomass pyrolysis processes," *Journal of Thermal Analysis and Calorimetry,* vol. 119, pp. 1429-1438, 2015.

[390] J. D. S. Castro, E. G. P. da Silva and C. F. Virgens, "Evaluation of models to predict the influence of chemical pretreatment on the peels of Nephelium lappaceum L. based on pyrolysis kinetic parameters obtained using a combined Fraser-Suzuki function and

Friedman's isoconversional method," *Journal of Analytical and Applied Pyrolysis,* vol. 149, p. 104827, 2020.

[391] C. F. das Virgens and J. D. Castro, "Screening of slow pyrolysis routes for maximum biofuel production from Syzygium malaccense biomass by TGA-FSD and chemometric tools," *Journal of Thermal Analysis and Calorimetry,* vol. 146, pp. 2005-2014, 2021.

[392] D. Carpenter, T. L. Westover, S. Czernik and W. Jablonski, "Biomass feedstocks for renewable fuel production: a review of the impacts of feedstock and pretreatment on the yield and product distribution pretreatment on the yield and product distribution," *Green Chemistry,* vol. 16, pp. 384-406, 2014.

[393] H. L. Friedman, "Kinetics of Thermal Degradation of Char-Forming Plastics from Thermogravimetry. Application to a Phenolic Plastic," *Journal of Polymer Science: Part C Polymer Symposia,* vol. 6, no. 1, pp. 183-195, 1964.

[394] J. M. Criado, J. Malek and A. Ortega, "Applicability of the Master Plots in Kinetic Analysis of Non-Isothermal Data," *Thermochimica Acta,* vol. 147, pp. 377-385, 1989.

[395] P. E. Sánchez-Jiménez, L. A. Pérez-Maqueda, A. Perejón and J. M. Criado, "Generalized master plots as a straightforward approach for determining the kinetic model: The case of cellulose pyrolysis," *Thermochimica Acta,* vol. 552, pp. 54-59, 2013.

[396] H. Khaliliyan, C. Schuster, I. Sumerskii, M. Guggenberger, J. T. Oberlerchner, T. Rosenau, A. Potthast and B. Stefan, "Direct Quantification of Lignin in Liquors by High Performance Thin Layer Chromatography-Densitometry and Multivariate Calibration," *ACS Sustainable Chemistry & Engineering,* vol. 8, pp. 16766-16774, 2020.

[397] B. O. Myrvold and D. Pavlov, "Multivariate analysis for characterization of expanders," *Journal of Power Sources,* vol. 85, pp. 92-101, 2000.

[398] R. Huang, R. Su, W. Qi and Z. He, "Understanding the Key Factors for Enzymatic Conversion of Pretreated Lignocellulose by Partial Least Square Analysis," *Biotechnology progress,* vol. 26, no. 2, pp. 384-392, 2009.

[399] A. E. Ray, C. L. Williams, A. N. Hoover, C. Li, K. L. Sale, R. M. Emerson, J. Klinger, E. Oksen, A. Narani, J. Yan, C. M. Beavers, D. Tanjore, M. Yunes, E. Bose, J. H. Leal, J. L. Bowen, E. J. Wolfrum, M. G. Resch, T. A. Semelsberger and B. S. Donohoe, "Multiscale Characterization of Lignocellulosic Biomass Variability and Its Implications to Preprocessing and Conversion: a Case Study for Corn Stover," *ACS Sustainable Chemistry & Engineering,* vol. 8, pp. 3218-3230, 2020.

[400] B. Li, H. Liu, H. Xu, B. Pang, H. Mou, H. Wang and X. Mu, "Characterization of the Detailed Relationships of the Key Variables in the Process of the Alkaline Sulfite Pretreatment of Corn Stover by Multivariate Analysis," *BioResources,* vol. 9, no. 2, pp. 2757-2771, 2014.

[401] H. Xu, X. Che, Y. Ding, Y. Kong, B. Li and W. Tian, "Effect of crystallinity on pretreatment and enzymatic hydrolysis of lignocellulosic biomass based on multivariate analysis," *Bioresource Technology,* vol. 279, pp. 271-280, 2019.

[402] H. Xu, Y. Kong, J. Peng, X. Song, X. Che, S. Liu and W. Tian, "Multivariate analysis of the process of deep eutectic solvent pretreatment of lignocellulosic biomass," *Industrial Crops & Products,* vol. 150, p. 112363, 2020.

[403] H. Xu, Y. Kong, J. Peng, X. Song, Y. Liu, Z. Su, B. Li, C. Gao and W. Tian, "Comprehensive analysis of important parameters of choline chloride-based deep eutectic solvent pretreatment of lignocellulosic biomass," *Bioresource Technology,* vol. 319, p. 124209, 2021.

[404] A. Brandt, J. Gräsvik, J. P. Hallett and T. Welton, "Deconstruction of lignocellulosic biomass with ionic liquids," *Green Chemistry,* vol. 15, pp. 550-583, 2013.

[405] H. Malaeke, M. R. Housaindokht, H. Monhemi and M. Izadyar, "Deep eutectic solvent as an efficient molecular liquid for lignin solubilization and wood delignification," *Journal of Molecular Liquids,* vol. 263, pp. 193-199, 2018.

[406] M. Karlsson, N. Giummarella, P. A. Linden and M. Lawoko, "Toward a Consolidated Lignin Biorefinery: Preserving the Lignin Structure through Additive-Free Protection Strategies," *ChemSusChem,* vol. 13, pp. 4666-4677, 2020.

[407] G. E. Box and D. W. Behnken, "Some New Three Level Designs for the Study of Quantitative Variables," *Technometrics,* vol. 2, no. 4, pp. 455-475, 1960.

[408] S. Rangarajan, A. Bhan and P. Daoutidis, "Rule-Based Generation of Thermochemical Routes to Biomass Conversion," *Industrial & Engineering Chemistry Research,* vol. 49, no. 21, pp. 10459-10470, 2010.

[409] Q. Li, G. Wittreich, Y. Wang, H. Bhattacharjee, U. Gupta and D. G. Vlachos, "Accurate Thermochemistry of Complex Lignin Structures via Density Functional Theory, Group Additivity, and Machine Learning," *ACS Sustainable Chemistry & Engineering,* vol. 9, pp. 3043-3049, 2021.

[410] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research,* vol. 12, pp. 2825-2830, 2011.

[411] J. Lym, G. R. Wittreich and D. G. Vlachos, "A Python Multiscale Thermochemistry Toolbox (pMuTT) for thermochemical and kinetic parameter estimation," *Computer Physics Communications,* vol. 247, p. 106864, 2020.

[412] N. Cohen and S. W. Benson, "Estimation of Heats of Formation of Organic Compounds by Additivity Methods," *Chemical Reviews,* vol. 93, pp. 2419-2438, 1993.

[413] F. Sattari, D. Tefera, K. Sivaramakrishnan, S. H. Mushrif and V. Prasad, "Chemoinformatic Investigation of the Chemistry of Cellulose and Lignin Derivatives in Hydrous Pyrolysis," *Industrial & Engineering Chemistry Research,* vol. 59, pp. 11582-11595, 2020.

[414] A. Rudin and P. Choi, "Mechanical Properties of Polymer Solids and Liquids," in *The Elements of Polymer Science & Engineering*, Academic Press, 2013, pp. 149-229.

[415] D. Vural, J. C. Smith and L. Petridis, "Dynamics of the lignin glass transition," *Physical Chemistry Chemical Physics,* vol. 20, pp. 20504-20512, 2018.

[416] J. D. S. Castro and C. F. d. Virgens, "Thermal decomposition of Nephelium lappaceum L. peel," *Journal of Thermal Analysis and Calorimetry,* vol. 138, pp. 3541-3549, 2019.

[417] P. Choi, "Molecular Modelling—an Enabling Technology for Chemical Engineers," *The Canadian Journal of Chemical Engineering,* vol. 84, pp. 265-268, 2006.

[418] F. J. Martin-Martinez, J. A. Dobado, A. Sanchez-Gonzalez, D. L. Barreiro and M. J. Barreiro, "Multiscale Modeling of Lignocellulosic Biomass," in *In: Andreoni W., Yip S. (eds) Handbook of Materials Modeling. Springer*, Springer, Cham, 2020, pp. 1628-1644.

[419] J. A. Poveda-Giraldo, J. C. Solarte-Toro and C. A. C. Alzate, "The potential use of lignin as a platform product in biorefineries: A review," *Renewable and Sustainable Energy Reviews,* vol. 138, p. 110688, 2021.

[420] R. L. Silveira, S. R. Stoyanov, S. Gusarov, M. S. Skaf and A. Kovalenko, "Supramolecular Interactions in Secondary Plant Cell Walls: Effect of Lignin Chemical Composition Revealed with the Molecular Theory of Solvation," *The Journal of Physical Chemistry Letters,* vol. 6, pp. 206-211, 2015.

[421] K. Houston, M. R. Tucker, J. Chowdhury, N. Shirley and A. Little, "The Plant Cell Wall: A Complex and Dynamic Structure As Revealed by the Responses of Genes under Stress Conditions," *Frontiers in Plant Science,* vol. 7, no. 984, 2016.

[422] O. M. Terrett, J. J. Lyczakowski, L. Yu, D. Luga, W. T. Franks, S. P. Brown, R. Dupree and P. Dupree, "Molecular architecture of softwood revealed by solid-state NMR," *Nature Communications,* vol. 10, p. 4978, 2019.

[423] N. D. Bonawitz and C. Chapple, "The Genetics of Lignin Biosynthesis: Connecting Genotype to Phenotype," *Annual Review of Genetics,* vol. 44, pp. 337-363, 2010.

[424] R. Vanholme, K. Morreel, C. Darrah, P. Oyarce, J. H. Grabber, J. Ralph and W. Boerjan, "Metabolic engineering of novel lignin in biomas crops," *New Phytologist,* vol. 196, pp. 978-1000, 2012.

[425] L. Dessbesell, M. Paleologou, M. Leitch, R. Pulkki and C. Xu, "Global lignin supply overview and kraft lignin potential as an alternative for petroleum-based polymers," *Renewable and Sustainable Energy Reviews,* vol. 123, p. 109768, 2020.

[426] A. Ekielski and P. K. Mishra, "Lignin for Bioeconomy: The Present and Future Role of Technical Lignin," *International Journal of Molecular Sciences,* vol. 22, no. 63, 2021.

[427] M. V. Galkin, D. D. Francesco, U. Edlund and J. S. Samec, "Sustainable sources need reliable standards," *Faraday Discussions,* vol. 202, pp. 281-301, 2017.

[428] J. E. Bidlack and W. V. Dashek, Plant Cells and their Organelles, John Wiley & Sons, Ltd., 2017.

[429] R. R. Sederoff, J. J. MacKay, J. Ralph and R. D. Hatfield, "Unexpected variation in lignin," *Current Opinion in Plant Biology,* vol. 2, pp. 145-152, 1999.

[430] W. Boerjan, J. Ralph and M. Baucher, "Lignin Biosynthesis," *Annual Review of Plant Biology,* vol. 54, pp. 519-546, 2003.

[431] R. Vanholme, B. De Meester, J. Ralph and W. Boerjan, "Lignin biosynthesis and its integration into metabolism," *Current Opinion in Biotechnology,* vol. 56, pp. 230-239, 2019.

[432] V. G. Lebedev and K. A. Shestibratov, "Genetic Engineering of Lignin Biosynthesis in Trees: Compromise between Wood Properties and Plant Viability," *Russian Journal of Plant Physiology,* vol. 68, no. 4, pp. 596-612, 2021.

[433] A. Chanoca, L. de Vries and W. Boerjan, "Lignin Engineering in Forest Trees," *Frontiers in Plant Science,* vol. 10, no. 912, 2019.

[434] R. Bhardwaj, N. Handa, R. Sharma, H. Kaur, S. Kohli, V. Kumar and P. Kaur, "Lignins and Abiotic Stress: An Overview," in *Physiological Mechanisms and Adaptation Strategies in Plants Under Changing Environment*, New York, NY, Springer, 2013, pp. 267-296.

[435] Q. Liu, L. Luo and L. Zheng, "Lignins: Biosynthesis and Biological Functions in Plants," *International Journal of Molecular Sciences,* vol. 19, p. 335, 2018.

[436] D. Vincent, C. Lapierre, B. Pollet, G. Cornic, L. Negroni and M. Zivy, "Water Deficits Affect Caffeate O-Methyltransferase, Lignification, and Related Enzymes in Maize Leaves. A Proteomic Investigation," *Plant Physiology,* vol. 137, pp. 949-960, 2005.

[437] J. L. Farrar, Trees in Canada, Canadian Forest Services., Natural Resources Canada. Co-published by Fizhenry & Whitesite Ltd. and the Canadian Forest Service, 1995.

[438] "Tree Species Compendium Index: Douglas-fir," [Online]. Available: https://www2.gov.bc.ca/gov/content/industry/forestry/managing-our-forest-resources/silviculture/tree-species-selection/tree-species-compendium-index/douglas-fir. [Accessed 06 09 2021].

[439] R. M. Rowell, R. Pettersen and M. A. Tshabalala, "Cell Wall," in *Handbook of Wood Chemistry and Wood Composites*, CRC PRess, 2012.

[440] A. Wagner, Y. Tobimatsu, L. Phillips, H. Flint, B. Geddes, F. Lu and J. Ralph, "Syringyl lignin production in conifers: proof of concept in a Pine tracheary element system," *PNAS,* vol. 112, no. 19, pp. 6218-6223, 2015.

[441] H.-M. Chang and X. Jiang, "Biphenyl structure and its impact on the macromolecular structure of lignin: A critical review," *Journal of Wood Chemistry and Technology,* vol. 40, no. 2, pp. 81-90, 2020.

[442] P. J. Van Soest, J. B. Robertson and B. A. Lewis, "Methods for Dietary Fiber, Neutral Detergent Fiber, and Nonstarch Polysaccharides In Relation to Animal Nutrition," *Journal of Dairy Science,* vol. 74, pp. 3583-3597, 1991.

[443] M. P. da Rosa, P. H. Beck, D. G. Müller, J. B. Moreira, J. S. da Silva and A. M. M. Durigon, "Extraction of Organosolv Lignin from Rice Husk under Reflux Conditions," *Biological and Chemical Research,* pp. 87-98, 2017.

[444] R. L. Johnson and K. Schmidt-Rohr, "Quantitative solid-state 13C NMR with signal enhancement by multiple cross polarization," *Journal of Magnetic Resonance,* vol. 239, pp. 44-49, 2014.

[445] A. E. Bennett, C. M. Rienstra, M. Auger, K. V. Lakshmi and R. G. Griffin, "Heteronuclear decouplingin rotating solids," *The Journal of Chemical Physics,* vol. 103, pp. 6951-6958, 1995.

[446] Y. Wu, "Intra-increment lignin content of five western Canadian coniferous woods," University of British Columbia, Retrieved from https://open.library.ubc.ca/collections/ubctheses/831/items/1.0105472, 1964.

[447] K. R. Oleson, "Extractives of Douglas-fir and Their Implications for Biofuel Saccharification," University of Washington, 2018.

[448] A. Tejado, C. Pena, J. Labidi, M. J. Echeverria and I. Mondragon, "Physico-chemical characterization of lignins from different sources for use in phenol–formaldehyde resin synthesis," *Bioresource Technology,* vol. 98, no. 8, pp. 1655-1663, 2007.

[449] T. You and F. Xu, "Applications of Molecular Spectroscopic Methods to the Elucidation of Lignin Structure," in *Applications of Molecular Spectroscopy to Current Research in the Chemical and Biological Sciences*, London, United Kingdom, IntechOpen, 2016, p. 235.

[450] M. F. Cherif, D. Trache, N. Brosse, F. Benaliouche and A. F. Tarchoun, "Comparison of the Physicochemical Properties and Thermal Stability of Organosolv and Kraft Lignins from Hardwood and Softwood Biomass for Their Potential Valorization," *Waste and Biomass Valorization,* vol. 11, pp. 6541-6553, 2020.

[451] C.-M. Popescu, M.-C. Popescu, G. Singurel, C. Vasile, D. S. Argyropoulos and S. Willfor, "Spectral Characterization of Eucalyptus Wood," *Applied Spectroscopy,* vol. 61, no. 11, pp. 1168-1177, 2007.

[452] J. Mao, X. Cao, D. C. Olk, W. Chu and K. Schmidt-Rohr, "Advanced solid-state NMR spectroscopy of natural organic matter," *Progress in Nuclear Magnetic Resonance Spectroscopy,* vol. 100, pp. 17-51, 2017.

[453] F. Lu and J. Ralph, "Solution-State NMR of Lignocellulosic Biomass," *Journal of Biobased Materials and Bioenergy,* vol. 5, pp. 169-180, 2011.

[454] J. D. Mao and K. Schmidt-Rohr, "Accurate Quantification of Aromaticity and Nonprotonated Aromatic Carbon Fraction in Natural Organic Matter by 13C Solid-State Nuclear Magnetic Resonance," *Environmental Science and Technology,* vol. 38, pp. 2680-2684, 2004.

[455] S. G. Kostryukov, P. S. Petrov, V. A. Kalyazin, P. P. Masterova, V. S. Tezikova, N. A. Khluchina, L. Y. Labzina and D. K. Alalvan, "Determination of Lignin Content in Plant Materials Using Solid-State 13C NMR Spectroscopy," *Polymer Science,* vol. 63, no. 5, pp. 544-552, 2021.

[456] X. Gao, D. D. Laskar, J. Zeng, G. L. Helms and S. Chen, "A 13C CP/MAS-Based Nondegradative Method for Lignin Content Analysis," *ACS Sustainable Chemistry & Engineering,* vol. 3, pp. 153-162, 2015.

[457] L. Fu, S. A. McCallum, J. Miao, C. Hart, G. J. Tudryn, F. Zhang and R. J. Linhardt, "Rapid and accurate determination of the lignin content of lignocellulosic biomass by solid-state NMR," *Fuel,* vol. 141, pp. 39-45, 2015.

[458] C. A. Rezende, M. A. de Lima, P. Maziero, E. R. deAzevedo, W. Garcia and I. Polikarpov, "Chemical and morphological characterization of sugarcane bagasse submitted to a delignification process for enhanced enzymatic digestibility," *Biotechnology for Biofuels,* vol. 4, p. 54, 2011.

[459] K. M. Holtman, N. Chen, M. A. Chappell, J. F. Kadla, L. Xu and J. Mao, "Chemical Structure and Heterogeneity Differences of Two Lignins from Loblolly Pine As Investigated by Advanced Solid-State NMR Spectroscopy," *Journal of Agricultural and Food Chemistry,* vol. 58, pp. 9882-9892, 2010.

[460] Y. Ono, M. Takeuchi and A. Isogai, "Changes in neutral sugar composition, molar mass and molar mass distribution, and solid-state structures of birch and Douglas fir by repeated sodium chlorite delignification," *Cellulose,* vol. 29, pp. 2119-2129, 2022.

[461] J. F. Haw, G. E. Maciel and H. A. Schroeder, "Carbon-13 nuclear magnetic resonance spectrometric study of wood and wood pulping with cross polarization and magic-angle spinning," *Analytical Chemistry,* vol. 56, no. 8, p. 1323–1329, 1984.

[462] G. Gilardi, L. Abis and A. E. Cass, "Carbon-13 CP/MAS solid-state NMR and FT-IR spectroscopy of wood cell wall biodegradation," *Enzyme and Microbial Technology,* vol. 17, pp. 268-275, 1995.

[463] K. Freudenberg and A. C. Neish, Constitution and biosynthesis of lignin, Springer Berlin Heidelberg, 1968.

[464] P. M.-A. Pawar, S. Koutaniemi, M. Tenkanen and E. J. Mellerowicz, "Acetylation of woody lignocellulose: significance and regulation," *Frontiers in plant science,* vol. 4, no. 118, 2013.

[465] P. Naidjonoka, M. A. Hernandez, G. K. Palsson, F. Heinrich, H. Stalbrand and T. Nylander, "On the interaction of softwood hemicellulose with cellulose surfaces in relation to molecular structure and physiochemical properties of hemicellulose," *Soft Matter,* vol. 16, p. 7063, 2020.

[466] J. Ralph and C. Rodger, "NMR of Lignin Model Trimers, or why you will never find crystalline regions in lignin," *Sixth International Symposium of Wood and Paper Chemistry,* vol. 1, p. 59, 1991.

[467] J. Ralph, J. M. Marita, S. A. Ralph, R. D. Hatfield, F. Lu, R. M. Ede, J. Peng, S. Quideau, R. F. Helm, J. H. Grabber, H. Kim, G. Jimenez-Monteon, Y. Zhang, H.-J. G. Jung, L. L. Landucci, J. J. MacKay, R. R. Sederoff, C. Chapple and A. M. Boudet, "Solution-State NMR of Lignins," in *Advances in Lignocellulosic Characterization*, Atlanta, GA, TAPPI Press, 1999, pp. 55-108.

[468] H. H. Nimz, D. Robert, O. Faix and M. Nemr, "Carbon-13 NMR Spectra of Lignins," *Holzforschung,* vol. 35, pp. 16-26, 1981.

[469] C. Chen, "Characterization of Milled Wood Lignins and Dehydrogenative Polymerisates from Monolignols by Carbon-13 NMR Spectroscopy," in *Lewis and Sarkanen; Lignin and Lignan Biosynthesis*, Washington, DC, ACS Symposium Series; American Chemical Society, 1998, pp. 255-275.

[470] K. M. Holtman, H.-m. Chang, H. Jameel and J. F. Kadla, "Quantitative 13C NMR Characterization of Milled Wood Lignins Isolated by Different Milling Techniques," *Journal of Wood Chemistry and Technology,* vol. 26, pp. 21-34, 2006.

[471] E. I. Evstigneyev, A. S. Mazur, A. V. Kalugina, A. V. Pranovich and A. V. Vasilyev, "Solid-State 13C CP/MAS NMR for Alkyl-O-Aryl Bond Determination in Lignin Preparations," *Journal of Wood Chemistry and Technology,* vol. 38, no. 2, pp. 137-148, 2018.

[472] S. Nakagame, R. P. Chandra, J. F. Kadla and J. N. Saddler, "The isolation, characterization and effect of lignin isolated from steam pretreated Douglas-fir on the enzymatic hydrolysis of cellulose," *Bioresource Technology,* vol. 102, no. 6, pp. 4507-4517, 2011.

Appendices

This thesis contains three appendices, Appendix A, B and C, of which provide supporting information for Chapters 2, 5 and 6, respectively. A references section which accounts for all three appendices is found following Appendix C.

Contents

Appendix A

Table A1. Summary of literature findings on the relationship lignin biosynthesis, content, and composition and drought stress response mechanisms.

Findings regarding lignin	Year	Author	Reference
Drought stress leads to altered lignin content and location, reduced lignification in middle lamella.	2002	Donaldson	[1]
Wall extensibility and root growth are inhibited by water deficit through alterations of phenolics in the cell wall in the root elongation zone.	2006	Fan et al.	[2]
Identified two distinct phases of drought stress which begins with a "mild endurance period" in which APOX and PAL are activated providing antioxidant protection, and a "severe injury period" in which lignifying peroxidases (GPOX, CPOX, and SPOX) are enhanced and lignin and lipid peroxidation is increased, restricting growth.	2007	Lee et al.	[3]
Leaf lignin content increased significantly under severe and moderate drought treatments. Significant differences in leaf lignification between drought-tolerant and drought-sensitive inbred lines.	2009	Hu et al.	[4]
Plants displayed varying responses to drought stress in terms of lignin content. Adjustment of total lignin resulted in an increased S/G ratio, either by increasing the amount of lignin S units or by the reduction of lignin G units.	2011	Moura-Sobczak et al.	[5]
Lignification decreased cell wall plasticity, inhibited cell wall growth, and assisted water retention. Lignin metabolism likely plays an important role in drought stress response and post-drought recovery.	2013	Li et al.	[6]
Increases in activity of various lignification-related enzymes in leaves in response to drought stress along with a positive correlation between leaf rolling and lignin content.	2013	Terzi et al.	[7]
Water stress plants featured significant accumulation of lignin. Lignification was dependant on internode maturation (immature or mature) and the tissue region in the stem (rind or pith). The mature internodes exhibited more lignification than immature internodes, and the rind more lignified than the pith.	2015	dos Santos et al.	[8]
Cinnamoyl CoA Reductase (CCR) accumulation, immune-cytolocalization of CCR protein and lignin deposition pattern of both root and stem tissues suggested that CCR is important for vascular tissue development under drought stress conditions in developing seedlings of *Leucaen.*	2015	Srivastava et al.	[9]
PtoMYB170 increases lignin deposition during wood formation in poplar by triggering lignin biosynthetic gene expression as well as promoting dark-induced stomata closure and thus drought tolerance in comparison to its divergent homologous PtoMYB216 gene.	2017	Xu et al. 2017	[10]
Lignin content of Miscanthus under drought stress was only mildly affected in the stem with no significant effect on the lignin in the leaf tissue.	2017	Van der Weijde et al.	[11]
Additionally, lignin content was not affected by drought stress – thus suggesting there may be a different mechanism of cell wall development for drought tolerance.	2017	Wildhagen et al.	[12]
Investigated the phenotypic traits of several engineered Arabidopsis plants for desired biofuel traits. Low lignin plants demonstrated reduced levels of water loss and better drought tolerance than wild-type plants. Results suggested a dependency between the drought tolerance of the plants with low lignin and ABA. Engineered plants with low lignin content had stomata closures more responsive to ABA, which reduces water loss and could be key for conferring drought tolerance. Examination of the expression of drought-responsive genes in the plants also indicated that the plants engineered for low lignin have increased induction of stress-responsive genes, potentially accounting for their superior drought tolerance.	2018	Yan et al.	[13]
Lignin monomer composition of the trees from their study was related to dry-season leaf life span and xylem embolism resistance. Total lignin content was not correlated with any variable in their study. They propose two main lines of trait variation for drought tolerance based on their results, one for keeping leaves around longer by increasing the S:G ratio, and the other for leafless trees to retain more water in the stem and to shed leaves to keep higher step water potential. They conclude that two lines are related and that the mechanisms that link S:G ratio with xylem embolisms vulnerability requires further investigation.	2018	Lima et al.	[14]
Created overexpression and knockdown transgenic lines of OsTF1L, a rice homeodomain-leucine zipper transcription factor gene, to further investigate their function in drought tolerance mechanisms. Overexpression of OsTF1L increased shoot lignification in the typically lignified tissues, with no ectopic lignification throughout the plant. Results indicated OsTF1L could increase drought tolerance through multiple molecular mechanisms such as stomatal movement and lignin biosynthesis.	2019	Bang et al.	[15]
Generated VlbZIP30-overexpressed transgenic grapevine plants and analyzed the plants response to drought treatment relative to control plants. Transgenic plants with overexpression of VlbZIP30 displayed improved drought tolerance, demonstrated by increased leaf relative water content, tuning of photosynthesis rate, and increased lignin content in the leaves. Increased leaf lignin content found to be primarily G units. Consistent with those results, VlbZIP30 regulated the expression of lignin biosynthetic (VvPRX N1, VvPRX4, and VvPRX72) and drought responsive (VvNAC17) genes, supporting lignin biosynthesis and improving drought resistance in grapevine. Overexpressing VlbZIP30 in A. *thaliana* did not result in the same regulation of lignin biosynthesis as in grapevine, which the authors speculate could be due to the difference in herbaceous vs woody vine biosynthesis mechanisms. Additionally, they found differences in lignin biosynthesis pathway regulation between stem and leaves, indicating that lignin biosynthesis may be regulated differently in certain tissues. First report of a bZIP transcriptional factor being directly involved in lignin biosynthesis and enhancing drought resistance.	2020	Tu et al.	[16]
Investigated the roles of important enzymes for lignin synthesis and lignins function in drought tolerance. Lignification is important for both melon and Arabidopsis in tolerating drought stress, and in particular CmCAD2 and 3 are key to this mechanism.	2020	Liu et al.	[17]
Explored drought response mechanism of foxtail millet by identifying a drought-induced R2R3-MYB transcriptional factor SiMYB56, and then comparing wild-type to transgenic plants with overexpressed SiMYB56. SiMYB56 may confer drought resistance in Foxtail millet by stimulating lignin biosynthesis under drought conditions.	2020	Xu et al.	[18]
Investigated the functional mechanisms of the DREB gene, DcDREB1A, cloned from carrots (*Daucus carota* L.), in response to drought stress in transgenic DcDREB1A overexpressed *Arabidopsis thaliana*. Results indicated that DcDREB1A is a nuclear protein involved in the regulation of plant drought stress tolerance by reducing water loss through stomata regulation and lignin deposition, decreasing oxidative lipid damage from ROS by increasing SOD and POD activities, and increasing the expression of other stress-responsive genes in plants.	2020	Li et al.	[19]
Looked at lignin deposition in the legume crop chickpea (Cicer arietinum L.) in response to drought stress. The authors concluded that root length and lignin content may be increased under drought conditions and that the LACCASE gene family may play a role in this stress response mechanism.	2020	Sharma et al.	[20]

Studied the proteomic profiles of stress proteins in tea plants under drought stress, specifically how they impact lignin. Tea plants respond to drought stress by inhibiting the accumulation of enzymes that catalyse lignin biosynthesis, while still promoting the accumulation of lignin, potentially through encouraging the activity of these enzymes.	2020	Gu et al.	[21]
Laccases are multicopper oxidases that play a role in lignin biosynthesis via oxidative coupling of monolignols. Positively correlated the expression of PeLAC10 with lignification in bamboo. PeLAC10 may be associated with lignification as shown by increased lignin content in the transgenic plants. Transgenic plants with overexpression of PeLAC10 had higher lignification, better survivability, relatively normal growth, reduced MDA content (MDA concentration is an indicator of membrane damage from ROS) compared to the control plants.	2020	Li et al.	[22]
Abiotic stress influenced cell wall polysaccharide content but not affect total lignin content. Lignin composition in the young shoot was significantly altered as shown by a decrease in S and increase in G lignin monomer units under both drought and salt stress conditions. They did not observe the same change in the xylem tissues. Transcriptional analysis revealed significant alterations in the expression of cell-wall related genes in young shoots in response to abiotic stress conditions. Core lignin biosynthesis pathway genes (F5H and COMT) were significantly decreased.	2020	Hori et al.	[23]
Transgenic plants with reduced lignin content in the vessels but not fibres had significant reduction in sap flow and hydraulic conductance in the xylem transportation system compared to wild type plants. These transgenic plants were also more vulnerable to drought, displaying dwarfism, low survival rate, and reduction of aboveground biomass yield. Transgenic plants with reduced lignin in the fibres instead of the vessels did not exhibit significantly reduced performance of xylem transportation system and vulnerability to drought stress, although they did show more significant reduced mechanical strength.	2020	Cao et al.	[24]
Isolated the full-length cDNA of a caffeoyl-CoA O-methyltransferase gene (CCoAOMT) from P. *ostii* which is an important enzyme for lignin biosynthesis and is suggested to play a role in abiotic stress tolerance. Under drought treatment, the expression levels of PoCCoAOMT in P. *ostii* was found to be overexpressed as compared to the control plants. PoCCoAOMT overexpressed transgenic tobacco plants were then used to further understand its role in drought stress response. In the transgenic lines they found PoCCoAOMT was expressed the in all three tissues Lignin content was significantly higher in the roots, stem, and leaves than in the wildtype plants, and highest in the roots. Four genes involved in lignin biosynthesis Phenylalanine ammonia gene (PAL), cinnamyl alcohol dehydrogenase gene (CAD), 4-coumarate: CoA ligase gene (4CL) and caffeic acid-O-methyltransferase gene (COMT) increased expression in the transgenic lines compared to the wildtype lines in all tissues, although inconsistently. Guaiacyl and syrignyl units were the predominant components of the plant's lignin, and the transgenic lines had increased levels of both monomer units, with a larger increase guaiacyl units. Drought stress tolerance specifically associated with higher levels of guaiacyl lignin units.	2021	Zhao et al.	[25]
Investigated the molecular mechanism of drought resistance in Cassava, a woody shrub important for both food and energy. Looked at the role of "Related to ABI/VP1" (RAV), an important subfamily of the APETALA2/ETHYLENE RESPONSIVE FACTOR (AP2/ERF) transcription factor family, in drought response. Proposing a model for the molecular mechanism of MeRAV5, MePOD and MeCAD15 action in modulating plant drought resistance, whereby drought stress induces the expression of these transcriptional factors, with MeRAV5 interacting with both MePOD and MeCAD15, promoting their activities and influencing the accumulation of H_2O_2 and lignin, conferring drought resistance to the plant.	2021	Yan et al.	[26]
Isolated a transcriptional factor gene, MsWRKY11, generally considered to be stress-inducible, and investigated its potential mechanism in drought tolerance in alfalfa. MsWRKY11 positively regulates lignin biosynthesis in stems and roots of alfalfa. The MsWRKY11 overexpressed lines had increased levels of all three main monomer units in the stem, with the greatest increase found in the S monomer units. MsWRKY11 is positively activated by MsWRKY22, increases stomatal density and lignin accumulation, conferring drought tolerance to alfalfa.	2021	Wen et al.	[27]
Overexpression of *PtrbHLH186* resulted in abnormal lignification, enhanced vessel cell development and altered wood composition, in particular, early and increased lignification, increased total lignin, increased G units, decreased polysaccharides, and more drought-tolerant phenotypes.	2022	Liu et al.	[28]

Appendix B

B1 Preprocessing

Table B1. Classical data preprocessing methods applicable to all types of multivariate data classified as column-wise methods. The most commonly used data preprocessing methods are mean centring and scaling. [29] Column scaling is typically used if the data variables are on different scales or are measured on different ranges leading to varying impacts on the analysis. [30]

Method	Brief Description
Mean Centering	Shifts the center of the data to the origin, mean-centered data will have a mean of zero which is better for visualization.
Scaling	Divides each variable by a number. This can make variables with different scales/units more comparable. When each variable is divided by its standard deviation, it is commonly referred to as unit variance scaling. Scaling should not be applied to spectroscopic data. [29]
Autoscaling/Standardization	Combining the mean centring and unit variance scaling is sometimes termed autoscaling or standardization. [31] Note that the term 'normalization' is also sometimes used in this context, but this has been recommended against because of its inconsistent use in the chemometric literature. [30] Autoscaling can sometimes amplify noise. [32]
Pareto Scaling	Pareto scaling is another form of column scaling that differs from autoscaling in that it uses the square root of the standard deviation as the scaling factor instead of just the standard deviation. [33]

Table B2. In contrast to the column-wise classical preprocessing methods described in Table B, signal correction methods are typically row-wise manipulations and are applied to reduce the impact of perturbations in the data while augmenting the useful information. [29]

Method	Brief Description
Baseline Correction	Continuous and low-frequency background signals independent of the sample under investigation in spectroscopic data are called baselines. [34] Baseline correction involves identifying the nature and magnitude of the baseline and removing it from the data leaving the baseline noise numerically centered about zero. [35] [36] Methods for baseline correction involve subtracting a constant value from the data or forming a line through two data points and subtracting the values of the line from its corresponding data points. [34] For chromatographic analysis, a "blank" chromatogram could also be subtracted from the sample runs to remove the baseline. [35] More complex techniques include the use of mathematical algorithms to identify and remove the baseline from the data; however, it is important to ensure that the variations in the data from the sample under investigation are not affected by the correction. [35]
Smoothing	Smoothing is the name generally used for data preprocessing that seeks to reduce the high-frequency noise in the data unrelated to the sample under investigation without corrupting the desired information. [34] [37] [35] Similar to baseline correction, there are several different techniques for smoothing experimental data. A popular method for addressing noise is the Savitzky-Golay smoothing algorithm [37] [38] which is a moving average algorithm based on simple local least-squares polynomial approximation. [39] [40]
Differentiation	Derivatives can eliminate background signals, assist in determining peak positions, and improve the visual resolution of peaks. [41] Common algorithms for differentiation include the Savitzky-Golay and Norris-Williams [42] algorithms which are used to smooth the noise that is enhanced by differentiation. [29] Derivatives are a common preprocessing strategy; however, they should be used with caution as they can enhance noise.
Row Scaling (Normalization)	Row scaling is a data preprocessing strategy to make the data more comparable by scaling the rows of a data set to a constant total. [32] This row-wise scaling transformation can also be referred to as normalization. The term normalization sometimes implies scaling using the sum of squares; however, this is inconsistently applied in the literature. [32]
Multiplicative Signal Correction (MSC)/Extended Multiplicative Signal Correction (EMSC)	The EMSC is essentially a method that mathematically corrects for light scattering variations resulting from factors such as particle size, shape, and other physical variations of the sample by separating the physical light-scattering effects from the vibrational spectra recorded. [43] The EMSC is a powerful data preprocessing approach for vibrational spectra that can separate and quantify different types of variations in the spectra as well as provide basic baseline correction and normalization. [44] A more in-depth description and tutorial for the EMSC can be found elsewhere. [44] [43]
Standard Normal Variate (SNV)	A similar and linearly related technique to MSC, SNV, is used to process data prior to further analysis by correcting for physical variations in the sample such as particle effects. [45] The SNV is a mathematical transformation applied to individual spectra that can reduce multiplicative interferences of scattering and particle size. [46] An in-depth description and comparison of MSC and SNV found insufficient evidence to conclude one technique is superior to the other. [45]

Table B3. Dimensionality reduction/variable selection methods

Method	Brief Description
Genetic Algorithms (GA)	Genetic Algorithms are an optimization/variable selection approach based on the concept of evolution by natural selection. [47] There are several variants of GA based on the methodology of a population-based stochastic algorithm that produces 'better' populations. [31] More detailed descriptions and implications for chemometric applications can be found elsewhere. [48] [49] [50]
Orthogonal Signal Correction (OSC)	Any systematic variation in **X** that is unrelated to **Y** and included in the multivariate model could result in an imprecise model. The basis of the OSC technique is to remove the parts of the data in **X** that are completely unrelated to **Y.** OSC does this by only removing the parts of **X** that are mathematically orthogonal to **Y** (or very close to orthogonal). The OSC algorithm follows the same process as the PLS algorithm, except for the calculation of the weights. [51]
Orthogonal Projections to Latent Structures (O-PLS)	Orthogonal projections to latent structures is another method that maintains the same objective as OSC but with a different approach. [52] It is a way to remove systematic variation from an input data set **X** that is not correlated (orthogonal) to the response set **Y**. O-PLS removes the non-correlated systematic variation in **X** by analyzing the variation explained in each PLS component. Advantages of O-PLS include the improved interpretation of PLS models and their parameters scores, loadings, and residuals.

B2 Performance Criteria

Performance criteria are based on residuals (e_i) obtained by subtracting the predicted value ($\hat{y}_i$) (Equation 1) from its true value (y_i).

$$e_i = y_i - \hat{y}_i \tag{1}$$

Performance criteria are derived from the residuals but use different mathematical strategies to define distinct criteria. For a more detailed overview of these performance criteria, Varmuza and Filzmoser provided a clear introduction. [31] Here a brief summary is provided. The standard deviation of the standard error of prediction (SEP) provides an estimate of the accuracy of predictions and is defined in Equation 2

$$SEP = \sqrt{\frac{1}{z-1}\sum_{i=1}^{z}(y_i - \hat{y}_i - \text{bias})^2} \tag{2}$$

where z is the number of predictions, and the bias is the arithmetic mean of the prediction errors. Bias can occur from systematic errors and is defined in Equation 3.

$$bias = \frac{1}{z}\sum_{i=1}^{z}(y_i - \hat{y}_i) \tag{3}$$

If the predicted values in Equation 2 are from the calibration set, it is sometimes referred to as the standard error of calibration. Another performance criteria commonly used is the mean squared error (MSE) (Equation 4) which is the arithmetic mean of the squared errors.

$$MSE = \frac{1}{z}\sum_{i=1}^{z}(y_i - \hat{y}_i)^2 \tag{4}$$

Taking the square of Equation 4 provides root mean squared error (RMSE) (Equation 5) which is almost the same as the SEP if bias can be neglected. The RMSE criteria are preferred for practical applications because their units match the original data, however, the non-squared MSE are useful for model optimization purposes. [29]

$$RMSE = \sqrt{\frac{1}{z}\sum_{i=1}^{z}(y_i - \hat{y}_i)^2} \tag{5}$$

Similar to MSE is the predicted residual error sum of squares (PRESS) defined in Equation 6 below.

$$PRESS = \sum_{i=1}^{z}(y_i - \hat{y}_i)^2 \tag{6}$$

Like the SEP, either indexing or altering the acronym are common approaches to specify the data set used to calculate the performance criteria. Sometimes R^2, the coefficient of determination (Equation 7) is used to represent the spread of the predictions.

$$R^2 = 1 - \frac{PRESS}{\sum_{i=1}^{z}(y_i - \bar{y}_i)^2} \tag{7}$$

B3 Literature Summaries

Table B4. Preprocessing methods, multivariate techniques, and data analyzed in papers that use NIR and chemometrics to study lignocellulosic biomass or lignin.[1]

Data Preprocessing	Multivariate Analysis	X Matrix (Objects and Observed Features/Predictor Variables)	Y Matrix (Properties of Objects/Response Variables)	Year Published	Reference
MSC	PCA, PLS	NIR, ^{13}C CP/MAS NMR, MIR spectra	Klason lignin, glucose, xylose, galactose, arabinose, mannose	1991	[53]
Log transformation, scatter-correction using SNV-detrending	Modified PLS	NIR spectra	Lignin and other biomass feedstock components	1996	[54]
First derivative	PLS	NIR spectra	Acid insoluble and soluble lignin, Kappa number, relative viscosity, ISO brightness, degree of polymerization, glucan, xylan	2002	[55]
-	PCA, PLS	Py-MB/MS spectra, NIR spectra	Weight loss during brown-rot biodegradation of spruce wood	2002	[56]
Mean centering, variance normalized	PCA, PLS	Py-MB/MS spectra, NIR spectra	Lignin, glucose, xylose, mannose, galactose, arabinose, rhamnose	2004	[57]
Second derivative	PLS	NIR spectra	Klason lignin, total lignin, acid-soluble lignin, extractives	2004	[58]
Second derivative	MLR, PLS	NIR spectra	Lignin content	2005	[59]
First and second derivative Savitzky-Golay algorithm	PCA, PLS	NIR spectra	Klason lignin	2006	[60]
Second derivative	PLS	NIR spectra	Acid-soluble lignin, Klason lignin, total lignin, extractives, cellulose	2006	[61]
Second derivative	PLS	NIR spectra	Klason lignin, thioacidolysis S/G ratio, cellulose content, xylose content	2006	[62]
First derivative MSC,	PLS	NIR spectra	H/G from analytical pyrolysis	2006	[63]
MSC, first and second derivative, vector normalization, constant offset	PLS	NIR spectra	Kappa number	2007	[64]
Centering, autoscaling, Saunderson correction, MSC, square, square root, and Savitzky-Golay transformation and differentiation, OSC	PLS	MIR spectra, NIR spectra	Lignin, arabinose, galactose, glucose, mannose, xylose, extractives	2007	[65]
Divided by background spectrum, Karl Norris derivative filter	PLS	NIR spectra	Klason lignin, hemicellulose, cellulose, moisture, insoluble ash, total ash	2007	[66]
Baseline correction, normalization, Second derivative with Savitzky-Golay algorithm	PLS	NIR spectra	Weight percentage gain due to acetylation	2008	[67]
-	Genetic inverse least squares	NIR spectra	Lignin and extractives	2009	[68]
First derivative with Savitzk-Golay algorithm	PLS	NIR spectra	H/G/S proportions from modified thioacidolysis	2009	[69]
Second derivative with Savitzky-Golay algorithm, unit vector normalization	PCA	NIR spectra	-	2010	[70]
Savitzky–Golay second derivative	PLS	NIR spectra	Saccharification ratio, glucose, xylose, arabinose, mannose, galactose, lignin, ash	2011	[71]
-	PLS	NIR spectra	Klason lignin, extractives, cellulose, hemicellulose,	2011	[72]
-	Genetic inverse least squares	NIR spectra	Lignin and extractives	2011	[73]
Constant offset elimination, straight line subtraction, normalization, MSC, first/second derivative with Savitzky-Golay algorithm	PLS	NIR spectra	Total lignin content	2011	[74] [75]
First/second derivative with Savitzky-Golay algorithm, normalization, MSC, straight line subtraction	PLS	NIR spectra	S/G from analytical pyrolysis	2011	[76]
Second derivative	PLS	NIR spectra	Lignin content, monomer composition (S/G ratio), cellulose, extractives, mannose, galactose, xylose	2011	[77]
MSC, mean centering	PCA, PLS, backwards interval partial least squares, HCA	NIR spectra	Enzymatic release of arabinose, xylose, glucose	2012	[78]
Savitzky–Golay second derivative	PCA, PLS	NIR spectra	Saccharification efficiency, cellulose, hemicellulose, lignin content	2012	[79]
Offset correction, MSC, vector normalization, derivatives, Savitzky-Golay algorithm	PLS	NIR spectra	Wood density from X-ray micro density data	2012	[80]
First and Second derivative with Savitzky-Golay algorithm, MSC, straight line subtraction	PCA, PLS	NIR spectra	Extractive's content	2012	[81]

Constant offset elimination, MSC, first/second derivative, normalization, baseline correction	PLS	NIR spectra	Lignin and extractives	2013	[82]
SNV, first/second derivatives, OSC, EMSC	PCA, PLS	Visible-NIR spectra	Klason lignin, acid soluble lignin, total lignin, extractives, moisture, ash, insoluble residue	2013	[83]
First derivative with Savitzky-Golay algorithm	PLS, PCR	NIR spectra	Lignin, extractives, cellulose, hemicellulose	2014	[84]
MSC, SNV, first/second derivative using Savitzky-Golay algorithm, mean centering	PLS, variable importance for projection	NIR spectra	Total dietary fibre (soluble and insoluble non-starch polysaccharides which include cellulose, hemicellulose, pectin, b-glucans, gums, and lignin)	2015	[85]
Normalization, detrend, first and second derivative	PCA, PLS	NIR spectra	H/S/G lignin, Klason lignin, Py-Lignin, Acid-soluble lignin, xylose, glucose, C5/C6, extractives	2017	[86]
GA, ordered predictors selection, interval PLS, Kennard and Stone algorithm	PLS	NIR spectra	Klason lignin	2017	[87]
MSC, derivatives and smoothing	PLS	NIR spectra	Cellulose, hemicellulose, lignin, cellulose conversion, cell wall content	2017	[88]
First and second derivative, Savitzky-Golay	PLS-1	NIR spectra	Klason lignin, arabinose, galactose, glucose, mannose, xylose, 4-O-methylglucuronic acid, galacturonic acid. and glucuronic acid	2018	[89]
SNV, MSC, first/second derivative	PLS, PLS and competitive adaptive reweighted sampling	NIR spectra	Acetyl bromide lignin content	2019	[90]
First derivative, moving average filtering, MSC, SNV	PLS, least square SVM, back-propagation neural network, kernel extreme learning machine, PCA, particle swarm optimization	NIR spectra	Acid insoluble lignin, holocellulose	2020	[91]
MSC, first/second derivatives,	PLS, Variable selection methods	NIR spectra	Lignin and holocellulose content	2020	[92]
Moving average, baseline, SNV and MSC	Backward interval partial least-squares, synergy interval partial least squares, uninformative variable elimination algorithms,	NIR spectra	Klason lignin	2020	[93]
SNV, normalization, detrending, first and second derivative with Savitzky-Golay smoothing	PLS	NIR spectra	Glucan, xylan, lignin, extractives, and ash content	2020	[94]
Savitzky-Golay smoothing, normalization, SNV, first derivative,	PLS	NIR spectra	Lignin content	2021	[95]
EMSC, SNV, mean centering	PCA, PLS-DA	NIR Spectra	Acid insoluble-soluble lignin	2021	[96]
Normalization, second derivative Savitzky-Golay smoothing	PCA, PLS	NIR spectra	Carbonization characteristics (Carbon content wt%, Oxygen/Carbon ratio, hydrogen/carbon ratio)	2021	[97]

[1]When information is not available or non-applicable '-' is used as a placeholder

Table B5. Preprocessing methods, multivariate techniques, and data analyzed in papers that use MIR and chemometrics to study lignocellulosic biomass or lignin.[1]

Data Preprocessing	Multivariate Analysis	X Matrix (Objects and Observed Features/Predictor Variables)	Y Matrix (Known Properties of Objects/Response Variables)	Year Published	Reference
Baseline correction	MLR	MIR spectra	Lignin, cellulose, and hemicellulose content	1985	[98]
MSC	PCA, PLS	NIR, ^{13}C CP/MAS NMR, MIR spectra	Klason lignin, glucose, xylose, galactose, arabinose, mannose	1991	[53]
Baseline correction, normalization	PCR, PLS	MIR spectra	Phenolic OH Groups	1993	[99]
Normalization	PCA, SIMCA	MIR spectra	-	1999	[100]
Baseline correction	PLS	MIR spectra	Lignin, extractives, total carbohydrate, basic density	1999	[101]
Baseline correction, normalization, mean centering, variance normalization	PCR, PLS	MIR spectra	Lignin and other wood component concentration	2000	[102]
Normalization	PLS, PCR	MIR spectra	Glucose, xylose, galactose, mannose	2001	[103]
-	PLS	MIR spectra	Glucan, mannan, galactan, xylan, and Klason and acid-soluble lignin compositions of pretreated softwood solid residues	2001	[104]
OSC	PLS	MIR spectra	Kappa number (residual lignin content) and carbohydrate contents	2002	[105]
Baseline correction, centering, OSC	PCA, PLS	MIR spectra, ^{13}C CPMAS NMR spectra	Weight loss from decay using soft rot fungi	2003	[106]
Baseline correction	PCA, PLS	MIR spectra	Lignin content, carboxyl and phenolic OH groups, antioxidant and emulsifying properties of lignin	2004	[107]
Normalization, MSC	PCA, PLS	MIR spectra, Py-MB/MS spectra	Distance from bark to pith	2005	[108]
Centering, autoscaling, Saunderson correction, MSC, square, square root, and Savitzky-Golay transformation and differentiation, OSC	PLS	MIR spectra, NIR spectra	Lignin, arabinose, galactose, glucose, mannose, xylose, extractives	2007	[65]
Baseline correction, moving average smoothing	-	MIR spectra	-	2007	[109]
Column centering, normalization, baseline correction,	PLS	MIR spectra	Kappa number	2007	[110]
Baseline correction, normalization, second derivative	Non-linear least squares	MIR spectra	Mathematical model of deconvoluted lignin spectra	2008	[111]
Mean normalization, MSC	PLS	MIR spectra, Py-MB/MS spectra	Lignin content from UV-Vis and gravimetric analysis	2009	[112]
Baseline correction, normalization, first derivative, variance scaling	PLS	MIR spectra	Contents of components (i.e., lignin, cellulose, and hemicellulose)	2010	[113]
Baseline correction, vector normalization, second derivative	PCA	MIR spectra	-	2010	[114]
Normalization, MSC	PCA	MIR spectra	-	2010	[115]
First/Second derivative, SNV, MSC	PLS	MIR spectra	Lignin content	2011	[116]
First derivative, vector normalization, baseline correction (Rubber band method)	PCA, PLS	MIR spectra	Lignin content Acetyl Bromide of extractive-free poplar, bomb calorimetry energy content	2011	[117]
Baseline correction	Linear Regression	MIR spectra	S Ratio (S/(S+G))	2012	[118]
Smoothing, first derivative, second derivative, baseline correction, SNV, detrending, unit vector normalization, MSC, OSC	PCA, PLS	MIR spectra	Surrogate mixtures of cellulose, xylan, and lignin	2012	[119]
Baseline correction, normalization, first/second derivative with Savitzky-Golay smoothing	Linear Regression	MIR spectra	Concentration of S and G units	2013	[120]
Normalization, MSC	PCA, PLS	MIR spectra	Glass transition temperature of lignin from differential scanning calorimetry	2013	[121]
-	PLS	MIR spectra	^{31}P NMR determined alkyl-OH, condensed-OH, S, G, H, COOH, aryl-OH, S/G	2014	[122]
First derivative	PLS, PCR	MIR spectra	Wood chemistry contents (i.e., lignin, cellulose, hemicellulose, extractives)	2015	[123]
Baseline correction, first derivative, mean centering	PLS	MIR spectra	Lignin, Cellulose, Hemicelluloses, Extractives, thermal reactivity, energy content	2016	[124]
-	PLS, Monte Carlo Sampling	MIR spectra	Cellulose, hemicellulose, lignin Van Soest	2018	[125]
-	PCA, Discriminant analysis	MIR, Py-GC/MS	-	2018	[126]
SNV, mean centering	PCA	MIR spectra	-	2018	[127]
Baseline correction, normalization, mean centering, first/second derivative	PCA, PLS	MIR spectra	GPC data (M_n, M_w) and HSQC NMR data (inter-unit linkage abundance) in technical lignins	2019	[128]
Normalization	PCA, linear regression	MIR spectra	Lignin content	2019	[129]
Autoscaling	PCA	MIR spectra	-	2019	[130]
-	PCA	MIR spectra	-	2020	[131]
Filtered using Savitzky-Golay algorithm including scattering corrections and smoothing, derivatives, normalization	PCA, HCA	MIR spectra	-	2020	[132]
-	PCA, HCA	MIR spectra	-	2020	[133]
-	PCA	HSQC NMR spectra, MIR spectra	-	-	[134]
-	PCA, PLS	ATR-IR spectra	Temperatures of aqueous solvent	2021	[135]
MSC	PCA, PLS	ATR-IR spectra	Weight average molecular weight and polydispersity	2021	[136]
Background subtraction, baseline correction, vector normalization, SNV	PCA, HCA	MIR spectra	-	2022	[137]

[1]When information is not available or non-applicable '-' is used as a placeholder

Table B6. Preprocessing methods, multivariate techniques, and data analyzed in papers that use Raman and chemometrics to study lignocellulosic biomass or lignin.[1]

Data Preprocessing	Multivariate Analysis	X Matrix (Objects and Observed Features/Predictor Variables)	Y Matrix (Properties of Objects/Response Variables)	Year Published	Reference
Second derivative MSC	PLS	FT-Raman	S/G ratio by thioacidolysis	1998	[138]
MSC, first derivative	PLS	FT-Raman	Lignin, S/G ratio, wood constituents, anatomy, and basic density	2003	[139]
-	PCA, PLS	UVRR model compound spectra	Quantity of lignin model compound structures	2003 2005	[140] [141]
Smoothing, baseline correction	Linear Regression	FT-Raman spectra	S/G ratio from Py-GC/MS	2011	[142]
Baseline correction, normalization	-	FT-Raman spectra	-	2011	[143]
Background subtraction using blank, mean centering, first derivative	PLS	1064 nm dispersive multichannel Raman spectra	Lignin monomer composition	2011	[144]
Wavelet decomposition, second derivative, PCA,	Spectral enthalpy minimization methodology [145] with stimulated annealing optimization algorithm [146], SVD	Raman micro spectra	-	2012	[147]
Baseline correction, Savitzky-Golay smoothing, mean centering	PCA, PCR	1064 nm dispersive multichannel Raman spectra	S and G from thioacidolysis,	2012	[148]
First/second derivatives with Savitzky-Golay algorithm, smoothing, SNV, MSC. EMSC	PLS	FT-Raman/NIR/MIR spectra	S/G ratio from Py-MB/MS	2014	[149]
Baseline correction	PCA, cluster analysis, VCA	Confocal Raman spectra	-	2014	[527]
PCA filtering using SVD	MCR-ALS	Raman spectra	Soluble and insoluble lignin	2015	[151]
First/second derivative, EMSC	PLS	FT-Raman spectra	S/G ratio from Py-MB/MS	2015	[152]
-	MCR of hyperspectral images	Stimulated Raman Scattering spectra	-	2015	[153]
Baseline correction, normalization	VCA	Confocal Raman spectra	-	2016	[154]
Cosmic ray removal, background subtraction	VCA, NMF, MCR ALS, PCA filtering	Confocal Raman microscopy spectra	-	2018	[155]
Normalization	PCA, clustering analysis	Micro-Raman spectra	-	2021	[156]
Savitzky-Golay Smoothing, second derivative baseline correction,	PCR, PLS, LASSO regression, ridge regression	FT-Raman spectra	Klason lignin, acid-soluble lignin,	2021	[157]
Asymmetrical least square smoothing, linear baseline correction, first derivatives, vector normalization, Savitzky-Golay algorithm	PCA, HCA	Raman micro spectra	-	2021	[158]
Savitzky-Golay smoothing, adaptive iteratively reweighted penalized least squares baseline correction	PCR, PLS, ridge regression, LASSO regression, elastic net regression	FT-Raman spectra	Holocellulose content	2022	[159]
Savitzky-Golay smoothing, adaptive iteratively reweighted penalized least squares baseline correction	Linear regression	FT-Raman spectra	Kappa number	2022	[160]

[1]When information is not available or non-applicable '-' is used as a placeholder

Table B7. Preprocessing methods, multivariate techniques, and data analyzed in papers that use NMR and chemometrics to study lignocellulosic biomass or lignin.[1]

Data Preprocessing	Multivariate Analysis	X Matrix (Objects and Observed Features/Predictor Variables)	Y Matrix (Properties of Objects/Response Variables)	Year Published	Reference
Normalization	PCA, PLS	^{13}C CP-MAS NMR spectra	Klason lignin content	1989	[161]
MSC	PCA, PLS	NIR, ^{13}C CP/MAS NMR, MIR spectra	Klason lignin, glucose, xylose, galactose, arabinose, mannose	1991	[53]
Unit variance scaling, mean centering	PCA, PLS	^{13}C and ^{31}P NMR spectra, SEC (M_n, M_w, M_z)	Combustion Parameters (Burning time and swelling of black liquor)	1999	[162]
Centering	PCA	^{13}C CP/MAS NMR spectra, FT-IR spectra	-	2003	[106]
Centering and normalization	PCA-ANN	Atom based features of lignin atoms (descriptors)	^{13}C NMR shifts	2004	[163]
Baseline correction, normalization, scaling, mean centering, pareto scaling, scaling to unit variance	PCA, OPLS-DA	2D ^{13}C-^{1}H HSQC NMR spectra	-	2009	[164]
Mean centred, normalized to total spectral intensity, Pareto scaling	PCA	^{13}C NMR spectra	-	2011	[165]
Automatic phase correction, baseline correction	PCA, PLS	^{1}H NMR spectra	HPLC Sugar concentrations (glucose, xylose, other minor sugars) in biomass hydrolysates	2012	[166]
-	PCA	^{31}P NMR spectra	-	2014	[122]
Baseline correction, normalization, Pareto scaling	PCA, PLS-DA and linear models, HCA	^{1}H NMR spectra	Lignin, cellulose, hemicellulose	2019	[167]
-	PCA	HSQC NMR spectra, FT-IR spectra	-	-	[134]
Pareto scaling	PCA	^{13}C CP-MAS NMR	-	2020	[168]
Baseline correction	PLS, MLR	^{1}H and DOSY NMR spectra	Weight and number average molecular weight, polydispersity	2021	[169]

[1]When information is not available or non-applicable '-' is used as a placeholder

Table B8. Preprocessing methods, multivariate techniques, and data analyzed in papers that use MS and chemometrics to study lignocellulosic biomass or lignin.[1]

Data Preprocessing	Multivariate Analysis	X Matrix (Objects and Observed Features/Predictor Variables)	Y Matrix (Properties of Objects/Response Variables)	Year Published	Reference
Baseline correction, normalization, scaling to unit variance	PCA, PLS	Py-GC/MS spectra	Klason Lignin and carbohydrate content from acid hydrolysis/derivatization-GC	1993	[170]
Normalization	Factor analysis	Py-MBMS spectra	-	1994	[171]
Mean centering, variance normalized	PCA, PLS	Py-MBMS spectra	Bark source, acid concentration, phenolysis temperature	2002	[172]
-	PCA, PLS	Py-MBMS spectra, NIR spectra	Weight loss during brown-rot biodegradation of spruce wood	2002	[56]
Mean centering, variance normalized	PCA, PLS	Py-MBMS spectra, NIR spectra	Lignin, glucose, xylose, mannose, galactose, arabinose, rhamnose	2004	[57]
Normalization, MSC	PCA, PLS	Py-MBMS spectra, FT-IR spectra	Distance from bark to pith	2005	[108]
Normalization	PCA	Py-MBMS spectra	-	2006	[173]
Baseline correction, normalized to ^{38}Ar isotope of carrier gas	PCA	TG/MS spectra	-	2007	[174]
Centering, row normalization	PCA	Py-DE-MS	-	2008	[175]
Background subtraction, normalized to total ion content	PCA	Py-MBMS spectra	-	2008	[176]
Background subtraction, normalized to total ion content, PCA	PCA	Py-M/MS spectra	-	2009	[177]
Normalization, MSC	PLS	Py-MBMS spectra, FT-IR spectra	Lignin content from UV-Vis and gravimetric analysis	2009	[112]
Normalization	PCA	Py-MBMS spectra	-	2009	[178]
Chromatograph moving average smoothing and alignment, baseline correction	MCR-AR	Py-GC/MS spectra	-	2012	[179]
Normalization	PCA	Py-GC/MS spectra	-	2013	[180]
-	PCA, Discriminant analysis with automatic backward variable selection, PLS, multiple regression	Py-GC/MS spectra	Active alkali (Kappa number)	2014	[181]
Normalization	Partial correlation analysis, multivariate correlated components regression	Py-GC/MS spectra	Antioxidant activity	2015	[182]
-	PCA	Py-MBMS spectra	-	2015	[183]
Normalization	PCA, *k*-means clustering	Py-GC/MS spectra	-	2016	[184]
Baseline correction, peak alignment, normalization,	PCA	GC/MS spectra	-	2016	[185]
-	PCA, Discriminant analysis	Py-GC/MS spectra, FT-IR spectra	-	2018	[126]
Autoscaling, normalization	PCA, Parallel factor analysis 2, PLS-DA	Py-GC/MS spectra	Class of trees	2018	[186]
Mean centering, scaling to unit variance	PCA-quadratic discriminant analysis	Ultra-high-performance liquid chromatography/high-resolution multiple-stage tandem MS	-	2018	[187]
Baseline correction, MCR-AR, normalization	OPLS-DA	Py-GC/MS	Cellulose, hemicellulose, softwood lignin, wheat straw lignin	2019	[188]
Normalization, peak alignment	PCA	Py-GC/MS	-	2019	[189]
-	PCA	Py-GC/MS	-	2019	[190]
Spectral deconvolution, peak alignment, Savitzky-Golay filter smoothing, normalization	PCA, HCA, hierarchical clustering on the principal components with *k*-means partition	Py-GC/MS	-	2020	[191]
Normalization, power transformation, variable stability scaling	PCA	Pyrolysis direct insertion probe with atmospheric pressure chemical ionization and atmospheric pressure photoionization coupled to ultrahigh-resolution MS	-	2020	[192]
TIC-normalized	PCA, Hierarchical clustering, *k*-means clustering	Py-MBMS	-	2021	[193]
-	PCA	Py-GC/MS	-	2021	[194]

[1]When information is not available or non-applicable '-' is used as a placeholder

Table B9. Preprocessing methods, multivariate techniques, and data analyzed in papers do not fall in any one of the major categories presented above or are not associated with a particular analytical technique.[1]

Data Preprocessing	Multivariate Analysis	X Matrix (Objects and Observed Features/Predictor Variables)	Y Matrix (Properties of Objects/Response Variables)	Year Published	Reference
Mean centering	PCA, PCR, PLS	Visible spectra	Kappa number	1998	[195]
Mean centering, scaling to unit variance	PCA, PCR, PLS	Capillary zone electrophoresis data	Total cooking yield, kappa number, iso brightness	2000	[196]
-	PCA	Parameters of battery expanders and lignin properties	-	2000	[197]
Baseline correction	PCA	TG/MS differential curve	-	2007	[174]
Standardization	PLS	Physicochemical characteristics (Accessible interior surface area, exterior surface area, amorphous fraction, absorbance ratio, destruction of hydrogen bonds)	Enzymatic digestibility	2009	[198]
Normalization	PCA	Quantified lignin characteristics	-	2010	[199]
-	PCA, PLS	Pretreatment process parameters and chemical composition variables	Pretreatment effectiveness variables	2014	[200]
Mean centering and scaling to unit variance	PCA	Literature-based input data sets of the thermochemical conversion processes of various biomass	-	2016	[201]
Normalized to dry mass	PCA, PCR, PLS	TGA data	Lignin, cellulose, ash, volatile matter, fixed carbon	2017	[202]
-	PCA	Single lignin molecules	-	2018	[203]
-	PCA	Tree traits (lignin content, S unit content, S/G ratio, wood density, stem water potential in dry season, and leaf life span)	-	2018	[204]
-	PCA	Motion of atoms from molecular dynamics simulations	-	2018	[203]
-	Multivariable linear estimation	Lignin structural features (methoxy, oxygen, aliphatic OH, ether linkage content, weight average molecular weight)	Conversion reactivity toward oxidative depolymerization to monomers	2018	[205]
-	MLR, real valued genetic algorithm	Lignin and extractives content	Higher heating values	2018	[206]
Fraser-Suzuki deconvolution	PCA	Energy-dispersive X-ray fluorescence (EDXRF)	-	2019	[207]
-	PCA, PLS	Pretreatment conditions, Composition of pretreated samples, morphological characterizations, Pretreatment evaluation	Pretreatment conditions, Composition of pretreated samples, morphological characterizations, Pretreatment evaluation, Enzymatic effect evaluation	2019	[208]
Denoising using Savitzky-Golay algorithm, baseline correction, autoscaling	PLS	High performance thin layer chromatography-densitometry	Lignin content	2020	[209]
Friedmans method, Fraser-Suzuki deconvolution, modified criado method	SOM neural network, SVM learning algorithm	TGA-DTG	-	2020	[210]
-	MLR, multinomial regression, *k*-means clustering	Red-Green-Blue Colour space images	Quality assignments, chemical composition, inorganic speciation	2020	[211]
-	PCA, PLS	54 important variables of deep eutectic solvent pretreatment	54 important variables of deep eutectic solvent pretreatment	2020	[212]
-	PCA	Properties of the feedstock to the RCF depolymerization yields (lignin content (S, G, H, coumarate, and ferulate), monomer selectivities, total monomer yields)	-	2020	[213]
Baseline correction and background correction, Savitzky-Golay smoothing	PCA, self modeling MCR, Bayesian hierarchical clustering, data fusion	FTIR and ^{1}H NMR	Identifying reaction networks of complex mixtures	2020	[214]
Normalization, Fraser-Suzuki deconvolution	Heat map, *k*-means clustering, Self organizing map, ANN	TGA differential curve	-	2021	[215]
Subtraction of dry mass, moving average smoothing, first derivative, mean centering	PLS	TGA differential curve	Lignin content	2021	[216]
-	PCA, PLS	42 important variables of deep eutectic solvent pretreatment	Pretreatment evaluation parameters: Xylan, lignin, solid recovery rate, recovery rate of glucan, recovery rate of xylan, delignification rate	2021	[217]
-	ANN	Proximate analysis (fixed carbon, volatile matter, ash and moisture)	Cellulose, hemicellulose, lignin	2021	[218]

-	Group additivity model based on PCA	Gas-phase density functional theory data of 4100 species at the M06-2X/6-311++G(d,p) level	-	2021	[219]
-	MLR, generalized additive models, random forests	Fluorescence lifetime imaging microscopy parameters	Wood cell wall or saccharification traits	2021	[220]
-	Box–Behnken Experimental Design, MLR	$\beta - O - 4$ inter-unit linkages, hydroxyl functionalities, M_w, molar mass dispersity, lignin yield	Temperature, acidity, ethanol/water ratio	2022	[221]
-	Random forest regression	Solvent, active metal/solvent ratio, temperature, solvent ethanol, active metal/lignin ratio, catalyst to solvent ratio, reaction time and more	Bio-oil yield, char yield and reaction time in catalytic lignin depolymerization	2022	[222]
-	PCA	Variables of two lignin degradation methods (CuO-NaOH oxidation and tetramethyl ammonium hydroxide thermochemolysis)	-	2022	[223]
Locally weighted regression smoothing	Linear regression, SVM, decision tree, ANN	Material wt %, retention during cyclic charge/discharge	Specific capacitance variation of a lignin-based supercapacitor	2022	[224]

[1] When information is not available or non-applicable '-' is used as a placeholder

Appendix C

C1 Wet chemistry

Table C1. Lignin proximate analysis and modified organosolv extraction results.[1]

Component	Control	Drought	Cedar Test Sample						
	wt%	wt%	wt%						
			1	2	3	4	5	6	Average[9]
NDF extractible[2]	15.6	17.6	11.7	11.3	10.6	12.0			11.4 ± 1.4
NDF[3]	13.2	11.1	5.1	4.5	4.6	6.0			5.0 ± 1.5
ADF[4]	44.7	44.7	51.9	52.8	52.6	51.4			52.2 ± 1.4
ADF Lignin[5]	25.9	25.9	30.9	31.2	31.3	30.0			30.8 ± 1.2
Ash[6]	0.6	0.8	0.5	0.1	1.0	0.6			0.6 ± 0.9
Organosolv Lignin[7]	3.7	4.9					5.9	5.6	5.8 ± 0.3
Relative Recovery[8]	14.3	18.8					19.1	18.2	18.7 ± 0.9

[1]Results reported on a 70°C oven dry basis.
[2]Protein, starch, waxes, polar & non-polar extractable.
[3]Hemicellulose
[4]Cellulose + acid soluble lignin
[5]Acid resistant lignin
[6]Residue after ashing
[7]Modified organic solvent lignin extraction
[8]Lignin
[9]Average of four test runs for sequential NDF, ADF, acid hydrolysis treatment, average of two test runs for organosolv extraction

C2 FT-IR Spectroscopy

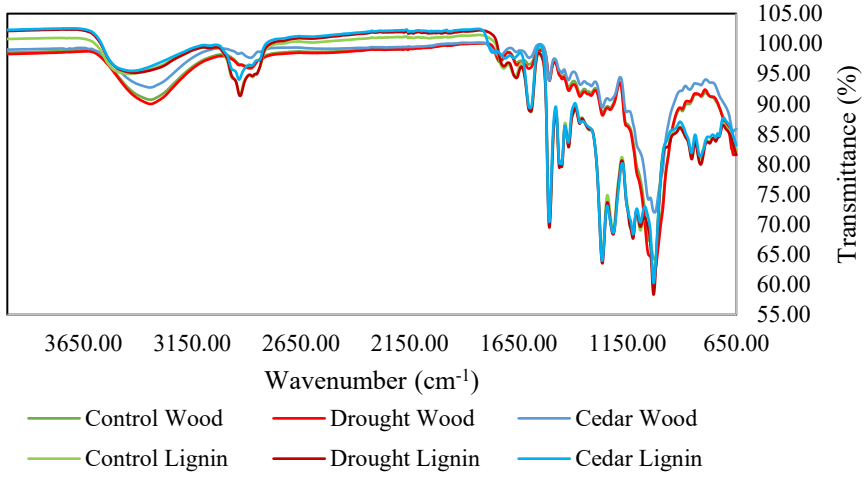


Figure C1. FT-IR spectra of wood and lignin samples in the spectral range 4000 – 650 cm^{-1}.

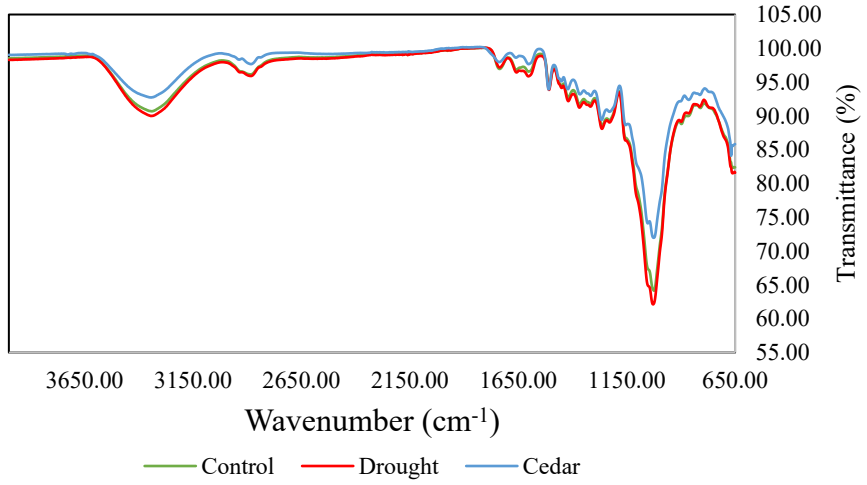


Figure C2. FT-IR spectra of wood samples in the spectral range 4000 – 650 cm^{-1}.

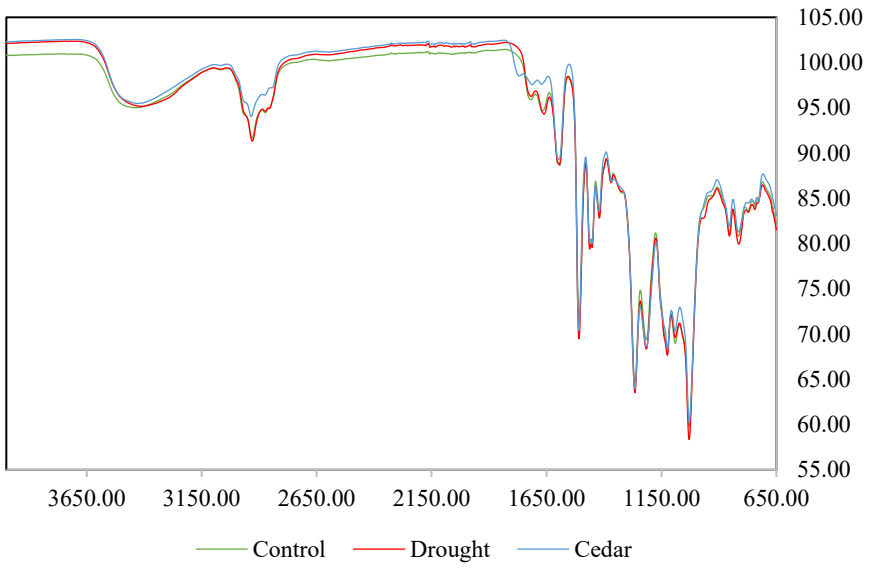


Figure C3. FT-IR of lignin samples in the spectral range 4000 – 650 cm^{-1}.

C3 Solid-state ^{13}NMR spectra

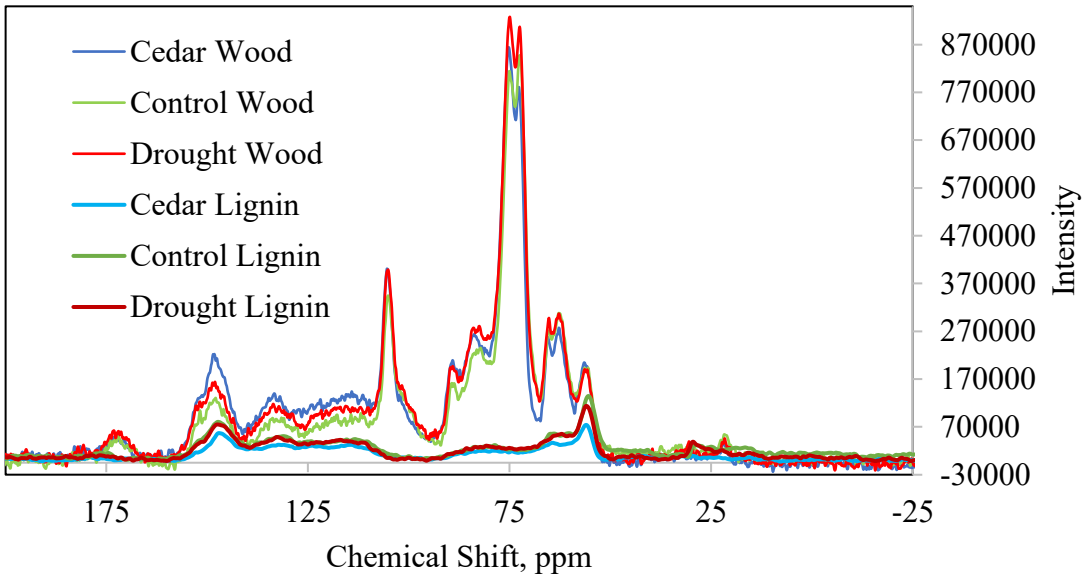


Figure C4. ^{13}C NMR shift of wood and lignin samples. The three wood samples with thinner line width are clearly distinct from the lignin spectra because of their higher signal intensity, particularly in the carbohydrate region (~110 – 58 ppm).

References

[1] L. A. Donaldson, "Abnormal Lignin Distribution in Wood from Severely Drought Stressed Pinus Radiata Trees," *IAWA Journa,* vol. 23, no. 2, pp. 161-178, 2002.

[2] L. Fan, R. Linker, S. Gepstein, E. Tanimoto, R. Yamamoto and P. M. Neumann, "Progressive Inhibition by Water Deficit of Cell Wall Extensibility and Growth along the Elongation Zone of Maize Roots Is Related to Increased Lignin Metabolism and Progressive Stelar Accumulation of Wall Phenolics," *Plant Physiology,* vol. 140, pp. 603-612, 2006.

[3] B.-R. Lee, K.-Y. Kim, W.-J. Jung, J.-C. Avice, A. Ourry and T.-H. Kim, "Peroxidases and lignification in relation to the intensity of water-deficit stress in white clover (Trifolium repens L.)," *Journal of Experimental Botany,* vol. 58, no. 6, pp. 1271-1279, 2007.

[4] Y. Hu, W.-C. Li, Y.-Q. Xu, G.-J. Li, Y. Liao and F.-L. Fu, "Differential expression of candidate genes for lignin biosynthesis under drought stress in maize leaves," *Journal of Applied Genetics,* vol. 50, no. 3, pp. 213-223, 2009.

[5] U. Moura-Sobczak, U. Souza and P. Mazzafera, "Drought stress and changes in the lignin content and composition in Eucalyptus," in *BMC Proceedings*, Arraial d'Ajuda, Bahia, Brazil, 2011.

[6] Z. Li, Y. Peng and X. Ma, "Different response on drought tolerance and post-drought recovery between the small-leafed and the large-leafed white clover (Trifolium repens L.) associated with antioxidative enzyme protection and lignin metabolism," *Acta Physiol Plant,* vol. 35, pp. 213-222, 2013.

[7] R. Terzi, N. S. Guler, N. K. Caliskan and A. Kadioglu, "Lignification response for rolled leaves of Ctenanthe setosa under long-term drought stress," *Turkish Journal of Biology,* vol. 37, pp. 614-619, 2013.

[8] A. B. dos Santos, A. Bottcher, E. Kiyota, J. L. S. Mayer, R. Vincentini, M. d. S. Brito, S. Creste, M. G. Landell and P. Mazzafera, "Water Stress Alters Lignin Content and Related Gene Expression in Two Sugarcane Genotypes," *Journal of Agricultural and Food Chemistry,* vol. 63, pp. 4708-4720, 2015.

[9] S. Srivastava, R. K. Vishwakarma, Y. A. Arafat, S. K. Gupta and B. M. Khan, "Abiotic stress induces change in Cinnamoyl CoA Reductase (CCR) protein abundance and lignin deposition in developing seedlings of Leucaena leucocephala," *Physiology and Molecular Biology of Plants,* vol. 21, no. 2, pp. 197-205, 2015.

[10] C. Xu, X. Fu, R. Liu, L. Guo, L. Ran, C. Li, Q. Tian, B. Jiao, B. Wang and K. Luo, "PtoMYB170 positively regulates lignin deposition during wood formation in poplar and confers drought tolerance in transgenic Arabidopsis," *Tree Physiology,* vol. 37, pp. 1713-1726, 2017.

[11] T. Van der Weijde, L. M. Huxley, S. Hawkins, E. H. Sembiring, K. Farrar, O. Dolstra, R. G. Visser and L. M. Trindade, "Impact of drought stress on growth and quality of miscanthus for biofuel production," *Global Change Biology Bioenergy,* vol. 9, pp. 770-782, 2017.

[12] H. Wildhagen, S. Paul, M. Allwright, H. K. Smith, M. Malinowska, S. K. Schnabel, M. J. Paulo, F. Cattonaro, V. Vendramin, S. Scalabrin, D. Janz, C. Douthe, O. Brendel, C. Buré, D. Cohen, I. Hummel, D. Le Thiec, F. van Eeuwijk, J. J. Keurentjes, J. Flexas, M. Morgante, P. Robson, M.-B. Bogeat-Triboulot, G. Taylor and A. Polle, "Genes and gene clusters related to genotype and drought-induced variation in saccharification potential, lignin content and wood anatomical traits in Populus nigra," *Tree Physiology,* vol. 38, pp. 320-339, 2017.

[13] J. Yan, A. Aznar, C. Chalvin, D. S. Birdseye, E. E. Baidoo, A. Eudes, P. M. Shih, D. Loque, A. Zhang and H. V. Scheller, "Increased drought tolerance in plants engineered for low lignin and low xylan content," *Biotechnology for Biofuels,* vol. 11, no. 195, 2018.

[14] T. R. Lima, E. C. Carvalho, F. R. Martins, R. S. Oliveira, R. S. Miranda, C. S. Müller, L. Pereira, P. R. Bittencourt, J. C. Sobczak, E. Gomes-Filho, R. C. Costa and F. S. Araújo, "Lignin composition is related to xylem embolism resistance and leaf life span in trees in a tropical semiarid climate," *New Phytologist,* vol. 219, pp. 1252-1262, 2018.

[15] S. W. Bang, D.-K. Lee, H. Jung, P. J. Chung, Y. S. Kim, Y. D. Choi, J.-W. Suh and J.-K. Kim, "Overexpression of OsTF1L, a rice HD-Zip transcriptionfactor, promotes lignin biosynthesis and stomatal closurethat improves drought tolerance," *Plant Biotechnology Journal,* vol. 17, pp. 118-131, 2019.

[16] M. Tu, X. Wang, W. Yin, Y. Wang, Y. Li, G. Zhang, Z. Li, J. Song and X. Wang, "Grapevine VlbZIP30 improves drought resistance by directly activating VvNAC17 and promoting lignin biosynthesis through the regulation of three peroxidase genes," *Horticulture Research,* vol. 7, no. 150, 2020.

[17] W. Liu, Y. Jiang, C. Wang, L. Zhao, Q. Xing, M. Li, T. Lv and H. Qi, "Lignin synthesized by CmCAD2 and CmCAD3 in oriental melon (Cucumis melo L.) seedlings contributes to drought tolerance," *Plant Molecular Biology,* vol. 103, pp. 689-704, 2020.

[18] W. Xu, W. Tang, C. Wang, L. Ge, J. Sun, X. Qi, Z. He, Y. Zhou, J. Chen, Z. Xu, Y.-Z. Ma and M. Chen, "SiMYB56 Confers Drought Stress Tolerance in Transgenic Rice by Regulating Lignin Biosynthesis and ABA Signaling Pathway," *Frontiers in Plant Science,* vol. 11, no. 785, 2020.

[19] T. Li, Y. Huang, A. Khadr, Y.-H. Wang, Z.-S. Xu and A.-S. Xiong, "DcDREB1A, a DREB-binding transcription factor from Daucus carota, enhances drought tolerance in transgenic Arabidopsis thaliana and modulates lignin levels by regulating lignin-biosynthesis-related genes," *Environmental and Experimental Botany,* vol. 169, p. 103896, 2020.

[20] N. K. Sharma, S. K. Gupta, V. Dwivedi and D. Chattopadhyay, "Lignin deposition in chickpea root xylem under drought," *Plant Signaling & Behavior,* vol. 15, no. 6, p. e1754621, 2020.

[21] H. Gu, Y. Wang, H. Xie, C. Qiu, S. Zhang, J. Xiao, H. Li, L. Chen, X. Li and Z. Ding, "Drought stress triggers proteomic changes involving lignin, favonoids and fatty acids in tea plants," *Nature Scientific Reports,* vol. 10, p. 15504, 2020.

[22] L. Li, K. Yang, S. Wang, Y. Lou, C. Zhu and Z. Gao, "Genome-wide analysis of laccase genes in moso bamboo highlights PeLAC10 involved in lignin biosynthesis and in response to abiotic stresses," *Plant Cell Reports,* vol. 39, pp. 751-763, 2020.

[23] C. Hori, X. Yu, J. C. Mortimer, R. Sano, T. Matsumoto, J. Kikuchi, T. Demura and M. Ohtani, "Impact of abiotic stress on the regulation of cell wall biosynthesis in Populus trichocarpa," *Plant Biotechnology,* vol. 37, pp. 273-283, 2020.

[24] C. Shumin, C. Huang, L. Luo, S. Zheng, Y. Zhong, J. Sun, J. Gui and L. Li, "Cell-Specific Supression of 4-Coumarate-CoA Ligase Gene Reveals Differential Effect of Lignin on Cell Physiological Function in Populus," *Frontiers in Plant Science,* vol. 11, p. 17, 2020.

[25] D. Zhao, Y. Luan, W. Shi, X. Zhang, J. Meng and J. Tao, "A Paeonia ostii caffeoyl-CoA O-methyltransferase confers drought stress tolerance by promoting lignin synthesis and ROS scavenging," *Plant Science,* vol. 303, p. 110765, 2021.

[26] Y. Yan, P. Wang, Y. Lu, Y. Bai, Y. Wei, G. Liu and H. Shi, "MeRAV5 promotes drought stress resistance in cassava by modulating hydrogen peroxide and lignin accumulation," *The Plant Journal,* vol. 107, pp. 847-860, 2021.

[27] W. Wen, R. Wang, L. Su, A. Lv, P. Zhou and Y. An, "MsWRKY11, activated by MsWRKY22, functions in drought tolerance and modulates lignin biosynthesis in alfalfa (Medicago sativa L.)," *Environmental and Experimental Botany,* vol. 184, p. 104373, 2021.

[28] H. Liu, J. Gao, J. Sun, S. Li, B. Zhang, Z. Wang, C. Zhou, D. B. Sulis, J. P. Wang, V. L. Chiang and W. Li, "Dimerization of PtrMYB074 and PtrWRKY19 mediates transcriptional activation of PtrbHLH186 for secondary xylem development in Populus trichocarpa," *New Phytologist,* vol. 234, pp. 918-933, 2022.

[29] S. Roussel, S. Preys, F. Chauchard and J. Lallemand, "Chapter 2 Multivariate Data Analysis (Chemometrics)," in *In: O'Donnell C., Fagan C., Cullen P. (eds) Process Analytical Technology for the Food Industry. Food Engineering Series.*, New York, Springer, 2014, pp. 7-60.

[30] R. G. Brereton, "Chapter 4 - Preprocessing," in *Chemometrics for Pattern Recognition* , John Wiley & Sons, Ltd, 2009, pp. 107-176.

[31] K. Varmuza and P. Filzmoser, Introduction to Multivariate Statistical Analysis in Chemometrics, CRC Press, 2009.

[32] R. G. Brereton, Applied Chemometrics for Scientists, Chichester, England: John Wiley & Sons, 2007.

[33] P. Ebrahimi, N. Vierkeck, R. Bro and S. B. Engelsen, "Chemometric Analysis of NMR Spectra," in *In: Webb G. (eds) Modern Magnetic Resonance*, Springer Cham, 2017.

[34] J.-M. Roger, J.-C. Boulet, M. Zeaiter and D. N. Rutledge, "3.01 - Pre-processing Methods," in *Comprehensive Chemometrics (Second Edition)*, Elsevier, 2020, pp. 1-75.

[35] K. M. Pierce, B. Kehimkar, L. C. Marney, J. C. Hoggard and R. E. Synovec, "Review of chemometric analysis techniques for comprehensive two dimensional separations data," *Journal of Chromatography A,* vol. 1255, pp. 3-11, 2012.

[36] J. Sandak, A. Sandak and R. Meder, "Assessing trees, wood and derived products with near infrared spectroscopy: hints and tips," *Journal of Near Infrared Spectroscopy,* vol. 24, pp. 485-505, 2016.

[37] A. Savitzky and M. J. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry,* vol. 36, no. 8, pp. 1627-1639, 1964.

[38] J. Steinier, Y. Termonia and J. Deltour, "Comments on Smoothing and Differentiation of Data by Simplified Least Square Procedure," *Analytical Chemistry,* vol. 44, no. 11, pp. 1906-1909, 1972.

[39] G. Vivo´-Truyols and P. J. Schoenmakers, "Automatic Selection of Optimal Savitzky-Golay Smoothing," *Analytical Chemistry,* vol. 78, pp. 4598-4608, 2006.

[40] R. W. Schafer, "What is a Savitzky-Golay filter?," *IEEE Signal processing magazine,* vol. 28, no. 4, pp. 111-117, 2011.

[41] M. Otto, "3 Signal Processing and Time Series Analysis," in *Chemometrics: statisticsand computer application in analytical chemistry, 3rd ed*, John Wiley & Sons, 2016, pp. 55-92.

[42] K. H. Norris and P. C. Williams, "Optimization of mathematical treatments of raw near-infrared signal in the measurement of protien in hard red spring wheat. I. Influence of particle size.," *Cereal Chemistry,* vol. 61, no. 2, pp. 158-165, 1984.

[43] H. Martens, J. P. Nielsen and S. B. Engelsen, "Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures," *Analytical Chemistry,* vol. 75, pp. 394-404, 2003.

[44] N. K. Afseth and A. Kohler, "Extended multiplicative signal correction in vibrational spectroscopy, a tutorial," *Chemometrics and Intelligent Laboratory Systems,* vol. 117, pp. 92-99, 2012.

[45] M. S. Dhanoa, S. J. Lister, R. Sanderson and R. J. Barnes, "The link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) transformations of NIR spectra," *Journal of Near Infrared Spectroscopy,* vol. 2, pp. 43-47, 1994.

[46] R. J. Barnes, M. S. Dhanoa and S. J. Lister, "Standard Normal Variate Transformation and De-trending Near-Infrared Diffuse Reflectance Spectra," *Applied Spectroscopy,* vol. 43, no. 5, pp. 772-777, 1989.

[47] S. Katoch, S. S. Chauchan and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimedia Tools and Applications,* vol. 80, pp. 8091-8126, 2021.

[48] S. Mirjalili, "Genetic Algorithm," in *Evolutionary Algorithms and Neural Networks: Theory and Applications*, Springer, 2019, pp. 43-53.

[49] R. Leardi, R. Boggia and M. Terrile, "Genetic Algorithms as a Stretegy for Feature Selection," *Journal of Chemometrics,* vol. 6, pp. 267-281, 1992.

[50] R. Leardi and A. L. Gonzalez, "Genetic algorithms applied to feature selection in PLS regression: how and when to use them," *Chemometrics and Intelligent Laboratory Systems,* vol. 41, pp. 195-207, 1998.

[51] S. Wold, H. Antti, F. Lindgren and J. Ohman, "Orthogonal signal correction of near-infrared spectra," *Chemometrics and intelligent laboratory systems,* vol. 44, pp. 175-185, 1998.

[52] J. Trygg and S. Wold, "Orthogonal projections to latent structures (O-PLS)," *Journal of Chemometrics,* vol. 16, pp. 119-128, 2002.

[53] L. Wallbacks, U. Edlund and B. Norden, "Multivariate characterization of pulp using solid-state 13C NMR, FTIR, and NIR," *Tappi Journal,* pp. 201-206, 1991.

[54] M. A. Sanderson, F. Agblevor, M. Collins and D. K. Johnson, "Compositional Analysis of Biomass Feedstocks by Near Infrared Reflectance Spectroscopy," *Biomass and Bioenergy,* vol. 11, no. 5, pp. 365-370, 1996.

[55] P. Fardim, M. M. Ferreira and N. Duran, "Multivariate Calibration for Quantitative Analysis of Eucalypt Kraft Pulp by NIR Spectrometry," *Journal of Wood Chemistry and Technology,* vol. 22, no. 1, pp. 67-81, 2002.

[56] S. S. Kelley, J. Jellison and B. Goodell, "Use of NIR and pyrolysis-MBMS coupled with multivariate analysis for detecting the chemical changes associated with brown-rot biodegradation of spruce wood," *FEMS Microbiology Letters,* vol. 209, pp. 107-111, 2002.

[57] S. S. Kelley, R. M. Rowell, M. Davis, C. K. Jurich and R. Ibach, "Rapid analysis of the chemical composition of agricultural fibers using near infrared spectroscopy and pyrolysis molecular beam mass spectrometry," *Biomass and Bioenergy,* vol. 27, pp. 77-88, 2004.

[58] F. S. Poke, J. K. Wright and C. A. Raymond, "Predicting Extractives and Lignin Contents in Eucalyptus globulus Using Near Infrared Reflectance Analysis," *Journal of Wood Chemistry and Technology,* vol. 2004, no. 1, pp. 55-67, 2004.

[59] T.-F. Yeh, H.-m. Chang and J. F. Kadla, "Rapid Prediction of Solid Wood Lignin Content Using Transmittance Near-Infrared Spectroscopy," *Journal of Agricultural and Food Chemistry,* vol. 52, pp. 1435-1439, 2004.

[60] J. Rodrigues, A. Alves, H. Pereira, D. d. S. Perez, G. Chantre and M. Schwanninger, "NIR PLSR results obtained by calibration with noisy, low-precision reference values: Are the results acceptable?," *Holzforschung,* vol. 60, pp. 402-408, 2006.

[61] F. S. Poke and C. A. Raymond, "Predicting Extractives, Lignin, and Cellulose Contents Using Near Infrared Spectroscopy on Solid Wood in Eucalyptus globulus," *Journal of Wood Chemistry and Technology,* vol. 26, no. 2, pp. 187-199, 2006.

[62] T. Yamada, T.-F. Yeh, H.-M. Chang, L. Li, J. F. Kadla and V. L. Chiang, "Rapid analysis of transgenic trees using transmittance near-infrared spectroscopy," *Holzforschung,* vol. 60, pp. 24-28, 2006.

[63] A. Alves, M. Schwanninger, H. Pereira and J. Rodrigues, "Calibration of NIR to assess lignin composition (H/G ratio) in maritime pine wood using analytical pyrolysis as the reference method," *Holzforschung,* vol. 60, pp. 29-31, 2006.

[64] A. Alves, A. Santos, D. d. S. Perez, J. Rodrigues, H. Pereira, R. Simoes and M. Schwanninger, "NIR PLSR model selection for Kappa number prediction of maritime pine Kraft pulps," *Wood science and technology,* vol. 41, pp. 491-499, 2007.

[65] T.-J. Toivanen and R. Alen, "A FTIR/PLS method for determining variations in the chemical composition of birch (Betula pendula/B. pubescens) stem wood," *Appita : Technology, Innovation, Manufacturing, Environment,* vol. 60, no. 2, pp. 155-160, 2007.

[66] S. Jin and H. Chen, "Near-infrared analysis of the chemical composition of rice straw," *Industrial Crops and Products,* vol. 26, pp. 207-211, 2007.

[67] B. Stefke, E. Windeisen, M. Schwanninger and B. Hinterstoisser, "Determination of the Weight Percentage Gain and of the Acetyl Group Content of Acetylated Wood by Means of Different Infrared Spectroscopic Methods," *Analytical Chemistry,* vol. 80, no. 4, pp. 1272-1279, 2008.

[68] B. Uner, I. Karaman, H. Tanriverdi and D. Ozdemir, "Prediction of Lignin and Extractive Content of Pinus nigra Arnold. var. Pallasiana Tree Using Near Infrared Spectroscopy and Multivariate Calibration," *Journal of Wood Chemistry and Technology,* vol. 29, no. 1, pp. 24-42, 2009.

[69] A. R. Robinson and S. D. Mansfield, "Rapid analysis of poplar lignin monomer composition by a streamlined thioacidolysis procedure and near-infrared reflectance-based prediction modeling," *The Plant Journal,* vol. 58, no. 4, pp. 706-714, 2009.

[70] C. Krongtaew, K. Messner, T. Ters and K. Fackler, "Characterization of Key Parameters for Biotechnological Lignocellulose Conversion Assessed by FT-NIR Spectroscopy. Part I: Qualitative Analysis of Pretreated Straw," *BioResources,* vol. 5, no. 4, pp. 2063-2080, 2010.

[71] Y. Horikawa, T. Imai, R. Takada, T. Watanabe, K. Takabe, Y. Kobayashi and J. Sugiyama, "Near-Infrared Chemometric Approach to Exhaustive Analysis of Rice Straw Pretreated for Bioethanol Conversion," *Applied Biochemistry and Biotechnology,* vol. 164, pp. 194-203, 2011.

[72] C. Lepoittevin, J.-P. Rousseau, A. Guillemin, C. Gauvrit, F. Besson, F. Hubert, D. d. S. Perez, L. Harvengt and C. Plomion, "Genetic parameters of growth, straightness and wood chemistry traits in Pinus pinaster," *Annals of Forest Science,* vol. 68, pp. 873-884, 2011.

[73] B. Uner, I. Karaman, H. Tanriverdi and D. Ozdemir, "Determination of lignin and extractive content of Turkish Pine (Pinus brutia Ten.) trees using near infrared spectroscopy and multivariate calibration," *Wood Science and Technology,* vol. 45, pp. 121-134, 2011.

[74] M. Schwanninger, J. C. Rodrigues, N. Gierlinger and B. Hinterstoisser, "Determination of lignin content in Norway spruce wood by Fourier transformed near infrared spectroscopy and partial least infrared spectroscopy and partial least squares regression. Part 1: Wavenumber selection and evaluation of the selected range," *Journal of Near Infrared Spectroscopy,* vol. 19, pp. 319-329, 2011.

[75] M. Schwanninger, J. C. Rodrigues, N. Gierlinger and B. Hinterstoisser, "Determination of lignin content in Norway spruce wood by Fourier transformed near infrared spectroscopy and partial least squares regression analysis. Part 2: Development and evaluation of the final model," *Journal of Near Infrared Spectroscopy,* vol. 19, pp. 331-341, 2011.

[76] A. Alves, R. Simoes, D. J. Stackpole, R. E. Vaillancourt, B. M. Potts, M. Schwanninger and J. Rodrigues, "Determination of the syringyl/guaiacyl ratio of Eucalyptus globulus wood lignin by near infrared-based partial least squares regression models using analytical pyrolysis as the reference method," *Journal of Near Infrared Spectroscopy,* vol. 19, pp. 343-348, 2011.

[77] D. J. Stackpole, R. E. Vaillancourt, A. Alves, J. Rodrigues and B. M. Potts, "Genetic Variation in the Chemical Components of Eucalyptus globulus Wood," *Genes | Genomes | Genetics,* vol. 1, no. 2, pp. 151-159, 2011.

[78] A. Baum, J. Agger, A. S. Meyer, M. Egebo and J. D. Mikkelsen, "Rapid near infrared spectroscopy for prediction of enzymatic hydrolysis of corn bran after various pretreatments," *New Biotechnology,* vol. 29, no. 3, pp. 1871-6784, 2012.

[79] Y. Horikawa, T. Imai, R. Takada, T. Watanabe, K. Takabe, Y. Kobayashi and J. Sugiyama, "Chemometric Analysis with Near-Infrared Spectroscopy for Chemically Pretreated Erianthus toward Efficient Bioethanol Production," *Applied Biochemistry and Biotechnology,* vol. 166, pp. 711-721, 2012.

[80] A. Alves, A. Santos, P. Rozenberg, L. E. Paques, J.-P. Charpentier, M. Schwanninger and J. Rodrigues, "A common near infrared—based partial least squares regression model for the prediction of wood density of Pinus pinaster and Larix x eurolepis," *Wood Science and Technology,* vol. 46, pp. 157-175, 2012.

[81] A. M. Alves, R. F. Simoes, C. A. Santos, B. M. Potts, J. Rodrigues and M. Schwanninger, "Determination of Eucalyptus globulus wood extractives content by near infrared-based partial least squares regression models: comparison between extraction procedures," *Journal of Near Infrared Spectroscopy,* vol. 20, pp. 275-285, 2012.

[82] W. He and H. Hu, "Rapid Prediction of Different Wood Species Extractives and Lignin Content Using Near Infrared Spectroscopy," *Journal of Wood Chemistry and Technology,* vol. 33, pp. 52-64, 2013.

[83] M. K. Rambo, E. P. Amorim and M. M. Ferreira, "Potential of visible-near infrared spectroscopy combined with chemometrics for analysis of some constituents of coffee and banana residues," *Analytica Chimica Acta,* vol. 775, pp. 41-49, 2013.

[84] B. K. Via, C. Zhou, G. Acquah, W. Jiang and L. Eckhardt, "Near Infrared Spectroscopy Calibration for Wood Chemistry: Which Chemometric Technique Is Best for Prediction and Interpretation?," *Sensors,* vol. 14, no. 8, pp. 13532-13547, 2014.

[85] D. S. Ferreira, R. J. Poppi and J. A. L. Pallone, "Evaluation of dietary fiber of Brazilian soybean (Glycine max) using near-infrared spectroscopy and chemometrics," *Journal of Cereal Science,* vol. 64, pp. 43-47, 2015.

[86] M. N. Gebreselassie, K. Ader, N. Boizot, F. Millier, J.-P. Charpentier, A. Alves, R. Simoes, J. C. Rodrigues, G. Bodineau, F. Fabbrini, M. Sabatti, C. Bastien and V. Segura, "Near-infrared spectroscopy enables the genetic analysis of chemical properties in a large set of wood samples from Populus nigra (L.) natural populations," *Industrial Crops & Products,* vol. 107, pp. 159-171, 2017.

[87] C. Assis, R. S. Ramos, L. A. Silva, V. Kist, M. H. Barbosa and R. F. Teofilo, "Prediction of Lignin Content in Different Parts of Sugarcane Using Near-Infrared Spectroscopy (NIR), Ordered Predictors Selection (OPS), and Partial Least Squares (PLS)," *Applied Spectroscopy,* vol. 71, no. 8, pp. 2001-2012, 2017.

[88] T. Van der Weijde, L. M. Huxley, S. Hawkins, E. H. Sembiring, K. Farrar, O. Dolstra, R. G. Visser and L. M. Trindade, "Impact of drought stress on growth and quality of miscanthus for biofuel production," *GCB Bioenergy,* vol. 9, pp. 770-782, 2017.

[89] L. M. Fahey, M. K. Nieuwoudt and P. J. Harris, "Using near infrared spectroscopy to predict the lignin content and monosaccharide compositions of Pinus radiata wood cell walls," *International Journal of Biological Macromolecules,* vol. 113, pp. 507-514, 2018.

[90] O. Elle, R. Richter, M. Vohland and A. Weigelt, "Fine root lignin content is well predictable with near-infrared spectroscopy," *Scientific Reports,* vol. 9, no. 6396, 2019.

[91] H. Yang, Y. Liu, Z. Xiong and L. Liang, "Rapid Determination of Holocellulose and Lignin in Wood by Near Infrared Spectroscopy and Kernel Extreme Learning Machine," *Analytical Letters,* vol. 53, no. 7, pp. 1140-1154, 2020.

[92] L. Liang, L. Wei, G. Fang, F. Xu, Y. Deng, K. Shen, Q. Tian, T. Wu and B. Zhu, "Prediction of holocellulose and lignin content of pulp wood feedstock using near infrared spectroscopy and variable selection," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* vol. 225, p. 117515, 2020.

[93] X. Sheng, Z. Li, Z. Li, J. Dong, J. Wang and J. Yin, "Nondestructive determination of lignin content in Korla fragrant pear based on near-infrared spectroscopy," *Spectroscopy Letters,* vol. 53, no. 4, pp. 306-314, 2020.

[94] E. J. Wolfrum, C. Payne, A. Schwartz, J. Jacobs and R. W. Kressin, "A Performance Comparison of Low-Cost Near-Infrared (NIR) Spectrometers to a Conventional Laboratory Spectrometer for Rapid Biomass Compositional Analysis," *BioEnergy Research,* vol. 13, pp. 1121-1129, 2020.

[95] X. Wu, G. Li, X. Liu and F. He, "Rapid non-destructive analysis of lignin using NIR spectroscopy and chemo-metrics," *Food and Energy Security,* p. e289, 2021.

[96] V. Ponnuchamy, O. Gordobil, R. H. Diaz, A. Sandak and J. Sandak, "Fractionation of lignin using organic solvents: A combined experimental and theoretical study," *International Journal of Biological Macromolecules,* vol. 168, pp. 792-805, 2021.

[97] S.-W. Hwang, U. T. Hwang, K. Jo, T. Lee, J. Park, J.-C. Kim, H. W. Kwak, I.-G. Choi and H. Yeo, "NIR-chemometric approaches for evaluating carbonization characteristics of hydrothermally carbonized lignin," *Scientific Reports,* vol. 11, p. 16979, 2021.

[98] T. P. Schultz, M. C. Templeton and G. D. McGinnis, "Rapid Determination of Lignocellulose by Diffuse Reflectance Fourier Transform Infrared Spectrometry," *Analytical Chemistry,* vol. 57, pp. 2867-2869, 1985.

[99] O. Faix and J. H. Bottcher, "Determination of Phenolic Hydroxyl Group Contents in Milled Wood Lignins by FTIR Spectroscopy Applying Partial Least-Squares (PLS) and Principal Components Regression (PCR)," *Holzforschung,* vol. 47, pp. 45-49, 1993.

[100] A. R. Cotrim, A. Ferraz, A. R. Goncalves, F. T. Silva and R. E. Bruns, "Identifying the origin of lignins and monitoring their structrual changes by means of FTIR-PCA and -SIMCA," *Bioresource Technology,* vol. 68, pp. 29-34, 1999.

[101] R. Meder, S. Gallagher, K. L. Mackie, H. Bohler and R. R. Meglen, "Rapid Determination of the Chemical Composition and Density of Pinus radiata by PLS Modelling of Transmission and Diffuse Reflectance FTIR Spectra," *Holzforschung,* vol. 53, pp. 261-266, 1999.

[102] A. Ferraz, J. Baeza, J. Rodriguez and J. Freer, "Estimating the chemical composition of biodegraded pine and eucalyptus wood by DRIFT spectroscopy and multivariate analysis," *Bioresource Technology,* vol. 74, pp. 201-212, 2000.

[103] J. Rodrigues, J. Puls, O. Faix and H. Pereira, "Determination of Monosaccharide Composition of Eucalyptus globulus Wood by FTIR Spectroscopy," *Holzforschung,* vol. 55, pp. 265-269, 2001.

[104] M. P. Tucker, Q. A. Nguyen, F. P. Eddy, K. L. Kadam, L. M. Gedvilas and J. D. Webb, "Fourier Transform Infrared Quantitative Analysis of Sugars and Lignin in Pretreated Softwood Solid Residues," in *Twenty-Second Symposium on Biotechnology for Fuels and Chemicals*, Humana Press Inc., 2001, pp. 51-61.

[105] S. Bjarnestad and O. Dahlman, "Chemical Compositions of Hardwood and Softwood Pulps Employing Photoacoustic Fourier Transform Infrared Spectroscopy in Combination with Partial Least-Squares Analysis," *Analytical Chemistry,* vol. 74, no. 22, pp. 5851-5858, 2002.

[106] H. Sivonen, M. Nuopponen, S. L. Maunu, F. Sundholm and T. Vuorinen, "Carbon-Thirteen Cross-Polarization Magic Angle Spinning Nuclear Magnetic Resonance and Fourier Transform Infrared Studies of Thermally Modified Wood Exposed to Brown and Soft Rot Fungi," *Applied Spectroscopy,* vol. 57, no. 3, pp. 266-273, 2003.

[107] C. G. Boeriu, D. Bravo, R. J. Gosselink and J. E. van Dam, "Characterisation of structure-dependent functional properties of lignin with infrared spectroscopy," *Industrial Crops and Products an International Journal,* vol. 20, pp. 205-218, 2004.

[108] N. Labbe, T. G. Rials, S. S. Kelley, Z.-M. Cheng, J.-Y. Kim and Y. Li, "FT-IR imaging and pyrolysis-molecular beam mass spectrometry: new tools to investigate wood tissues," *Wood Sci Technol,* vol. 39, pp. 61-77, 2005.

[109] J. C. del Rio, A. Gutierrez, I. M. Rodriguez, D. Ibarra and A. T. Martinez, "Composition of non-woody plant lignins and cinnamic acids by Py-GC/MS, Py/TMAH and FT-IR," *Journal of Analytical and Applied Pyrolysis,* vol. 79, pp. 39-46, 2007.

[110] V. Q. Dang, N. K. Bhardwaj, V. Hoang and K. L. Nguyen, "Determination of lignin content in high-yield kraft pulps using photoacoustic rapid scan Fourier transform infrared spectroscopy," *Carbohydrate Polymers,* vol. 68, pp. 489-494, 2007.

[111] O. Derkacheva and D. Sukhov, "Investigation of Lignins by FTIR spectroscopy," *Macromolecular Symposia,* vol. 265, pp. 61-68, 2008.

[112] D. G. Mann, N. Labbe, R. W. Sykes, K. Gracom, L. Kline, I. M. Swamidoss, J. N. Burris, M. Davis and C. N. Stewart Jr., "Rapid Assessment of Lignin Content and Structure in Switchgrass (Panicum virgatum L.) Grown Under Different Environmental Conditions," *Bioenergy Resources,* vol. 2, pp. 246-256, 2009.

[113] H. Chen, C. Ferrari, M. Angiuli, J. Yao, C. Raspi and E. Bramanti, "Qualitative and quantitative analysis of wood samples by Fourier transform infrared spectroscopy and multivariate analysis," *Carbohydrate Polymers,* vol. 82, pp. 772-778, 2010.

[114] R. Rana, R. Langenfeld-Heyser, R. Finkeldey and A. Polle, "FTIR spectroscopy, chemical and histochemical characterisation of wood and lignin of five tropical timber wood species of the family of Dipterocarpaceae," *Wood Science and Technology,* vol. 44, pp. 225-242, 2010.

[115] L. M. Kline, D. G. Hayes, A. R. Womac and N. Labbe, "Simplified determination of lignin content in hard and soft woods via UV-spectrophotometric analysis of biomass dissolved in ionic liquids," *Bio Resources,* vol. 5, no. 3, pp. 1366-1383, 2010.

[116] Y. Tamaki and G. Mazza, "Rapid Determination of Lignin Content of Straw Using Forier Transform Mid-Infrared Spectroscopy," *Journal of agricultural and food chemistry,* vol. 59, no. 2, pp. 504-512, 2011.

[117] G. Zhou, G. Taylor and A. Polle, "FTIR-ATR-based prediction and modelling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars," *Plant Methods,* vol. 7, no. 9, 2011.

[118] Y. Huang, L. Wang, Y. Chao, D. S. Nawawi, T. Akiyama, T. Yokoyama and Y. Matsumoto, "Analysis of Lignin Aromatic Structure in Wood Based on the IR Spectrum," *Journal of Wood Chemistry and Technology,* vol. 32, no. 4, pp. 294-303, 2012.

[119] D. J. Krasznai, P. Champagne and M. F. Cunningham, "Quantitative characterization of lignocellulosic biomass using surrogate mixtures and multivariate techniques," *Bioresource Technology,* vol. 110, pp. 652-661, 2012.

[120] O. Y. Derkacheva, "Estimation of Aromatic Structure Contents in Hardwood Lignins from IR Absorption Spectra," *Journal of Apploed Spectroscopy,* vol. 80, pp. 670-676, 2013.

[121] R. J. Sammons, D. P. Harper, N. Labbe, J. J. Bozell, T. Elder and R. G. Timothy, "Characterization of Organosolv Lignins using Thermal and FT-IR Spectroscopic Analysis," *BioResources,* vol. 8, no. 2, pp. 2752-2767, 2013.

[122] C. G. Boeriu, F. I. Fitigau, R. J. Gosselink, A. E. Frissen, J. Stoutesdijk and F. Peter, "Fractionation of five technical lignins by selective extraction in green solvents and characterisation of isolated fractions," *Industrial Crops and Products,* vol. 62, pp. 481-490, 2014.

[123] C. Zhou, W. Jiang, Q. Cheng and B. K. Via, "Multivariate Calibration and Model Integrity for Wood Chemistry Using Fourier Transform Infrared Spectroscopy," *Journal of Analytical Methods in Chemistry,* 2015.

[124] G. E. Acquah, B. K. Via, O. O. Fasina and L. G. Eckhardt, "Rapid Quantitative Analysis of Forest Biomass Using Fourier Transform Infrared Spectroscopy and Partial Least Squares Regression," *Journal of Analytical Methods in Chemistry,* vol. 2016, no. 1839598, 2016.

[125] X. Li, Y. Wei, J. Xu, N. Xu and Y. He, "Quantitative visualization of lignocellulose components in transverse sections of moso bamboo based on FTIR macro- and micro-spectroscopy coupled with chemometrics," *Biotechnology for Biofuels,* vol. 11, no. 263, 2018.

[126] M. Traoré, J. Kaal and A. M. Cortizas, "Chemometric tools for identification of wood from different oak species and their potential for provenancing of Iberian shipwrecks (16th-18th centuries AD)," *Journal of Archaeological Science,* vol. 100, pp. 62-73, 2018.

[127] E. Lazzari, T. Schena, M. C. A. Marcelo, C. T. Primaz, A. N. Silva, M. F. Ferrão, T. Bjerk and E. B. Caramão, "Classification of biomass through their pyrolytic bio-oil composition using FTIR and PCA analysis," *Industrial Crops & Products,* vol. 111, pp. 856-864, 2018.

[128] C. S. Lancefield, S. Constant, P. d. Peinder and P. C. Bruijnincx, "Linkage Abundance and Molecular Weight Characteristics of Technical Lignins by Attenuated Total Reflection-FTIR Spectroscopy Combined with Multivariate Analysis," *ChemSusChem,* vol. 12, no. 6, pp. 1139-1146, 2019.

[129] Y. Horikawa, S. Hirano, A. Mihashi, Y. Kobayashi, S. Zhai and J. Sugiyama, "Prediction of Lignin Contents from Infrared Spectroscopy: Chemical Digestion and Lignin/Biomass Ratios of Cryptomeria japonica," *Applied Biochemistry and Biotechnology,* vol. 188, pp. 1066-1076, 2019.

[130] M. Bergs, G. Volkering, T. Kraska, R. Pude, X. T. Do, P. Kusch, Y. Monakhova, C. Konow and M. Schulze, "Miscanthus x giganteus Stem Versus Leaf-Derived Lignins Differing in Monolignol Ratio and Linkage," *International Journal of Molecular Sciences,* vol. 20, p. 1200, 2019.

[131] M. F. Cherif, D. Trache, N. Brosse, F. Benaliouche and A. F. Tarchoun, "Comparison of the Physicochemical Properties and Thermal Stability of Organosolv and Kraft Lignins from Hardwood and Softwood Biomass for Their Potential Valorization," *Waste and Biomass Valorization,* vol. 11, pp. 6541-6553, 2020.

[132] R. Herrera-Diaz, L. Bodi-Paul, O. Gordobil and J. Labidi, "Fast methods for the identification of suitable chemo-enzymatic treatments of Kraft lignin to obtain aromatic compounds," *Biofuels, Bioproducts & Biorefining,* vol. 14, pp. 521-532, 2020.

[133] C.-M. Popescu, D. Jones, D. Krzisnik and M. Humar, "Determination of the effectiveness of a combined thermal/chemical wood modification by the use of FT-IR spectroscopy and chemometric methods," *Journal of Molecular Structure,* vol. 1200, p. 127133, 2020.

[134] J. Rumpf, X. T. Do, R. Burger, M. Schulze, B. Michel, Y. B. Monakhova, B. Diehl, R. Pude and M. Rehahn, "Lignins from Low-Input Crops: Biomass Origin Assignment via Chemometric Data Analysis".

[135] K. N. Khalili, P. de Peinder, P. C. Bruijnincx and B. M. Weckhuysen, "Monitoring Aqueous Phase Reactions by Operando ATR-IR Spectroscopy at High Temperature and Pressure: A Biomass Conversion Showcase," *Chemistry—Methods,* vol. 1, pp. 468-476, 2021.

[136] K. N. Khalili, P. de Peinder, J. Donkers, R. J. Gosselink, P. C. Bruinjnincx and B. M. Weckhuysen, "Monitoring Molecular Weight Changes during Technical Lignin Depolymerization by Operando Attenuated Total Reflectance Infrared Spectroscopy and Chemometrics," *ChemSusChem,* vol. 14, pp. 5517-5524, 2021.

[137] G. Toscano, M. Vittorio, E. Leoni, P. Stipa, E. Laudadio and S. Sabbatini, "FTIR spectroscopy for determination of the raw materials used in wood pellet production," *Fuel,* vol. 313, p. 123017, 2022.

[138] T. Ona, T. Sonoda, K. Ito, M. Shibata, T. Katayama, T. Kato and Y. Ootake, "Non-Destructive Determination of Lignin Syringyl Quaiacyl Monomeric Composition in Native Wood by Fourier Transform Raman Spectroscopy," *Journal of Wood Chemistry and Technology,* vol. 18, no. 1, pp. 43-51, 1998.

[139] T. Ona, T. Sonoda, J. Ohshima, S. Yokota and N. Yoshizawa, "A Rapid Quantitative Method to Assess Eucalyptus Wood Properties for Kraft Pulp Production by FT-Raman Spectroscopy," *Journal of Pulp and Paper Science,* vol. 29, no. 1, pp. 6-10, 2003.

[140] A.-M. Saariaho, A.-S. Jaaskelainen, M. Nuopponen and T. Vuorinen, "Ultra Violet Resonance Raman Spectroscopy in Lignin Analysis: Determination of Characteristic Vibrations of p-Hydroxyphenyl, Guaiacyl, and Syringyl Lignin Structures," *Applied Spectroscopy,* vol. 57, pp. 58-66, 2003.

[141] A.-M. Saariaho, D. S. Argyropoulos, A.-S. Jaaskelainen and T. Vuorinen, "Development of the partial least squares models for the interpretation of the UV resonance Raman spectra of lignin model compounds," *Vibrational Spectroscopy,* vol. 37, pp. 111-121, 2005.

[142] L. Sun, P. Varanasi, F. Yang, D. Loque, B. A. Simmons and S. Singh, "Rapid determination of syringyl: Guaiacyl ratios using FT-Raman Spectroscopy," *Biotechnology and Bioengineering,* vol. 109, no. 3, pp. 647-656, 2011.

[143] U. P. Agarwal, J. D. McSweeny and S. A. Ralph, "FT–Raman Investigation of Milled-Wood Lignins: Softwood, Hardwood, and Chemically Modified Black Spruce Lignins," *Journal of Wood Chemistry and Technology,* vol. 31, pp. 324-344, 2011.

[144] M. W. Meyer, J. S. Lupoi and E. A. Smith, "1064 nm dispersive multichannel Raman spectroscopy for the analysis of plant lignin," *Analytica Chimica Acta,* vol. 706, pp. 164-170, 2011.

[145] E. Widjaja and M. Garland, "Use of Raman Microscopy and Band-Target Entropy Minimization Analysis To Identify Dyes in a Commercial Stamp. Implications for Authentication and Counterfeit Detection," *Analytical Chemistry,* vol. 80, pp. 729-733, 2008.

[146] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, "Optimization by Simulated Annealing," *Science,* vol. 220, no. 4598, pp. 671-680, 1983.

[147] P. N. Perera, M. Schmidt, V. L. Chiang, P. J. Schuck and P. D. Adams, "Raman-spectroscopy-based noninvasive microanalysis of native lignin structure," *Analytical and Bioanalytical Chemistry,* vol. 402, pp. 983-987, 2012.

[148] J. S. Lupoi and E. A. Smith, "Characterization of Woody and Herbaceous Biomasses Lignin Composition with 1064 nm Dispersive Multichannel Raman Spectroscopy," *Applied Spectroscopy,* vol. 66, no. 8, pp. 903-910, 2012.

[149] J. S. Lupoi, S. Singh, M. Davis, D. J. Lee, M. Shepherd, B. A. Simmons and R. J. Henry, "High-throughput prediction of eucalypt lignin syringyl/guaiacyl content using multivariate analysis: a comparison between mid-infrared, near-infrared, and Raman spectroscopies for model development," *Biotechnology for Biofuels,* vol. 7, p. 93, 2014.

[150] N. Gierlinger, "Revealing chnages in molecular composition of plant cell walls on the micron-level by Raman mapping and vertex component analysis (VCA)," *Frontiers in Plant Science,* vol. 5, no. 306, 2014.

[151] C. J. Colares, T. C. Pastore, V. T. Coradin, J. A. Camargos, A. C. Moreira, J. C. Rubim and J. W. Braga, "Exploratory Analysis of the Distribution of Lignin and Cellulose in Woods by Raman Imaging and Chemometrics," *Journal of the Brazilian Chemical Society,* vol. 26, no. 6, pp. 1297-1305, 2015.

[152] J. S. Lupoi, A. Healey, S. Singh, R. Sykes, M. Davis, D. J. Lee, M. Shepherd, B. A. Simmons and R. J. Henry, "High-Throughput Prediction of Acacia and Eucalypt Lignin Syringyl/Guaiacyl Content Using FT-Raman Spectroscopy and Partial Least Squares Modeling," *Bioenergy Resources,* vol. 8, pp. 953-963, 2015.

[153] B. Liu, P. Wang, J. I. Kim, D. Zhang, Y. Xia, C. Chapple and J.-X. Cheng, "Vibrational Fingerprint Mapping Reveals Spatial Distribution of Functional Groups of Lignin in Plant Cell Wall," *Analytical Chemistry,* vol. 87, no. 18, pp. 9436-9442, 2015.

[154] B. Prats-Mateu, M. T. Hauser, A. Heredia and N. Gierlinger, "Waterproofing in Arabidopsis: Following Phenolics and Lipids In situ by Confocal Raman Microscopy," *Frontiers in Chemistry,* vol. 4, no. 10, 2016.

[155] B. Prats-Mateu, M. Felhofer, A. de Juan and N. Gierlinger, "Multivariate unmixing approaches on Raman images of plant cell walls: new insights or overinterpretation of results?," *Plant Methods,* vol. 14, no. 52, 2018.

[156] X. Zhang, "Visualising lignin quantitatively in plant cell walls by micro-Raman spectroscopy," *RSC Advances,* vol. 11, p. 131124, 2021.

[157] W. Gao, T. Shu, Q. Liu, S. Ling, Y. Guan, S. Liu and L. Zhou, "Predictive Modeling of Lignin Content for the Screening of Suitable Poplar Genotypes Based on Fourier Transform−Raman Spectrometry," *ACS Omega,* vol. 6, pp. 8578-8587, 2021.

[158] I. Liedtke, S. Diehn, Z. Heiner, S. Seifert, S. Obenaus, C. Büttner and J. Kneipp, "Multivariate Raman mapping for phenotypic characterization in plant tissue sections," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* vol. 251, p. 119418, 2021.

[159] W. Gao, T. Shu, Y. Guan, S. Ling, S. Liu and L. Zhou, "Novel strategy for establishment of an FT-Raman spectroscopy based quantitative model for poplar holocellulose content determination," *Carbohydrate Polymers,* vol. 277, p. 118793, 2022.

[160] W. Gao, L. Zhou, Y. Guan, H. Gao and S. Liu, "Monitoring the kappa number of bleached pulps based on FT-Raman spectroscopy," *Cellulose,* vol. 29, pp. 1069-1080, 2022.

[161] L. Wallbacks, U. Edlund and B. Norden, "Multivariate Data Analysis of In Situ Pulp Kinetics," *Journal of Wood Chemistry and Technology,* vol. 9, no. 2, pp. 235-249, 1989.

[162] P. Malkavaara, R. Alen and E. Kolehmainen, "Multivariate correlation between 13C and 31P NMR spectral data on dissolved lignin and the combustion properties of kraft black liqour," *Magnetic Resonance in Chemistry,* vol. 37, pp. 407-412, 1999.

[163] M. Jalali-Hevari, S. Masoum and P. Shahbazikhah, "Simulation of 13C nuclear magnetic resonance spectra of lignin compounds using principal component analysis and artificial neural networks," *Journal of Magnetic Resonance,* vol. 171, no. 1, pp. 176-185, 2004.

[164] M. Hedenstrom, S. Wiklund-Lindstrom, T. Oman, F. Lu, L. Gerber, P. Schatz, B. Sundberg and J. Ralph, "Identification of Lignin and Polysaccharide Modifications in Populus Wood by Chemometric Analysis of 2D NMR Spectra from Dissolved Cell Walls," *Molecular Plant,* vol. 2, no. 5, pp. 933-942, 2009.

[165] G. D. Strahan, C. A. Mullen and A. A. Boateng, "Characterizing Biomass Fast Pyrolysis by 13C NMR and Chemometric Analysis," *Energy Fuels,* vol. 25, pp. 5452-5461, 2011.

[166] E. Gjersing, R. M. Happs, R. W. Sykes, C. Doeppke and M. F. Davis, "Rapid Determination of Sugar Content in Biomass Hydrolysates Using Nuclear Magnetic Resonance Spectroscopy," *Biotechnology and Bioengineering,* vol. 110, no. 3, pp. 721-728, 2012.

[167] L. M. Aguilera-Saez, F. M. Arrabal-Campos, A. J. Callejon-Ferre, M. D. S. Medina and I. Fernandez, "Use of multivariate NMR analysis in the content prediction of hemicellulose, cellulose and lignin in greenhouse crop residues," *Phytochemistry,* vol. 158, pp. 110-119, 2019.

[168] L. G. Raymond, S. J. Hill, W. J. Grigsby and B. R. Bogun, "A chemometric approach for the segregation of bark biomass based on tree height and geographic location," *Journal of Wood Chemistry and Technology,* vol. 40, no. 6, pp. 361-369, 2020.

[169] R. Burger, J. Rumpf, X. T. Do, Y. B. Monakhova, B. W. Diehl, M. Rehahn and M. Schuzle, "Is NMR Combined with Multivariate Regression Applicable for the Molecular Weight Determination of Randomly Cross-Linked Polymers Such as Lignin?," *ACS Omega,* vol. 6, pp. 29516-29524, 2021.

[170] M. Kleen and G. Lindblad, "Quantification of lignin and carbohydrates in kraft pulps using analytical pyrolysis and multivariate data analysis," *Journal of Analytical and Applied Pyrolysis,* vol. 25, pp. 209-227, 1993.

[171] F. A. Agblevor, R. J. Evans and K. D. Johnson, "Molecular-beam mass-spectrometric analysis of lignocellulosic materials I. Herbaceous biomass," *Journal of Analytical and Applied Pyrolysis,* vol. 30, pp. 125-144, 1994.

[172] M. H. Alma and S. S. Kelley, "The application of pyrolysis-molecular beam mass spectrometry for characterization of barkphenolysis products," *Biomass and Bioenergy,* vol. 22, pp. 411-419, 2002.

[173] M. F. Davis, G. A. Tuskan, P. Payne, T. J. Tschaplinski and R. Meilan, "Assessment of Populus wood chemistry following the introduction of a Bt toxin gene," *Tree Physiology,* vol. 26, pp. 557-564, 2006.

[174] E. Mészáros, E. Jakab, G. Várhegyi and P. Tóvári, "Thermogravimetry/Mass Spectrometry Analysis of Energy Crops," *Journal of Thermal Analysis and Calorimetry,* vol. 88, no. 2, pp. 477-482, 2007.

[175] F. Modungo, E. Ribechini, M. Calderisi, G. Giachi and M. P. Colombini, "Analysis of lignin from archaeological waterlogged wood by direct exposure mass spectrometry (DE-MS) and PCA evaluation of mass spectral data," *Microchemical Journal,* vol. 88, pp. 186-193, 2008.

[176] R. Sykes, B. Kodrzycki, G. Tuskan, K. Foutz and M. Davis, "Within tree variability of lignin composition in Populus," *Wood Science Technology,* vol. 42, pp. 649-661, 2008.

[177] R. Sykes, M. Yung, E. Novaes, M. Kirst, G. Peter and M. Davis, "High-Throughput Screening of Plant Cell-Wall Composition Using Pyrolysis Molecular Beam Mass Spectroscopy," in *Jonathan R. Mielenz (ed.), Biofuels: Methods and Protocols, Methods in Molecular Biology, vol. 581*, Humana Press, a part of Springer Science+Business Media, 2009, pp. 169-183.

[178] V. Arantes, Y. Qian, S. S. Kelley, A. M. Milagres, T. R. Filley, J. Jellison and B. Goodell, "Biomimetic oxidative treatment of spruce wood studied by pyrolysis–molecular beam mass spectrometry coupled with multivariate analysis and 13C-labeled tetramethylammonium hydroxide thermochemolysis: implications for fungal degradation of wood," *JBIC Journal of Biological Inorganic Chemistry,* vol. 14, pp. 1253-1263, 2009.

[179] L. Gerber, M. Eliasson, J. Trygg, T. Moritz and B. Sundberg, "Multivariate curve resolution provides a high-throughput data processing pipeline for pyrolysis-gas chromatography/mass spectrometry," *Journal of Analytical and Applied Pyrolysis,* vol. 95, pp. 95-100, 2012.

[180] T. Ohra-aho, F. J. Gomes, J. L. Colodette and T. Tamminen, "S/G ratio and lignin structure among Eucalyptus hybrids determined by Py-GC/MS and nitrobenzene oxidation," *Journal of Analytical and Applied Pyrolysis,* vol. 101, pp. 166-171, 2013.

[181] M. A. Galetta, L. Reina, F. Resquin, C. Mantero, J. A. González-Pérez, G. Almendros, M. P. Menéndez and F. J. González-Vila, "Chemometric appraisal of lignin pyrolytic assemblages from Eucalyptus woods relevant for pulping in Uruguay," *Journal of Analytical and Applied Pyrolysis,* vol. 109, pp. 296-303, 2014.

[182] J. Ponomarenko, T. Dizhbite, M. Lauberts, A. Volperts, G. Dobele and G. Telysheva, "Analytical pyrolysis – A tool for revealing of lignin structure-antioxidant activity relationship," *Journal of Analytical and Applied Pyrolysis,* vol. 113, pp. 360-369, 2015.

[183] S. R. Decker, R. W. Sykes, G. B. Turner, J. S. Lupoi, C. Doepkke, M. P. Tucker, L. A. Schuster, K. Massa, M. E. Himmel, M. F. Davis and E. Gjersing, "High-throughput Screening of Recalcitrance Variations in Lignocellulosic Biomass: Total Lignin, Lignin Monomers, and Enzymatic Sugar Release," *Journal of Visualized Experiments,* vol. 103, 2015.

[184] H. E. Toraman, R. Vanholme, E. Borén, Y. Vanwonterghem, M. R. Djokic, G. Yildiz, F. Ronsse, W. Prins, W. Boerjan, K. M. Van Geem and G. B. Marin, "Potential of genetically engineered hybrid poplar for pyrolytic production of bio-based phenolic compounds," *Bioresource Technology,* vol. 207, pp. 229-236, 2016.

[185] R. B. Madsen, E. Lappa, P. S. Christensen, M. M. Jensen, M. Klemmer, J. Becker, B. B. Iversen and M. Glasius, "Chemometric analysis of composition of bio-crude and aqueous phase from hydrothermal liquefaction of thermally and chemically pretreated Miscanthus x giganteus," *Biomass and Bioenergy,* vol. 95, pp. 137-145, 2016.

[186] H. E. Toraman, V. Abrahamsson, R. Vanholme, R. Van Acker, F. Ronsse, G. Pilate, W. Boerjan, K. M. Van Geem and G. B. Marin, "Application of Py-GC/MS coupled with PARAFAC2 and PLS-DA to study fast pyrolysis of genetically engineered poplars," *Journal of Analytical and Applied Pyrolysis,* vol. 129, pp. 101-111, 2018.

[187] J. Prothmann, P. Spégel, M. Sandahl and C. Turner, "Identification of lignin oligomers in Kraft lignin using ultra-high-performance liquid chromatography/high-resolution multiple-

stage tandem mass spectrometry (UHPLC/HRMSn)," *Analytical and Bioanalytical Chemistry,* vol. 410, pp. 7803-7814, 2018.

[188] A. Trubetskaya, N. Souihi and K. Umeki, "Categorization of tars from fast pyrolysis of pure lignocellulosic compounds at high temperature," *Renewable Energy,* vol. 141, pp. 751-759, 2019.

[189] G. SriBala, H. E. Toraman, S. Symoens, A. Dejardin, G. Pilate, W. Boerjan, F. Ronsse, K. M. Van Geem and G. B. Marin, "Analytical Py-GC/MS of Genetically Modified Poplar for the Increased Production of Bio-aromatics," *Computational and Structural Biotechnology Journal,* vol. 17, pp. 599-610, 2019.

[190] X. Xin, S. Pang, F. d. M. Mercader and K. M. Torr, "The effect of biomass pretreatment on catalytic pyrolysis products of pine wood by Py-GC/MS and principle component analysis," *Journal of Analytical and Applied Pyrolysis,* vol. 138, pp. 145-153, 2019.

[191] J. Reyes-Rivera, E. Solano, T. Terrazas, M. Soto-Hernandez, S. Arias, Y. C. Almanza-Arjona and L. A. Polindara-Garcia, "Classification of lignocellulosic matrix of spines in Cactaceae by Py-GC/MS combined with omic tools and multivariate analysis: A chemotaxonomic approach," *Journal of Analytical and Applied Pyrolysis,* vol. 148, p. 1047956, 2020.

[192] C. Castilla, C. P. Ruger, S. Marcotte, H. Lavanant and C. Afonso, "Direct Inlet Probe Atmospheric Pressure Photo and Chemical Ionization Coupled to Ultrahigh Resolution Mass Spectrometry for the Description of Lignocellulosic Biomass," *Journal of the American Society for Mass Spectrometry,* vol. 31, pp. 822-831, 2020.

[193] A. E. Harman-Ware, D. Macaya-Sanz, C. R. Aberyratne, C. Doeppke, K. Haiby, G. A. Tuskan, B. Stanton, S. P. DiFazio and M. F. Davis, "Accurate determination of genotypic variance of cell wall characteristics of a Populus trichocarpa pedigree using high-throughput pyrolysis-molecular beam mass spectrometry," *Biotechnology for Biofuels,* vol. 14, no. 59, 2021.

[194] M. Wądrzyk, R. Janus, M. Lewandowski and A. Magdziarz, "On mechanism of lignin decomposition – Investigation using microscale techniques: Py-GC-MS, Py-FT-IR and TGA," *Renewable Energy,* vol. 177, pp. 942-952, 2021.

[195] P. Malkavaara and R. Alen, "A spectroscopic method for determining lignin content of softwood and hardwood kraft pulps," *Chemometrics and Intelligent Laboratory Systems,* vol. 44, pp. 287-292, 1998.

[196] P. Malkavaara, P. Harjula, R. Alen and J. Knuutinen, "Chemometric investigation on structural changes in pine kraft lignin during pulping," *Chemometrics and Intelligent Laboratory Systems,* vol. 52, pp. 117-122, 2000.

[197] B. O. Myrvold and D. Pavlov, "Multivariate analysis for characterization of expanders," *Journal of Power Sources,* vol. 85, pp. 92-101, 2000.

[198] R. Huang, R. Su, W. Qi and Z. He, "Understanding the Key Factors for Enzymatic Conversion of Pretreated Lignocellulose by Partial Least Square Analysis," *Biotechnology progress,* vol. 26, no. 2, pp. 384-392, 2009.

[199] R. J. Gosselink, J. E. Dam, E. de Jong, E. L. Scott, J. P. Sanders, J. Li and G. Gellerstedt, "Fractionation, analysis, and PCA modeling of properties of four technical lignins for prediction of their application potential in binders," *Holzforschung,* vol. 64, pp. 193-200, 2010.

[200] B. Li, H. Liu, H. Xu, B. Pang, H. Mou, H. Wang and X. Mu, "Characterization of the Detailed Relationships of the Key Variables in the Process of the Alkaline Sulfite Pretreatment of Corn Stover by Multivariate Analysis," *BioResources,* vol. 9, no. 2, pp. 2757-2771, 2014.

[201] N. Đurišić-Mladenović, B. D. Škrbić and A. Zabaniotou, "Chemometric interpretation of different biomass gasification processes based on the syngas quality: Assessment of crude glycerol co-gasification with lignocellulosic biomass," *Renewable and Sustainable Energy Reviews,* vol. 59, pp. 649-661, 2016.

[202] G. E. Acquah, B. K. Via, O. O. Fasina, S. Adhikari, N. Billor and L. G. Eckhardt, "Chemometric modeling of thermogravimetric data for the compositional analysis of forest biomass," *PLoS ONE,* vol. 12, no. 3, p. e0172999, 2017.

[203] D. Vural, J. C. Smith and L. Petridis, "Dynamics of the lignin glass transition," *Physical Chemistry Chemical Physics,* vol. 20, pp. 20504-20512, 2018.

[204] T. R. Lima, E. C. Carvalho, F. R. Martins, R. S. Oliveira, R. S. Miranda, C. S. Müller, L. Pereira, P. R. Bittencourt, J. C. Sobczak, E. Gomes-Filho, R. C. Costa and F. S. Araújo, "Lignin composition is related to xylem embolism resistance and leaf life span in trees in a tropical semiarid climate," *New Phytologist,* vol. 219, pp. 1252-1262, 2018.

[205] R. Ma, X. Zhang, Y. Wang and X. Zhang, "New Insights Toward Quantitative Relationships between Lignin Reactivity to Monomers and Their Structural Characteristics," *ChemSusChem,* vol. 11, pp. 2146-2155, 2018.

[206] F. Akdeniz, M. Biçil, Y. Karadede, F. E. Ozbek and G. Ozdemir, "Application of real valued genetic algorithm on prediction of higher heating values of various lignocellulosic materials using lignin and extractive contents," *Energy,* vol. 160, pp. 1047-1054, 2018.

[207] J. D. S. Castro and C. F. d. Virgens, "Thermal decomposition of Nephelium lappaceum L. peel," *Journal of Thermal Analysis and Calorimetry,* vol. 138, pp. 3541-3549, 2019.

[208] H. Xu, X. Che, Y. Ding, Y. Kong, B. Li and W. Tian, "Effect of crystallinity on pretreatment and enzymatic hydrolysis of lignocellulosic biomass based on multivariate analysis," *Bioresource Technology,* vol. 279, pp. 271-280, 2019.

[209] H. Khaliliyan, C. Schuster, I. Sumerskii, M. Guggenberger, J. T. Oberlerchner, T. Rosenau, A. Potthast and B. Stefan, "Direct Quantification of Lignin in Liquors by High Performance Thin Layer Chromatography-Densitometry and Multivariate Calibration," *ACS Sustainable Chemistry & Engineering,* vol. 8, pp. 16766-16774, 2020.

[210] J. D. S. Castro, E. G. P. da Silva and C. F. Virgens, "Evaluation of models to predict the influence of chemical pretreatment on the peels of Nephelium lappaceum L. based on pyrolysis kinetic parameters obtained using a combined Fraser-Suzuki function and Friedman's isoconversional method," *Journal of Analytical and Applied Pyrolysis,* vol. 149, p. 104827, 2020.

[211] A. E. Ray, C. L. Williams, A. N. Hoover, C. Li, K. L. Sale, R. M. Emerson, J. Klinger, E. Oksen, A. Narani, J. Yan, C. M. Beavers, D. Tanjore, M. Yunes, E. Bose, J. H. Leal, J. L. Bowen, E. J. Wolfrum, M. G. Resch, T. A. Semelsberger and B. S. Donohoe, "Multiscale Characterization of Lignocellulosic Biomass Variability and Its Implications to Preprocessing and Conversion: a Case Study for Corn Stover," *ACS Sustainable Chemistry & Engineering,* vol. 8, pp. 3218-3230, 2020.

[212] H. Xu, Y. Kong, J. Peng, X. Song, X. Che, S. Liu and W. Tian, "Multivariate analysis of the process of deep eutectic solvent pretreatment of lignocellulosic biomass," *Industrial Crops & Products,* vol. 150, p. 112363, 2020.

[213] E. O. Ebikade, N. Samulewicz, S. Xuan, J. D. Sheehan, C. Wu and D. G. Vlachos, "Reductive catalytic fractionation of agricultural residue and energy crop lignin and application of lignin oil in antimicrobials," *Green Chemistry,* vol. 22, pp. 7435-7447, 2020.

[214] F. Sattari, D. Tefera, K. Sivaramakrishnan, S. H. Mushrif and V. Prasad, "Chemoinformatic Investigation of the Chemistry of Cellulose and Lignin Derivatives in Hydrous Pyrolysis," *Industrial & Engineering Chemistry Research,* vol. 59, pp. 11582-11595, 2020.

[215] C. F. das Virgens and J. D. Castro, "Screening of slow pyrolysis routes for maximum biofuel production from Syzygium malaccense biomass by TGA-FSD and chemometric tools," *Journal of Thermal Analysis and Calorimetry,* 2021.

[216] C. Guizani, M. Trogen, H. Zahra, L. Pitkanen, K. Moriam, M. Rissanen, M. Makela, H. Sixta and M. Hummel, "Fast and quantitative compositional analysis of hybrid cellulose-based regenerated fibers using thermogravimetric analysis and chemometrics," *Cellulose,* 2021.

[217] H. Xu, Y. Kong, J. Peng, X. Song, Y. Liu, Z. Su, B. Li, C. Gao and W. Tian, "Comprehensive analysis of important parameters of choline chloride-based deep eutectic

solvent pretreatment of lignocellulosic biomass," *Bioresource Technology,* vol. 319, p. 124209, 2021.

[218] F. Kartal and U. Ozveren, "An improved machine learning approach to estimate hemicellulose, cellulose, and lignin in biomass," *Carbohydrate Polymer Technologies and Applications,* vol. 2, p. 100148, 2021.

[219] Q. Li, G. Wittreich, Y. Wang, H. Bhattacharjee, U. Gupta and D. G. Vlachos, "Accurate Thermochemistry of Complex Lignin Structures via Density Functional Theory, Group Additivity, and Machine Learning," *ACS Sustainable Chemistry & Engineering,* vol. 9, pp. 3043-3049, 2021.

[220] S. Escamez, C. Terryn, M. L. Gandla, Z. Yassin, G. Scheepers, T. Nasholm, O. Sundman, L. J. Jonsson, J. Lundberg-Felten, H. Tuominen, T. Niittyla and G. Paes, "Fluorescence Lifetime Imaging as an In Situ and Label-Free Readout for the Chemical Composition of Lignin," *ACS Sustainable Chemistry & Engineering,* vol. 9, pp. 17381-17392, 2021.

[221] M. Karlsson, V. L. Vegunta, R. Deshpande and M. Lawoko, "Protected lignin biorefining through cyclic extraction: gaining fundamental insights into the tuneable properties of lignin by chemometrics," *Green Chemistry,* 2022.

[222] A. C. Garcia, C. Shuo and J. S. Cross, "Machine learning based analysis of reaction phenomena in catalytic lignin depolymerization," *Bioresource Technology,* vol. 345, p. 126503, 2022.

[223] K. Younes, A. Moghrabi, S. Moghnie, O. Mouhtady, N. Murshid and L. Grasset, "Assessment of the Efficiency of Chemical and Thermochemical Depolymerization Methods for Lignin Valorization: Principal Component Analysis (PCA) Approach," *Polymers,* vol. 14, p. 194, 2022.

[224] S. Jha, S. Bandyopadhyay, S. Mehta, M. Yen, T. Chagouri, E. Palmer and H. Liang, "Data-Driven Predictive Electrochemical Behavior of Lignin-Based Supercapacitors *via* Machine Learning," *Energy and Fuels,* vol. 36, no. 2, pp. 1052-1062, 2022.