UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

UNDERSTANDING THERMOCHEMICAL PROCESS AND FEEDSTOCK COMPOSITIONAL IMPACTS ON STRATEGIES TO CONTROL PYROLYSIS AND TORREFACTION PRODUCT DISTRIBUTIONS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

CHRISTOPHER LEE WATERS
Norman, Oklahoma
2016

UNDERSTANDING THERMOCHEMICAL PROCESS AND FEEDSTOCK COMPOSITIONAL IMPACTS ON STRATEGIES TO CONTROL PYROLYSIS AND TORREFACTION PRODUCT DISTRIBUTIONS

A DISSERTATION APPROVED FOR THE
SCHOOL OF CHEMICAL, BIOLOGICAL AND MATERIALS ENGINEERING

BY

_____
Dr. Lance Lobban, Chair

_____
Dr. Laura Bartley

_____
Dr. Steven Crossley

_____
Dr. Richard Mallinson

_____
Dr. Bin Wang

© Copyright by CHRISTOPHER LEE WATERS 2016
All Rights Reserved.

Dedication

This work is dedicated to:

My God, who is the Father, and the Son, and the Holy Spirit,

who "has withheld many things, thus exercising our mind towards in-depth research,

by providing few things as the cause for pondering";

My wife, Christina, whose love, support, and patience are a gift beyond measure;

and our daughter, Elizabeth, who brings much joy and meaning to my life.

Acknowledgements

I would like to acknowledge the following people, without whose help, support, patience, and inspiration none of this would have been possible:

Drs. Richard Mallinson, Lance Lobban, Daniel Resasco, Laura Bartley, Steven Crossley, and Rolf Jentoft, whose guidance, correction, and collaboration were invaluable, and to have been their student is a great honor and privilege;

Christina Waters, whose love and support are beyond description;

My parents, Charles & Dr. Diana Waters, whose sacrifice, teaching, and values were and are the reason I am the person I am today, and to whom I owe my life;

Dr. Shaolong Wan, whose work ethic is a continuing inspiration;

Adam Stevens, Fan Lin, Tyler Vann, Issac Schneberger, Eddie Rich, David Bjerkass, Kenneth Teel, ZhenXi Liang, and Daniel Dixon, my tireless colleagues, sounding-boards, and friends;

Caleb and Joshua Krell, my best friends, cheering section, and occasional tormentors;

Clint Shore, for his unique viewpoints, tenacious questioning, and inspirational commitment to realizing his ideals;

Fr. Jeremy Davis, David Swyden, and Jeannine Crouch, for their prayers;

Kyle Caldwell, for his unfailing capacity for intellectual stimulation and assistance in keeping things in perspective;

Sherry Marshall, for her inspiration, dedication to education, and infectious attitude; and Don Haas, for his motivation and inspiration to strive for more and never settle.

Table of Contents

List of Tables

List of Figures

Abstract

Growing concern about the effects of post-Industrial revolution fossil carbon dioxide release has led to large-scale research efforts to understand how to address and mitigate future fossil CO_2 emissions. Within the scope of products and services that have been enabled by the extraction of fossil hydrocarbon sources, those that rely on energy release by combustion – that is, electricity generation and transportation fuels – are considered the primary culprit in the observed increase of atmospheric CO_2 levels. While various mature technologies exist for the generation of electrical power without the need for combustion of coal, natural gas, or liquid fuels, there are significantly more difficult engineering challenges involved in supplying so-called "zero emission" or "carbon neutral" energy storage solutions for transportation which can displace petroleum-based products.

Numerous studies point to the possibility of displacing or replacing petroleum use for transportation fuels via the industrial production of liquid fuels from thermochemical conversion of biomass. As biomass can be a renewable and sustainable non-fossil source of organic molecules, it is hypothesized that it is possible to develop conversion strategies and technologies that allow society to retain the advantages and existing infrastructure associated with liquid hydrocarbon fuels without the associated fossil CO_2 release. However, barring extensive regulatory requirements, any biomass-based technologies that seek to displace petroleum on a wide scale must be economically viable. For liquid biofuels, lowering processing costs and improving total fuel yields per acre are some of the significant engineering challenges that hinder economic viability.

In this contribution, we consider the effects of changing various process conditions during biomass conversions, as well as develop an understanding of what compositional features of the biomass are most important to the resulting thermochemical product distribution. By understanding how and why the process conditions and the biomass feedstock compositions affect the resulting thermochemical product distribution, parallel development of feedstocks and processes tailored for those feedstocks can occur. This combined approach may help accelerate progress towards more viable biofuels production and therefore help accelerate petroleum displacement in transportation fuels.

Chapter 1: Pyrolysis & Catalysis Literature Review

Portions adapted from a paper published in Frontiers in Energy Research

Fan Lin, Christopher L. Waters, Richard G. Mallinson, Lance L. Lobban, Laura E. Bartley

1.1 Introduction

Biomass can be a renewable and sustainable source of transportation fuels not associated with fossil CO_2 release. Numerous studies highlight the advantages of displacing petroleum fuels with industrial production of liquid fuels from thermochemical conversion of biomass (*1*-*3*). Thermochemical conversion entails heating of biomass in an anoxic environment; condensation of organic liquid products, known as bio-oil; and subsequent treatment of the products with catalysts to create liquid fuels, i.e., refined bio-oil, similar to petroleum-derived gasoline or diesel. This is in contrast to biochemical conversion, which utilizes enzymes to release sugars followed by microbial production of ethanol or other fuel molecules (*4*, *5*). Relative to biochemical approaches, thermal conversion has the potential to make use of all carbon (C)-containing biomass components, would allow society to retain existing infrastructure associated with liquid hydrocarbon fuels, and, due to the rapidity of the process, may reduce production costs by permitting scalability and distribution of production (*6*, *7*). For both thermochemical and biochemical biofuels, lowering processing costs and improving fuel yields per hectare are major engineering challenges that hinder economic viability. Thermochemical fuel production also faces challenges related to maintaining a high C-yield while obtaining a fungible fuel.

1.2 Challenges for Thermochemical Conversion

Two types of pyrolysis have been developed: fast pyrolysis and slow pyrolysis. Slow pyrolysis is usually performed over several hours and has a high solid yield, and as such has little relevance for liquid fuels production. Fast pyrolysis, however, is typically performed quickly, in seconds, at temperatures between 400 and 600°C and decomposes most of the solid biomass into a volatile mixture of various organic molecules, water, and CO/CO_2. Pyrolysis oil or bio-oil constitutes the condensable portion of this vapor. Non-condensable components (primarily CO_2 and CO) and a mineral rich solid (char) are other product classes that will not be addressed here, except in that they detract from the overall C-yield of raw and refined bio-oil. Bio-oil comprises water (15-30%) plus compounds from several chemical families including the following (Table 1.1): organic acids, light (C_1-C_3) oxygenates, furan and furan derivatives, phenolic species with various degrees of methyl and methoxy substituents, pyrones, and sugar derivatives like levoglucosan (*8*, *9*). Bio-oil's chemically complex nature prohibits its direct use in combustion applications or petroleum refining. The reasons for this include low heating value; ignition difficulty; high chemical reactivity which results in oligomerization and polymerization over time and upon heating, prohibiting distillative separation (*10*-*12*); immiscibility with petroleum; and high corrosivity (*10*). Many of these features are associated with the high oxygen content of biomass and the resulting bio-oil, relative to fossil fuels.

Category	Major Components	Dry Wt (%)
Light Oxygenates	Glycolaldehyde, Acetol	3-26
Organic Acids	Acetic acid, Formic acid, Propanoic	2-27
Aldehydes	Acetaldehyde, Formaldehyde, Ethanedial	3-18
Sugars	1,6-Anhydroglucose (Levoglucosan)	5-14
Phenols	Phenol, Catechol (Di-OH Benzene), Methyl Phenol, Dimethyl Phenol	3-13
Guaiacols	Isoeugenol, Eugenol, 4-Methyl Guaiacol	3-15
Furans	Furfurol, Hydroxymethyl furfural, Furfural	2-11
Syringols	2,6-Dimethoxy Phenol, Syringaldehyde, Propyl Syringol	2-9
Ketones	Acetone	4-6
Alcohols	Methanol, Ethylene Glycol, Ethanol	2-6
Esters	Methyl Formate, Butyrolactone, Methylfuranone	<1-3

Table 1.1. The percentage ranges and categories of major bio-oil components. Source: (*6*)

The tendency for compounds in the condensed vapors from torrefaction and pyrolysis to self-polymerize is an often-cited issue that lowers overall yield, both carbon and liquid, and increases processing costs (*10*, *13*). The multi-oxygenated phenol derivatives, which are the major product of lignin pyrolysis, readily form dimers, trimers, and other oligomers when condensed. This is primarily due to the high amounts of organic acids present in condensed bio-oil which catalyze these reactions (*12*). These molecules are primarily guaiacol & syringol derivatives, often substituted at the 4 position with vinyl or propyl moieties. These reactions are a major contributor to the thermally unstable nature of biomass and its increasing viscosity over time.

1.3 Catalytic upgrading

In order to obtain desirable fuel properties and allow integration with the existing transportation fuels infrastructure (gasoline and diesel engines), the bio-oil must be chemically converted to reduce the undesirable characteristics mentioned above.

Catalytic upgrading is typically used to refine bio-oil, improving its stability and making it an acceptable liquid fuel.

The simplest method is hydrotreating or hydrodeoxygenation, which removes oxygen via catalytic hydrogenation (*14*), decreasing both the chemical reactivity and corrosivity. However, the low H:C ratio of bio-oil leads to very high hydrogen input costs (*15*, *16*). Additionally, this process converts any C_1-C_5 oxygenates, representing as much as half of the carbon in bio-oil, to C_1-C_5 hydrocarbons that are too volatile for liquid fuels rather than liquids in the target fuel range (C_6-C_{14}) (*16*, *17*), greatly decreasing the process carbon efficiency.

Another straightforward approach is to 'crack' the pyrolysis vapors using acidic zeolite catalysts into light olefins and aromatic hydrocarbons (primarily benzene, toluene, and o/m/p-xylene) (*18*-*20*). This approach is appealing because of the lack of an external H_2 requirement and the simplicity of the product streams. Furthermore, since zeolite cracking is widely used in traditional petroleum refining/valorization (*21*), other advantages are the product compatibility with existing refinery infrastructure and the maturity of the process (*21*). However, zeolite cracking is crippled by poor usable carbon yield due to the high amounts of coke, CO, and CO_2 formed during the catalytic process (*19*) and the concomitant rapid catalyst deactivation. Regeneration of zeolites is typically done by combusting away the coke, and the carbon is lost to CO_2.

Additionally, further catalytic oligomerization and reforming for olefins and aromatics, respectively, are needed to make these products suitable for addition to refinery fuel product streams, increasing the process costs and further reducing overall carbon yield.

Recently, a variety of novel catalytic strategies have been proposed (*22*). Briefly, these include: selective hydrogenation of acetic acid to ethanol over oxide-supported metals; ketonization of acetic acid to acetone on both oxide & zeolite catalysts; condensation of light oxygenates and furanic compounds over a variety of catalysts; hydrolysis of levoglucosan followed by partial oxidation to gluconic acid; aldol condensation in biphasic water/organic emulsions using bifunctional nanohybrid metal/basic catalysts; a variety of options to reduce furanic compound reactivities; disproportionation and transalkylation of methoxy- substituents on phenolic species via metal-functionalized zeolites and oxide-supported metals; selective hydrogenation of the C-O phenolic bond using bimetallic alloy catalysts; and alkylation of phenolic species with ketones using functionalized zeolites.

While the studies cited above all show promising results, the underlying problem with any catalytic strategy to upgrade fast pyrolysis bio-oil to a suitable liquid fuels product is that any one catalyst cannot be optimized to convert the complex mixture that is bio-oil. A catalyst used for upgrading of one family of compounds (e.g. light oxygenates) may not be suitable for other families (e.g. substituted phenolics or furanics) which results in undesirable reactions (breaking C-C bonds unnecessarily, or increasing H:C ratios above the 2:1 optimum) or undergoing rapid deactivation due to reactions with other, non-targeted bio-oil oxygenates. For example, acidic catalysts which have been shown to be very efficient at producing alkyl aromatics from polysaccharide-derived molecules (*23*), are severely inhibited by the presence of lignin-derived methoxyphenols. These high molecular weight compounds have been shown to compete strongly for adsorption sites on these catalysts.

These upgrading challenges suggest the desirability of thermal conversion producing more selective product streams, i.e., each stream comprising fewer families of chemical compounds. As a result, some separation of bio-oil into catalytically favorable product streams must be achieved in order to improve the viability of this technology.

1.4 Staged Thermal Fractionation

One strategy which has been proposed to produce simpler intermediate streams is staged thermal fractionation (also referred to as staged degasification) (*24-26*). Staged fractionation leverages the inherent differences in the thermal stability and decomposition products of the biopolymers which constitute biomass. TGA studies have shown that hemicelluloses, cellulose, and lignin decompose over different temperature ranges; generally, hemicellulose decomposes at a lower temperature range (220-315°C) than cellulose (300-400°C), while lignin decomposes over a broad range of temperatures (150-900°C) (*27-29*). Additionally, In a series of publications, Brown and Patwardhan report the product distributions from the pyrolysis of the three biomass fractions (*12*, *30*, *31*). A summary of their results is reported in table 1.2.

Hemicellulose		**Cellulose**		**Lignin**	
Product	wt %	Product	wt %	Product	wt %
CO	2.8	Formic Acid	6.6	CO	1.8
CO_2	18.8	Furan/Acetone	0.7	CO_2	15.2
Acetaldehyde	0.7	Glycolaldehyde	6.7	Acetaldehyde	0.9
Formic Acid	11	Acetic acid	0	Formic acid/Acetone	0.7
2-methyl furan	1.5	2-Methyl furan	0.4	2-methyl furan	0.1
Acetic acid	1.1	Acetol	0.3	Acetic acid	11.5
Acetol	3	Furfural	1.3	2-furaldehyde	0.2
Furfural	2.2	2-Furanmethanol	0.5	Phenol	1.9
DAXP 1	1.6	3-Furanmethanol	0.3	2-methoxy phenol	0.9
DAXP 2	7	5-Methyl furfural	0.2	2-methyl phenol	0.1
Other DAXP	0.6	2-Hydroxy-3-methyl	0.2	4-methyl phenol	0.6
AXP	2	cyclopenten-1-one		2-methoxy-4-methyl phenol	0.7
Other AXP	1.4	Levoglucosenone	0.4	3,5-dimethyl phenol	0.1
Char	10.7	5-Hydroxymethyl furfural	2.8	3-ethyl phenol	0.6
Xylose	4.9	Anhydro xylopyranose	3	4-ethyl-2-methoxy phenol	0.4
Water	15.1	Levoglucosan - pyranose	58.8	4-vinyl phenol	3.5
		Levoglucosan - furanose	4.1	2-methoxy-4-vinyl phenol	1.8
		Other anhydrosugars	1.4	Unidentified (MW 114)	1.4
		Char	5.4	2,6-dimethoxyphenol	1.0
				Euginol	0.2
				4-methyl-2,6-dimethoxyphenol	0.8
				3,5-dimethoxy-4-hydroxy benzaldehyde	0.4
				3',4'-dimethoxy acetophenone	0.8
				4-allyl-2,6-dimethoxyphenol	0.2
				4-allyl-2,5-dimethoxyphenol	0.3
				3',5'-dimethoxy-4'-hydroxy acetophenone	0.3
				Sinapyl alcohol	0.7
				Unidentified (MW 280)	0.4
				Char	37

Table 1.2. Product distribution from fast pyrolysis of three biopolymer fractions. From (*12*, *30*, *31*)

As all three biopolymers yield distinct thermal decomposition products, it should be possible to develop a staged thermal fractionation strategy that is capable of producing several product streams of enhanced compositional purity (Figure 1.1). A low temperature step (at torrefaction conditions) targeting hemicellulose decomposition is followed by an intermediate step targeting cellulose decomposition, and finally a high

temperature step (similar to fast pyrolysis conditions) to decompose the remaining lignin. These purified streams could then be catalytically upgraded using the best catalyst for the job.

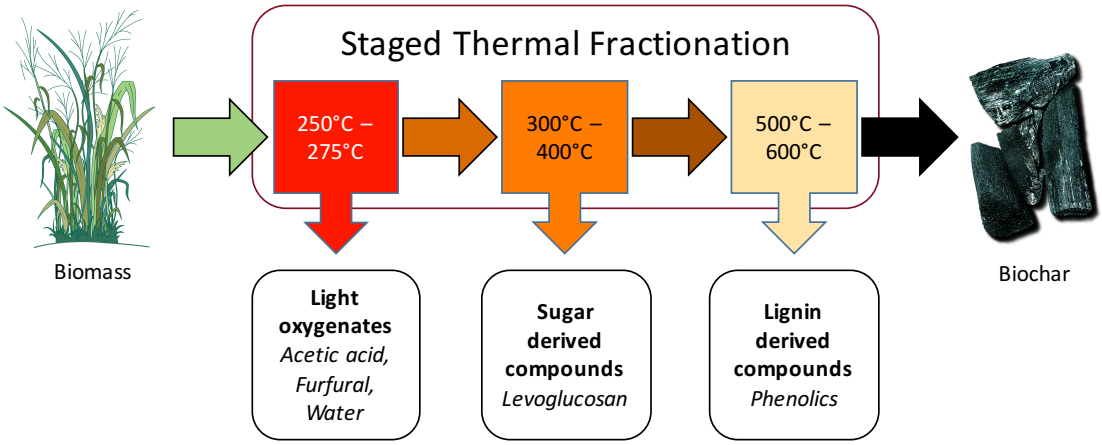


Figure 1.1. Hypothetical configuration for thermal fractionation of lignocellulosic biomass with resultant purified product streams

Achieving a purely thermal separation of the cellulose- and lignin- derived components from biomass (e.g., levoglucosan and phenolic species) is of particular interest due to the inherent overlapping thermal stability regimes of these components and challenges of downstream separation. As hemicellulose is the more unstable polysaccharide, lower temperature treatments are expected to decompose this component with minimal cellulose decomposition leaving a remaining solid enriched in cellulose and lignin. The major products of these two biopolymers – levoglucosan and variously substituted phenolics – are difficult to separate using thermal methods and typically involve solvent extraction of one form or another (*32*). Achieving a thermal segregation of these major pyrolysis products could enable implementation of the advanced catalytic strategies outlined above by eliminating the need for a costly downstream separation step. This latter challenge might be addressed by understanding the relationships between biomass

composition and bio-oil components, and using this information to alter biomass, either through genetic, chemical, or thermal means.

Chapter 2: Relationships between biomass composition and liquid products formed via pyrolysis

Adapted from a paper published in Frontiers in Energy Research

Fan Lin, Christopher L. Waters, Richard G. Mallinson, Lance L. Lobban, Laura E. Bartley

2.1 Biomass composition and chemical structures

Recent reviews have addressed the general relationships between biomass composition and thermal products, such as increasing the content of phenolics relative to carbohydrates to reduce the oxygen content of bio-oil (*33*). Here, we provide a more detailed description of the chemical structure and interactions among major cell wall components to aid in understanding more subtle relationships between biomass and bio-oil content. Biomass consists of cell walls that establish the structure of the plant and, to a lesser extent, nonstructural components (Table 2.1). Cell walls determine the shape of leaves and stems and the cells that compose them and consist of cellulose, hemicellulose, lignin, as well as structural proteins and wall-associated mineral components. Nonstructural components include sugars, proteins, and additional minerals (*33-35*). For example, in switchgrass, an important potential bioenergy crop, dry biomass consists of ~70% cell walls, 9% intrinsic water, 8% minerals, 6% proteins, and 5% nonstructural sugars (*35*). The relative fractions of different components, chemical linkages within and between polymers, and cellular patterning varies among plant species, organs, developmental stages, and growth conditions (*36-39*). Here, we review the components of secondary cell walls, which are formed as plant growth ceases, as they constitute the majority of plant biomass (*40*), and then discuss evidence for interactions among components. Table 2.1 lists the different major and minor

components of biomass and the broad ranges of their representation within biomass for biofuel conversion. Figure 2.1 shows the chemical structures and atom numbering of the most abundant cell wall monomeric species.

Biomass Component	**Dry Wt (%)[a]**
Cellulose	15-49 [b]
Hemicellulose:	12-50 [b, c, d]
Xylan	5-50 [c]
Mixed Linkage Glucan	0-5 [c, e]
Xyloglucan	Minor [c]
Mannan (and galactoglucomannan)	0-30 [c, f]
Soluble (Mainly sucrose)	9-67 [b, g]
Pectin	< 0.1 [h]
Lignin	6-28 [b]
Ferulic acid and *p*-coumaric acid	< 1.5 [h]
Protein	4-5 [b]
Ash (Mainly silicate)	0.4-14.4 [b]
Intrinsic moisture	11-34 [i]

[a]Percent mass composition of secondary cell walls. [b](*40*) [c](*41*)[d]As the highest percentage of xylan in (*41*) is higher than the highest percentage of hemicellulose in (*40*), the highest percentage of hemicellulose is set to the highest percentage of xylan. [e]MLG is only abundant in grasses, The maximum percentage of MLG we are aware of is that of the mature rice stem after flowering. (*42*) [f]Galactoglucomannan is only abundant in gymnosperm woods. Dicots and grasses possess <8% of mannan and galactoglucomannan. (*41*) [g]The high solubles abundance is only for sorghum biomass. Other plants usually have less than 15% soluble content. (*40*) [h](*43*) i (*44*)

Table 2.1. The variation of biomass components among vascular plants including grasses, softwoods, and hardwoods.

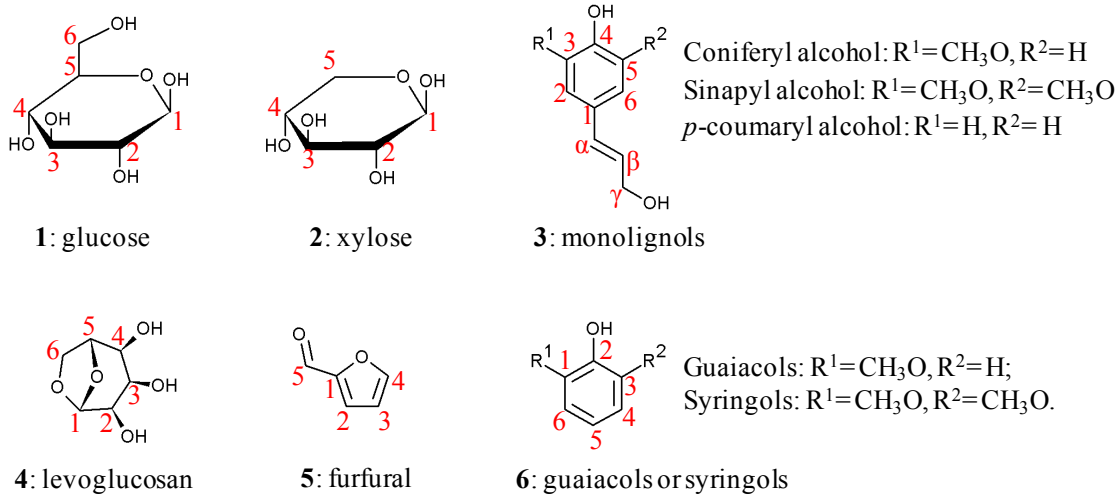


Figure 2.1. Structure of the major basic units of biomass polymers and related products

Cellulose and hemicellulose represent 15-49% and 12-50% of biomass by dry weight, respectively (*39*, *40*, *43*). Cellulose is an unbranched homopolymer of >500 β-(1,4)-linked glucose units. In plant cell walls, cellulose is primarily in the form of crystalline microfibrils consisting of approximately 36 hydrogen-bonded cellulose chains, but also has amorphous regions (*45*, *46*).

Hemicelluloses are typically branched polysaccharides substituted with various sugars and acyl groups. As discussed further in section 2.3, the different sugar composition and linkages of hemicelluloses influence thermal products (*47*, *48*). The structure and composition of hemicellulosic polysaccharides differ depending on plant species classification, i.e., taxonomy. Major taxonomic divisions with the relevance to bioenergy production are grasses, such as switchgrass and wheat; woody dicots, i.e., hardwoods, such as poplar; and woody gymnosperms, i.e., softwoods, such as pine. The most abundant grass hemicelluloses are mixed-linkage glucan (MLG) and glucuronoarabinoxylan (GAX) (*41*, *49*); the hemicelluloses of hardwood are primarily composed of glucuronoxylans (GX) but also contain a small amount of galactomannans (GM) (*40*); and softwood hemicelluloses are largely galactoglucomannan (GGM) and GAXs (*41*). MLG is an unbranched glucose polymer similar to cellulose but containing both β-(1-3)- and β-(1-4)-linkages (*42*). MLG is nearly unique to the order Poales, which includes the grasses, but has also been found in horsetail (*Equisetum*). Its abundance in mature tissues and secondary cell walls has recently been recognized(*49*). Xylans consist of a β-(1-4)-linked xylose backbone with various substitutions. GXs are xylans substituted mostly by glucuronic acid and 4-O-methyl glucuronic acid through α-(1-2)-linkages. GAXs are not only substituted by glucuronic acid, but also substituted

by arabinofuranoses at the O-3, which can be further substituted by the phenylpropanoid acids, to form feruloyl- and p-coumaryl esters linked at the O-5 (*41*). Acetyl groups are often attached to the O-3 of backbone xyloses but also attach to the O-2. Unlike xylans, which mainly consist of pentoses, mannans consist of hexoses like mannose, glucose and galactose. GM and GGM have a β-(1-4)-linked backbone with mannose or a combination of glucose and mannose, respectively. Both GM and GGM can be acetylated and substituted by α-(1-6)-linked galactoses (*41*, *50*, *51*). Relatively depleted in secondary walls, but rich in growing primary walls of dicot species, xyloglucan and pectins are two other polysaccharides in cell walls. Xyloglucan consists of β-(1-4)-linked glucose residues, modified by xylose and other sugar residues; and pectin is another branched or unbranched polymer that is rich in galacturonic acid, rhamnose, galactose, and several other monosaccharide residues (*5*, *41*).

Lignin is a crosslinked, heteropolyphenol mainly assembled from three monolignols—sinapyl (***S***), coniferyl (***G***), and p-coumaryl (***H***) alcohols. As waste products are often selected as biofuel feedstocks, it is also relevant to note that lignin derived from other monolignols such as caffeyl alcohol and 5-hydroxyconiferyl have been found in the seedcoat of both monocots and dicots (*52*, *53*). Lignin structural heterogeneity and various types of incorporated groups can lead to a variety of different depolymerization reactions during pyrolysis (*54*). Often traceable to the corresponding bio-oil components, the three major lignin units differ in the degree of methoxylation of their carbon ring. ***S***-units are methoxylated at both O-3 and O-5 ring positions; ***G***-units have one methoxy group at the O-3 position; and ***H***-units lack ring methoxy groups (3, Figure 2.1) (*55*). Lignin units undergo oxidative coupling in the cell wall to form many

types of dimers, including β–O–4, β–5, β–β, 5–5, 5–O–4, and β–1, leaving other atoms free to further polymerize, which significantly increases the structural heterogeneity of lignin. Lignin units can also be esterified with p-coumaryl, p-hydroxybenzoyl, and acetyl groups, primarily at the γ position of terminal units (*56*, *57*). Lignin compositions and the acylation groups vary among plant clades (*55*). Woody dicot lignins have ***G***- and ***S***-units, and trace amount of ***H***-units. Poplar wood, for example has a ***G:S:H*** ratio of 55:45:1 (*58*). The lignin of many hardwoods is acylated by p-hydroxybenzoates (*57*) and acetyl groups in low amounts (*59*). Biomass from other species, like palms and kenaf, possess a high degree of lignin acetylation (*60*). Grass lignins also contain ***G***- and ***S***-units with slightly higher amount of ***H***-units than woody dicots. Wheat straw, for example, has a ***G:S:H*** ratio of 64:30:6 (*61*). Grass lignin possesses high levels of p-coumarate esters (*62*) and can also be etherified by tricin and ferulic acid (*63*, *64*), as discussed further below. Woody gymnosperm lignins are different from angiosperm lignins, being primarily composed of ***G***-units and a lower amounts of ***H***-units (*55*). Biomass also contains inorganic elements including Ca, K, Si, Mg, Al, S, Fe, P, Cl, Na and some trace elements (<0.1%) such as Mn and Ti, according to ash analysis, formed by oxidation of biomass at 575°C (*65*, *66*). As with other biomass components, the abundance of mineral elements varies among species. In general, compared with grass biomass, woody biomasses contain less ash, Cl, K, N, S, and Si, but more Ca (*66*). Plant biomass components do not accumulate independently of each other, though their relationships are still an active area of research (*67-69*). Biomass components can correlate with each other because they are physically bound to each other through covalent and non-covalent interactions or because they accumulate in the same plant

organ or stage of plant development, though a physical interaction may not exist. Because the abundance of some biomass components is correlated, the thermal products from one biomass component may also correlate with other components. For example, the abundance of cellulose correlates with the abundance of lignin in five different biomass sources (Pearson's Correlation Coefficient = 0.83) and lignin-derived thermal products correlate with cellulosic glucose (*48*). Many mineral elements are also correlated with each other, for example, N, S, and Cl; Si, Al, Fe, Na, and Ti; Ca, Mg, and Mn; K, P, S, and Cl (*66*). Numerous interactions between lignins and hemicelluloses and among hemicelluloses have been observed. Among the best-studied examples, GAXs of grasses and other recently evolved monocot species covalently link to lignin through ether bonds with ferulate esters on arabinose moieties of arabinoxylan (*70*). In poplar and spruce wood, NMR results indicate that lignin and carbohydrates are directly bonded through several types of ether linkages (*71*, *72*). The data provide evidence for ether bonds between lignin and C_1, C_5, and C_6 atoms of pentoses and hexoses. Generally, xylan is the most closely associated polysaccharide to lignin, and NMR studies have also clearly identified lignin-glucuronic acid ester bonds (*71*). Also, MLGs closely coat low-substituted xylan regions, likely via non-covalent interactions (*73*, *74*). Furthermore, some components can also affect the distribution of other components. For example, rice plants that overexpress an enzyme that cleaves MLG exhibit reduced MLG and have an altered distribution profile of Si though maintain the same total amount of Si (*75*). In sum, mounting evidence supports covalent and noncovalent interactions among cell wall polymers and components, however these

connections have been difficult to study with questions persisting related to how different cell wall preparations and manipulation may alter observations.

2.2 Models for relationships between biomass components and bio-oil product composition

Reaction pathways of individual biomass components to formation of thermal products have been described (*76*). However, the pyrolysis literature suggests that biomass components tend to have more complex effects on bio-oil yield and product composition than simply their quantity. Here we introduce three possible "models" of how biomass components may influence the yield or composition of thermal products, and in section 2.3 we discuss evidence supporting each of them. Figure 2.2 provides schematic representations of the following models:

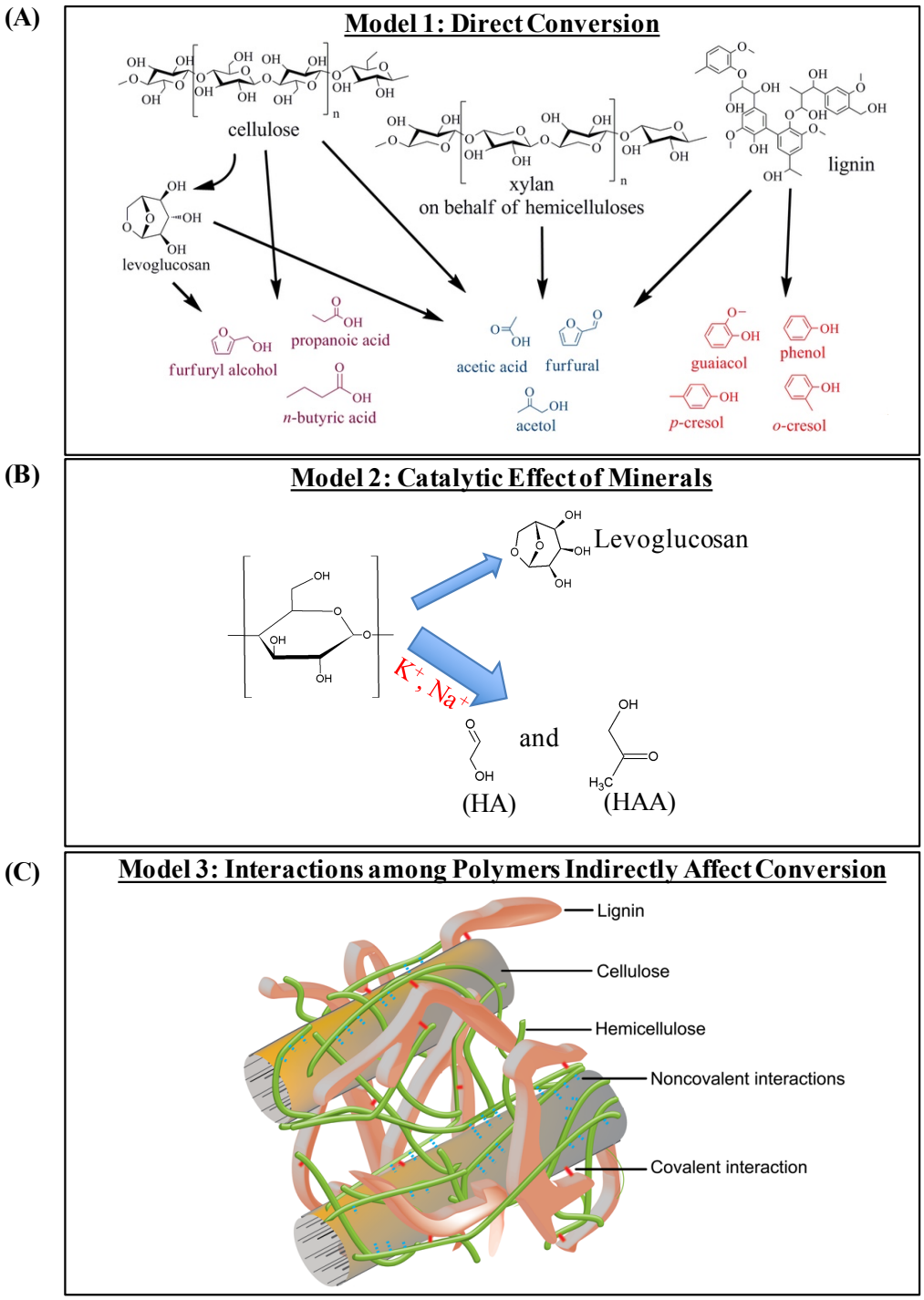


Figure 2.2. Three models of how biomass components and their interactions affect the formation of thermal products.
HA: Hydroxyacetone, HAA: Hydroxyacetaldehyde. (A) Direct conversion (B) Catalytic effect of minerals (C) Interactions among Polymers Indirectly Affect Conversion. (A) and (C) are adapted and modified from (*77*) and (*78*) with permission.

Model 1: Biomass components are the direct sources of thermal products. Components are converted to products through de-polymerization and secondary reactions such as cracking, i.e., splitting, and recombination (Figure 2.2A).

Model 2: Components or their derived products act as catalysts that accelerate thermal reactions of other components, altering product yields and ratios (Figure 2.2B).

Model 3: Chemical interactions or structural relationships among cell wall components alter bio-oil composition and/or yield (Figure 2.2C). This "indirect" model applies when variation in a biomass component alters the yield of a chemically unrelated product in a manner not easily explained by a catalytic effect. Chemical interactions that alter products may either be covalent or non-covalent chemical bonds between cell wall components. Structural relationships refer to correlations between components, often minor ones, and physical features of the biomass. For example, the abundance of a cell wall component may be indicative of the structure of the plant material, such as biomass bulk density differences caused by different leaf to stem ratios, but do not reflect chemical bonding between components. As of the preparation of this review, very little evidence addresses how biological correlations effect bio-oil products, so the discussion focuses on potential chemical interactions.

2.3 Evidence relating biomass content and bio-oil composition

Evidence in the literature for the three models described above is presented in table 2.2 and discussed below. In the reviewed experiments, relationships between biomass components and pyrolysis products have been identified by varying the starting biomass, either through experimentation on purified components, via naturally occurring variation among different biomass sources, or via pretreatment of the

biomass. Most studies included in this discussion report the chemical products derived from pyrolysis of biomass or biomass components. Studies that only reported weight losses or elemental balances were not considered. The two dominant techniques present in this corpus of literature are either pyrolysis-gas chromatography/mass spectroscopy where pyrolysis vapors from microgram- to milligram-scale samples are directly transported to a GC for analysis, or pyrolysis in a gram- to kilogram- scale reactor system followed by condensation of the vapors and subsequent chromatographic analysis of the liquid.

Possible model	Biomass component	Independent variable	Effect on products (Dependent variable)	Sample type	Thermal conversion condition	Reference
1	Hemicellulose	Arabinose, xylose, mannose, arabinitol	Arabinose and xylose produce similar products but slightly different yields and product ratios. Mannose gives more decomposition products than arabinose and xylose and a unique product, 5-hydroxymethyl-furfural.	Powdered high purity mono-saccharides (>99%)	500°C and 550°C, 10 s	(*79*)
1	Hemicellulose or nonstructural polysaccharides	Variation in sugar composition	Fermentated grain samples produce less acetic acid, furfural, and acetone than hull and straw.	Barley Straw, hull and yeast-fermentated grain	Fluidized bed , 500°C	(*80*)
1	Lignin	Lignin variation among hardwoods, softwoods, and grasses	Hardwood lignin produces guaiacyl and syringyl derivatives. Softwood lignin produces mostly guiaicyl derivatives but no hydroxyphenyl or syringyl compounds. Grass lignin uniquely produces vinylphenol, propenyl-phenols, and p-hydroxybenzaldehyde.	Solvent extracted lignin from milled spruce, beech, aspen, and bamboo	510°C, 10 s	(*81*)
1	Lignin	Natural variation and variation by hydrogen peroxide treatment in Klason lignin	Samples with high Klason lignin produce more 4-vinylguaiacol. Klason lignin positively correlate with 4-vinylguaiacol.	Hybrid maize, bm1 mutant, bm3 mutant and switchgrass biomass treated by different [H2O2] to remove lignin	650°C , 20 s	(*82*)

Table 2.2. Evidence in the literature for the three models described in section 2.2
CONTINUED ON FOLLOWING PAGES

Possible model	Biomass component	Independent variable	Effect on products (Dependent variable)	Sample type	Thermal conversion condition	Reference
1	Lignin	Natural variation in lignin abundance	High-lignin endoscarp biomasses produces more phenolic compounds, like phenol, 2-methoxyphenol, 2-methylphenol, 2-methoxy-4-methylphenol and 4-ethyl-2-methoxyphenol, compared to switchgrass biomass.	Walnut, olive, coconut husks, and peach drupe endocarp biomass and switchgrass biomass	650°C , 20 s	(*83*)
1	Lignin	Natural variation in lignin composition or structure	Compared with endocarp lignin, switchgrass lignin produces more acetic acid, toluene, furfural and 4-methylphenol, and lower or undetectable amounts of 4-ethyl-2-methoxyphenol, 2-methoxy-4-vinylphenol and 2-methoxy-4-(2-propenyl)-phenol. Compared with walnut and olive endocarp lignin, coconut endocarp lignin produces strikingly more phenol and less 2-methoxy-4-methylphenol. Coconut shell lignin produces unique compounds among the analyzed feedstocks, such as 2,6-dimethoxyphenol, 2-methoxy-4-(2-propenyl)-phenol and vanillin, but less 2-methoxy-4-(1-propenyl)-phenol compared to walnut shell and olive lignin.	Lignin extracted from various endocarps and switchgrass	650°C , 20 s	(*83*)
2	Mineral content	Variation of K+, Mg2+, and Ca2+ by demineral-ization and impregnation	Samples with more K+ produce more glycolaldehyde, acetic acid, acetol, butanedial, guaiacol, syringol, and 4-vinylsyringol, but less levoglucosan, furans, and pyrans. Samples with more Mg2+ produce less glycolaldehyde, levoglucosan and 3-methoxycatechol. Samples with more Ca2+ produce more pyrans and cyclopentenes but less levoglucosan and 3-methoxycatechol.	Poplar wood powder demineralized with hydrofluoric acid and impregnated with KCl, MgCl2 and CaCl2	550°C, 10 s	(*84*)
2	Mineral content	Increase of minerals by adding 1% and 5% NaCl	Samples with more NaCl added produce more furans, acids, ketones, and phenols than the samples with no NaCl added.	Rice straw and bamboo with NaCl	400°C-900°C, 4 min	(*85*)

Possible model	Biomass component	Independent variable	Effect on products (Dependent variable)	Sample type	Thermal conversion condition	Reference
2	Mineral content	Decrease minerals by acid wash	Cellulose with the acid washed char produces more levoglucosan but less formic acid, furane-type derivatives, and light oxygenates, than cellulose with unwashed char.	Cellulose with acid-washed and unwashed red oak char	400°C, 15 s	(*86*)
2	Mineral content	Decrease of K and Na by washing	Washed samples yield more levoglucosan, hydroxyl-acetaldehyde, and char than unwashed samples.	Washed and unwashed lolium, festuca, willow, and switchgrass biomass	500˚C, 10 s	(*87*)
2	Mineral content	Impregnation with NaCl, KCl, MgCl2, CaCl2, Ca(OH)2, Ca(NO3)2, CaCO3 and CaHPO4	Impregnated samples produce less levoglucosan, but more glycol-aldehyde and and acetol, compared to untreated cellulose.	Microcrystalline powdered cellulose impregnated with salt or switchgrass ash solution.	500 °C, 30 s and a range from 350°C to 600°C with 50°C increments	(*88*)
2	Mineral content	Natural variation in ash	Ash negatively correlates with hydroxyacetaldehyde and phenolic compounds, such as, trimethoxy-benzene, syringol, 4-allyl-2,6-di-methoxyphenol, syringaldehyde, 3,5-dimethoxy-4-hydroxy-cinnamaldehyde.	Poplar, willow, switchgrass, hot-water extracted sugar maple, and debarked sugar maple.	550˚C , 20 s	(*48*)
2	Mineral content	Increase mineral content by adding CaCl2,NaCl, KCl, K2SO4,KHCO3 individually to washed wood	Washed wood wtih calcium salt added produces more C6H8O2 compound compared to washed wood with no additive. Samples wtih sodium salt and potassium salt added produce less 2-Hydroxy-butanedial, 4-Hydroxy-5,6-dihydro-(2H)-pyran-2-one, and levoglucosan, but more tetrahydro-4-hydroxy-pyran-2-one, compared to washed samples.	Hornbeam wood washed and mixed with salts	280°C, 20 min , 100~500°C	(*89*)

Possible model	Biomass component	Independent variable	Effect on products (Dependent variable)	Sample type	Thermal conversion condition	Reference
2	Mineral content	Increase mineral contents by adding 1% NaCl	Xylans with NaCl added yield more glycolaldehyde and 3-deoxypentono-1,4-lactones, but less 1,4-anhydro-α-D-xylopyranose, 1,5-anhydro-4-deoxy-pent-I-en-3-ulose 2.4 and char	Synthesized xylan.	280°C, 30 min	(*90*)
2	Mineral content	Replace the cations in biomass with Na+ by ion-exchange or increase mineral content by adding Na2SO4 or NaHCO3	Sodium-containing pulps produce less polysaccharide-derived products and similar amounts of lignin-derived products compared to untreated pulps.	Spruce pulp, treated with ion exchange and salt impregnatation	620°C, 2s	(*91*)
3	Inter-component linkages	Biomass with hemicellulose selectively removed, compared with a physical mixtures of lignin and cellulose, and separate pyrolysis of lignin and cellulose.	The native grass cellulose-lignin samples produce more C1 to C3 products and furan, but less pyrans and levoglucosan.	cornstover, red oak, pine and switchgrass, depleted for hemicellulose and compared with different extracted component mixtures.	500˚C	(*92*)
1 & 3	Acetyl content	Natural variation in acetyl content among five species	Acetyl content positively correlates with acetic acid, but also methyl pyruvate and 2-furanone.	Poplar, willow, switchgrass, hot-water extracted sugar maple, and debarked sugar maple.	550˚C , 20s	(*48*)
1 & 3	Acetyl content	Variation of acetylation between acetylxylan and xylan	The acetylxylan produces more acetic acid, furan, and acetone, but less 2-furfural and acetaldehyde, than xylan.	Xylans extracted from cotton wood	500˚C	(*47*)

Possible model	Biomass component	Independent Variable	Effect on products (Dependent variable)	Sample type	Thermal conversion condition	Reference
1 & 3	Cellulose	Natural variation in glucans, mostly cellulose, among five species	Glucans positively correlate with levoglucosan and hydroxymethyl furfural, but also phenolic compounds like syringaldehyde, vanillin, and 3,5-dimethoxy-4-hydroxycinnamaldehyde.	Poplar, willow, switchgrass, hot-water extracted sugar maple, and debarked sugar maple.	550˚C , 20 s	(*48*)
1 & 3	Lignin	Natural variation in lignin among five species	Lignin positively correlates with hydroxyacetaldehyde and phenolic compounds, trimethoxybenzene; syringol; 4-allyl-2,6-dimethoxyphenol; syringaldehyde; and 3,5-dimethoxy-4-hydroxycinnamaldehyde.	Poplar, willow, switchgrass, hot-water extracted sugar maple, and debarked sugar maple.	550˚C , 20 s	(*48*)
1 & 3	Xylan	Natural variation in xylan among five species	Xylan positively correlates with hydroxyacetone, but negatively correlates with hydroxymethyl furfural and syringaldehyde.	Poplar, willow, switchgrass, hot-water extracted sugar maple, and debarked sugar maple.	550˚C , 20 s	(*48*)
2 & 3	Mineral content or other structure	Variation in weakly associated and strongly associated K, Na, Mg and Ca by washing with deionized water or nitric acid	Acid-washed samples produce bio-oil with a greater ration of organic products to water compared to unwashed samples. Washed samples with less mineral conent produce more levoglucosan and sugar compounds but less monophenols like phenol, 2-methyl-phenol, syrigol.	Eucalyptus loxophleba wood	Fluidized bed , 500°C	(*93*)
3	Water	Natural variation in moisture content (0%-20%)	Samples with greater moisture content produce more char and gas but less water comapred to lower moisture samples.	Pine wood	Fluidized bed , 480°C, 2 s	(*94*)

Possible model	Biomass component	Independent Variable	Effect on products (Dependent variable)	Sample type	Thermal conversion condition	Reference
3	Water	Natural variation in moisture content (2%-55%)	Samples with greater moisture content produce more condensible products, but less char than dry samples	Norway spruce	500°C, 30 min	(*95*)
3	Water	Natural variation in moisture content (5%-15%)	Samples with greater moistrue conent produce less levoglucosan at 450°C and 500°C than samples with lower moisture conent. Other products are also significantly affect by moisture conent, but trend sdepend on pyrolysis temperauture.	Switchgrass	Fluidized bed, 450°C, 500°C, 550°C	(*96*)
3	Inter-component linkages	Variation in polysaccharide amounts in lignin-carbohydrate complexes by enzyme treatment	Samples with lignin-associated polysaccharides removed produce more coniferyl alcohol than untreated samples.	Lignin-carbohydrate complexes extracted from spruce wood	500˚C, 1 min	(*72*)
1	Lignin interunit linkages (lignin dimer)	Variation of lignin interunit linkages, including β-O-4′ ethers, pinoresinols, phenylcoumaran, and dibenzodioxocins, among lignin-carbohydrate complexes with different compositions.	Glucomannan-associated lignin produces more guaiacol than xylan-associated lignin and glucan-associated lignin.	Lignin-carbohydrate complexes extracted from spruce wood	500˚C, 1 min	(*72*)

2.3.1 Model 1: Direct products of cellulose, hemicellulose, and lignin

Thermal breakdown of purified cellulose, hemicellulose, and lignin have been relatively well studied. Levoglucosan, a six carbon 1,6-anhyrosugar (see figure 2.1), was identified as the main product of cellulose pyrolysis nearly a century ago (*97*). Levoglucosan is formed alongside other smaller decomposition products, with maximum levoglucosan production occurring at 500°C (*98*). Minor products of cellulose pyrolysis are dominated by other anhydrosugars that retain all 6 carbons of glucose, such as 1,6-anhydroglucofuranose and 5-hydroxymethyl furfural, but also smaller molecules, like furfural (5, Figure 2.1), formic acid, and glycolaldehyde, among others (*31*).

As with cellulose, hemicellulose pyrolysis products depend mostly on the number of carbons in the monosaccharide residues of the starting polymer (*47*). Pentoses and hexoses produce similar light, C_1-C_3 oxygenates but differ in the types and selectivities (i.e., relative ratios) of heavier, C_4-C_6 products. Consistent with expectations, pyrolysis of monosaccharides reveals that hexoses can form more unique compounds than pentoses, including pyranic species; additionally, pentoses yield more lighter fragmentation products than hexoses and only trace amounts of C_6 and higher products (*79*).

Lignin thermal degradation products generally retain the characteristic ring decoration of the monolignols from which they originate (3, Figure 2.1). For example, syringol derivatives are bio-oil products derived from ***S***-lignin units and guaiacols are products derived from ***G***-lignin units (6, Figure 2.1). The derivative groups possess 1-3 carbons and/or oxygenate moieties at the 4 position (6, Figure 2.1). Consistent with

expectations, softwood lignins yield almost exclusively guaiacyl derivatives, while hardwood lignins yield both guaiacyl and syringyl derivatives. Grasses yield guaiacyl, syringyl, and p-hydroxyphenyl derivatives, but also vinylphenol, propenyl-phenols, and p-hydroxybenzaldehyde that are not produced during pyrolysis of softwood and hardwood (*48*, *81*), and are likely derived from ferulate and coumarate esters (*99*). Phenol derivatives are the large majority of the products formed from lignin pyrolysis; aromatic hydrocarbons and some furan derivatives are also detectable, but at very low amounts that might represent lignin sample contaminants (*81*). Lignins from spruce wood with different dimer compositions also show different product distributions, including variations in the yield of major products like guaiacol (*72*). This suggests that bonds between lignin units and the lignin structure determined by those bonds may impact pyrolysis, as well.

2.3.2 Model 2: Secondary reactions catalyzed by inorganic components

The biopolymers that make up the majority of the biomass by weight are established as the primary source of bio-oil products formed during thermal degradation. However, secondary reactions occur during the pyrolysis process involving other components present within the biomass (*48*, *85-91*). As products form, they can interact with catalytic minerals in the residual solid. For example, levoglucosan has been shown to react on minerals present in the residual char from pyrolysis of biomass. The products formed include levoglucosenone, furan derivatives, and lighter oxygenates such as acetic acid, acetone and acetol. Demineralization prohibits the formation of these products. (*86*, *87*)

Different inorganics are responsible for different kinds of secondary reactions. In general, the presence of metal cations enhances the homolytic cleavage of pyranose ring bonds over the heterolytic cleavage of glycosidic linkages, leading to the increased formation of light oxygenate decomposition products at the expense of levoglucosan formation. While Na^+, K^+, Mg^{2+}, and Ca^{2+} all catalyze levoglucosan decomposition, the effects of group 1, alkali metals and 2, alkaline elements differ. Increased Na^+ and K^+ alkali metal loading increased formic acid, glycolaldehyde and acetol more than similar amounts of the alkaline metals, Mg^{2+} and Ca^{2+}, though more furfural is produced with increasing concentrations of Mg^{2+} and Ca^{2+}. Additionally, the alkali metals reduce levoglucosan production at very low thresholds. This suggests that Na^+ and K^+ ultimately promote cracking reactions, while Mg^{2+} and Ca^{2+} promote dehydration reactions (*89*, *100*, *101*)

2.3.3 Model 3: Interactions and linkages between primary components

While the first two models address the direct conversion of biopolymer organic components to related bio-oil products and their further reaction catalyzed by biomass inorganics, the third addresses compositional and structural relationships among cell wall components and their impact on the products. Interactions between polysaccharides and lignin have been shown to alter pyrolysis products (*72*, *92*). The cellulose-lignin interaction can lead to a decrease in levoglucosan yield and an increase in light (C_1-C_3) compounds, especially glycoladehyde and furans. Based on the nature of the small products, Zhang et al. (*92*) hypothesized that the cellulose-lignin interaction occupies the C_6 position, disfavoring glycosidic bond cleavage that is required for the formation of levoglucosan and favoring light compound and furans formation through

ring scission, rearrangement, and dehydration reactions. The strength of this effect on pyrolysis products is most pronounced in grasses, followed by softwood and then hardwood, possibly due to the increased prevalence of covalent bonds between cellulose and lignin in grass cell walls (*102*, *103*). Hemicellulose-lignin interactions, especially the xylan-lignin interaction revealed in NMR experiments (*71*), may also affect pyrolysis. Indeed, enzymatic removal of hemicelluloses from lignin-carbohydrate complexes increased coniferyl alcohol yields (*72*).

An example of a compositional feature that may impact product distribution is the degree of acetylation of the biopolymers. As mentioned, acetyl groups decorate hemicellulose side-chains, and acetyl groups are also present in the lignin. The increased abundance of these groups in biomass correlates with increasing yields of acetic acid, methyl pyruvate, acetone, and furan; additionally, this acetylation correlates with decreasing yields of furfural and acetaldehyde (*47*, *104*). While the acetic acid and perhaps the methyl pyruvate can be explained by the direct production of these compounds upon pyrolytic decomposition (model 1), the nature of the relationship between acetate and the furanic and other 4-carbon species has not been clearly defined. The production of the 4-carbon species may be due to an indirect effect (model 3) or may be the result of catalytic reaction of acetate with itself (model 2).

Several investigations (*94-96*) suggest that feedstock moisture content can also play a role in the yield and product distribution of the organic fraction of the bio-oil. As previously discussed, the presence of water in bio-oil prohibits its direct use and creates challenges to catalytic valorization. For these reasons biomass is typically subjected to drying prior to pyrolysis, which both reduces the required energy of the pyrolysis step

and limits the water in the liquid condensate to water produced by decomposition reactions. However, the degree to which the feedstock moisture content should be eliminated is still under investigation. Burhenne et al. (*95*) found that higher feedstock moisture content led to slightly lower char & gas yields upon pyrolysis with minimal changes to the elemental composition of the char. However, this is in disagreement with Westerhof et al. (*94*), who observed slightly higher char yields with increasing moisture content. The water weight fraction distribution of the feedstocks in the two studies were quite different, 2.4% - 55.4% in Burhenne et al. versus 0% - 20% in Westerhof et al. Beyond impacts to the yields, He et al. (*96*) studied the change in selectivity to the organic fraction produced upon pyrolysis of switchgrass with 5, 10 and 15% feedstock moisture contents. The authors found that at 500°C, the lowest moisture content feedstock produced the highest amounts of levoglucosan and acetic acid. The authors note that while significant differences in pyrolysis products were observed, they could not identify clear trends in their data. Among these studies, the observable but sometimes contradictory or unclear trends suggest that the feedstock moisture content may have multiple impacts on the pyrolysis process, possibly related to the physical location of the water in biomass.

In addition to compositional factors, morphological factors also influence the bio-oil product distributions. Biomass undergoing thermal decomposition retains its morphology even in harsh thermal treatment regimes (*105*). Biomass is a poor conductor of heat (conductivity < 0.1 W m^{-1} K^{-1}) (*106*), and large temperature gradients occur in heated biomass particles (*107*). Most reactor systems for thermal degradation require size reduction of biomass particles; as an example, fluidized beds require

particle sizes no larger than 2 mm (*106*) to ensure rapid reaction. These particle sizes are larger than the tissue structures present in biomass. While the overall tissue and cellular morphology remain intact, micropore formation and shrinkage during the reaction process can occur in a non-uniform manner throughout the biomass (*105*, *108*). Piskorz and colleagues observed decreasing liquid yields with increasing particle size, attributed to increasing incidence of secondary reactions with wood particles (*109*). The principles of internal and external diffusion and the impacts of tortuosity, surface area, and diffusion path lengths are all fundamental to catalytic reaction engineering, and in the case of thermal biomass conversion these important parameters are all dictated by the reacting feedstock (*110*). Some evidence supports the notion that different plant developmental stages, which is related to the ratio of leaves to stems and biomass density, results in different pyrolysis products. For example, switchgrass harvested at later times during the growing season produced increased yields of condensable products, relative to that from younger, leafier material (*111*), though compositional and developmental differences of the starting material were not carefully assessed.

2.4 Conclusions

Years of research have led to understanding of the direct pyrolysis conversion pathways of the major monomeric and polymeric constituents of biomass (model 1, table 2.2). The observation that these constituents often represent minor components in raw bio-oil (table 2.1) highlights the importance of catalytic degradation (model 2) and possibly indirect effects (model 3) on pyrolysis products. The latter model is only recently receiving attention as knowledge of cell wall structures and analytical repertoires blossom (*48*, *92*). Detailed examination of the relationships between components and

products are still sparse, with the biological literature providing detailed characterization of cell wall components, while the engineering literature analyzes the chemical components, or often just total yields, of different pyrolysis fractions. We would argue that further investigations on the relationships between biomass components and thermal products will allow improvement of thermal product "quality." Short of attaining (or improving on) petroleum fuel-like properties, even the criteria for a high quality thermal product remains unclear. As discussed, this is in part because methods for up-grading are so dependent on bio-oil composition. Thus, methods that economically separate and/or simplify the different product streams, while still maintaining C-C bonds and overall C-content, are more likely to be amenable to catalytic upgrading.

Greater and more systematic analysis of biomass composition and pyrolysis products within species that show significant compositional variation will aid in better understanding biomass-bio-oil relationships. Much of the existing literature relies on comparisons of thermal degradation products across diverse taxonomic groups that vary greatly in cell wall composition beyond the biomass components measured (table 2.2). An analysis of more subtle compositional differences, in which compositional factors are varied across different samples may aid in refining biomass-bio-oil relationships. For example, genetic mutants that vary in only one component relative to near isogenic, un-mutated "wild-type" plants can directly address relationships between starting components and products (*82*). In addition to genetically determined compositional differences, biomass composition also depends on growth conditions and developmental stage, which relates to harvest time. Taken together, the scale of the problem points to

the value of developing high-throughput methods to help identify species and genotypes that are most suitable for production of specific thermal products and to guide the optimization of genetic stocks and growth condition for bioenergy crops. Methods available to identify such "high-quality" biomass include near-infrared reflectance spectroscopy (*35*), Fourier transform near-infrared spectroscopy (*112*), and pyrolysis molecular beam mass spectrometry, at least for lignin components (*99*, *113*). In general, these methods can be trained, either rationally or in a model-independent manner, to detect spectroscopic or molecular signatures in biological materials with linear, or non-linear, relationships to thermal products.

Besides selecting or breeding for natural variation in biomass composition (e.g., (*114*, *115*), it is also possible to genetically modify biomass composition (*116*). Most simply, genetic engineering of bioenergy plants can be achieved by modifying the plant's genome to 1) express genes from other organisms, 2) increase expression of native genes, or 3) reduce expression of native genes. More complex schemes are also possible, in which expression patterns of genes are altered through synthetic biology approaches that recombine various genetic elements (e.g., (*117*)). The most common method for plant genetic engineering co-opts the molecular machinery of a bacterial pathogen that introduces genes into plant chromosomes to facilitate its pathogenesis. Genetic engineering to improve bio-oil production would aim to increase biomass components that enhance the yield of favored products and/or to decrease components that produce disfavored products or interfere with up-grading strategies. Advances in understanding cell wall biosynthesis, including genes responsible for synthesizing the major polymer classes (*41*, *51*, *118*) and covalent interactions among them (e.g., (*119*-

121)); regulation of expression of the cell wall biosynthesis genes (*122*); and metal ion transport proteins that determine the abundance and distribution of plant mineral content (*123-125*), lay the foundation for genetically engineering bioenergy crop cell wall content and structure. For example, lignin is an important target for genetic engineering for pyrolysis since the major lignin-derived products have a lower O:C ratio, a higher energy value, and are more stable than sugar-derived products (*33*, *48*). Some important genes that participate in or regulate lignin synthesis have already been modified in energy crops without major interference with plant biomass yield (e.g., (*126*, *127*), reviewed in: (*128*)). However, current genetic engineering strategies are focused on developing low lignin biomass for saccharification and biochemical conversion to fuels. Therefore, more work is required to develop biomass with high lignin content for thermal conversion. Producing corrosive acetic acid in bio-oil (*48*), acetyl groups on cell wall polymers are another potential target for genetic engineering of "pyrolysis crops." Three enzyme classes, including the Reduced Wall Acetylation (RWA) proteins, acetylate cell wall polysaccharides (*121*, *129*, *130*). A mutant of the dicot reference plant, Arabidopsis thaliana, that lacks expression of all four RWA genes, shows a 40% reduction in secondary wall associated acetyl groups (*129*). Reducing expression of this family of genes in bioenergy crops may be help solve the problems caused by acetic acid in bio-oil produced from such plants.

Pretreatments such as washing/leaching and torrefaction are another class of strategies to improve biomass quality by changing biomass composition (*131*, *132*). For example, by washing biomass with detergent (Triton) or acid to remove minerals, the yield of bio-oil is increased and reaction water content is reduced (*132*). Coupling biochemical

conversion of biomass, which depletes the polysaccharide fraction, with pyrolysis of the resulting residue, or bagasse, is another avenue to explore further (*133*, *134*). Torrefaction is a low-temperature (200 to 400°C) thermal pretreatment that decomposes hemicellulose and may segregate disfavored products such as water and acid into intermediate streams before the next stage of pyrolysis (*131*). More efficient torrefaction may be achieved by changing the composition or chemical structure of hemicellulose through genetic methods to further separate the decomposition temperatures of hemicellulose from lignin and cellulose. By identifying and studying the roles of key biomass components during thermal conversion, it will be possible to maximize the economic and environmental benefits of plant biomass derived biofuels in the future.

Chapter 3: Experimental Methods

3.1 Bench scale torrefaction system

A bench scale torrefaction reactor (Figure 3.1), as described in Wan et al. (*135*), was used to perform torrefaction experiments as described in this text.

The reactor was made of a stainless steel tube with an ID of 0.930 in. and length of 12 in. The reactor was externally heated with a one-zone electric furnace, where the actual pyrolysis temperature was monitored by a Type K thermocouple inserted in the center of the reactor tube. Biomass was fed via a Coperion loss-in-weight twin screw auger feeder (*Coperion K Tron Pitman Inc., Sewell NJ*) over the course of one minute through a 1/2 inch stainless steel transport tube. A carrier gas, nitrogen at 300 mL/min, was introduced in the upper part of the transport tube to ensure good transport to the reactor, and to ensure no air was able to get into the reaction zone. A second flow of nitrogen preheated to 270°C was fed from the bottom of the reactor tube at 585 mL/min, for a total carrier flow rate of 885 mL/min, and a space-time of approximately 2 seconds. The reactor pressure was approximately 1 atm.

Biomass was reacted for 20 minutes with a temperature ramp to achieve 270°C for the last five minutes. When the biomass is introduced into this reactor, the measured temperature inside the reactor immediately decreases, and the temperature recovers to 270°C over the course of 15 minutes. The volatiles that evolved were condensed via an ice water trap followed by a liquid nitrogen trap. A wet test meter was used to measure the flow rate for the post-condensation gas mixture, before its release to a vent. After 20 minutes of reaction, the reactor was rapidly cooled and the residual solids were removed. The liquids in the two traps were collected and mixed thoroughly for analysis.

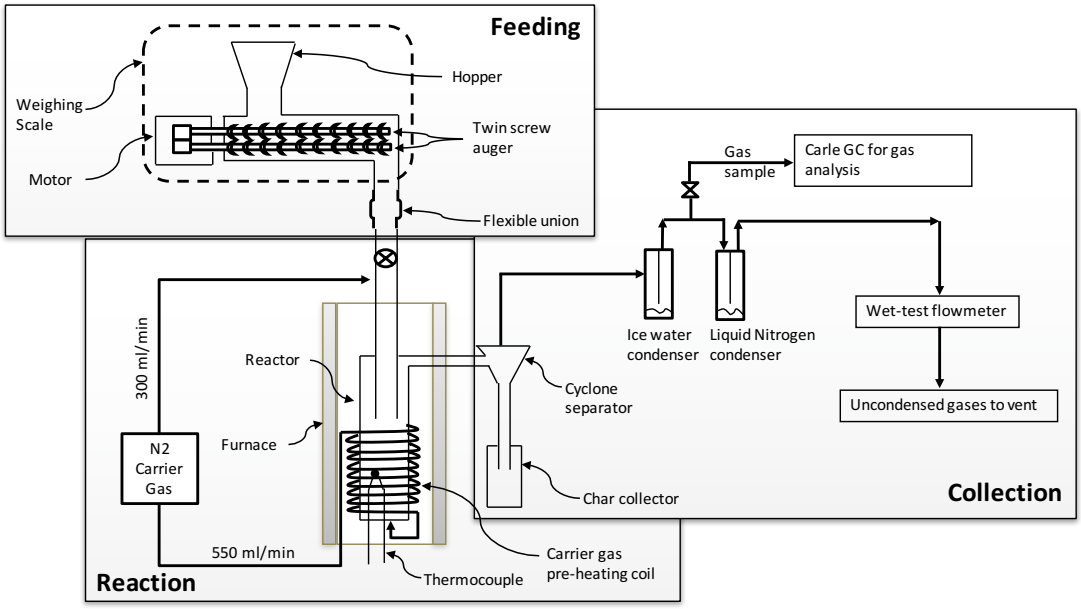


Figure 3.1. Bench scale torrefaction reactor

3.2 Pyrolysis GCMS/FID

3.2.1 Sample Preparation

To prepare biomass samples for analytical pyrolysis or torrefaction between 0.5 mg and 1.5mg of biomass was loaded into a fire polished quartz tube *(CDS Analytical, Oxford PA, Part No. 10A1-3015)* with a filler rod *(CDS Analytical, Oxford PA, Part No. 10A1-3016S)* and quartz wool *(CDS Analytical, Oxford PA, Part No. 0100-9014)* below to prevent the samples from falling out of the bottom (figure 3.2). The quartz tubes were weighed before and after sample loading, and the difference was taken as the mass of the sample in the tube. Larger sample masses (> 1 mg) were used in lower temperature experiments (below 450°C) in order to ensure a sufficient amount of material was detected.

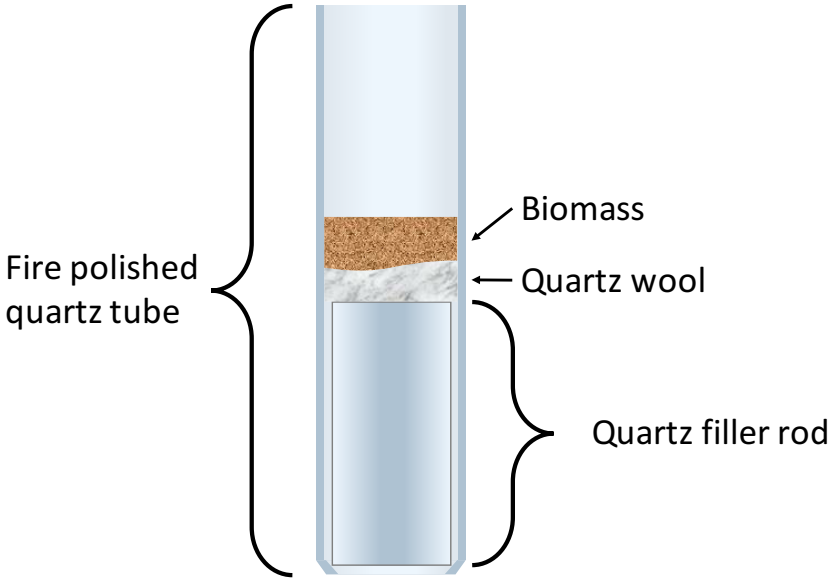


Figure 3.2. Quartz sample tube diagram (not to scale)

3.2.2 Pyroprobe Apparatus

A CDS Analytical Pyroprobe 5250T apparatus with autosampler (figures 3.3 & 3.4) was used to perform thermal decomposition experiments on raw and thermally pretreated biomasses. Samples were placed in the autosampler carousel and the software was configured to select the appropriate calibrated temperature and desired final hold time.

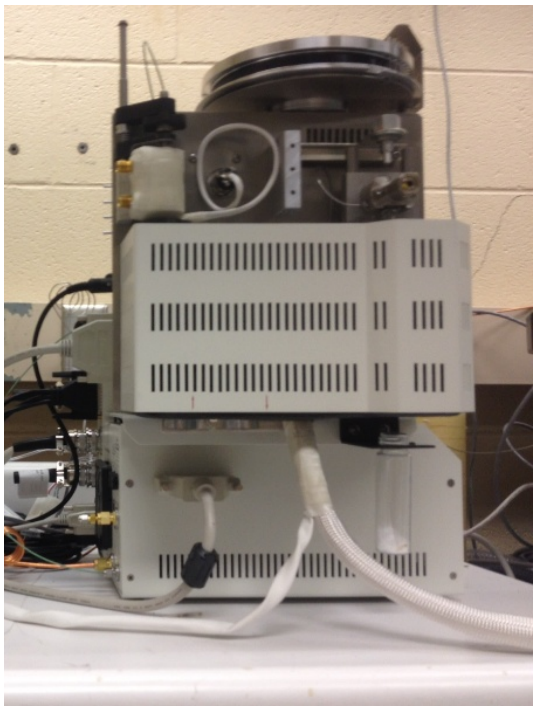


Figure 3.3. Pyroprobe apparatus.

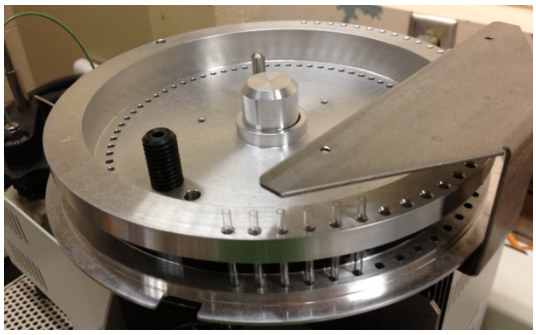


Figure 3.4. Pyroprobe autosampler.

When the GC ready signal was given, the autosampler carousel advanced, dropping the sample tube onto the top of an isolation valve, which rotates to allow the tube to drop further into the pyrolysis chamber (resting on top of a second, lower isolation valve). The top valves was closed and the pyrolysis chamber was then purged with helium for 20 seconds to remove any air which was introduced into the system. Following this purge, the platinum filament was resistively heated to the instrument set point utilizing a 1000°C / second temperature ramp. The evolved vapors passed through a valve oven

heated to 300°C, and were then swept into a 1/16” Silcosteel® (*Restek Corporation, U.S., Bellefonte, PA*) transfer line heated to 300°C which is connected to the injection port of the attached gas chromatograph (described in detail below). After the final hold time elapsed, the filament was deactivated and the lower isolation valve opened to allow the sample tube to be discharged from the instrument by gravity. For samples which have undergone successive thermal treatment regimes, the tube was left in the pyrolysis chamber until the gas chromatography program was completed, at which time the sequence repeats, resuming from the purge step (as no new tube is loaded into the chamber). After a tube had been discharged from the pyrolysis chamber in this manner, the pyrolysis chamber was cleaned by reactivating the filament, which was held at a temperature of 1200°C for 20 seconds while a purge flow of approximately 30 mL/min of helium flowed through the chamber. Samples were held online from 30 to 300 seconds depending on the experiment.

Prior to the use of the pyroprobe apparatus, a calibration was performed with a type K, 1/16" diameter, ungrounded, 6" long thermocouple (*Omega Engineering, Stamford, CT; Part no. TJ36-CASS-116U-6*) to relate the instrument set point to the actual temperature in the chamber. This was necessary because the filament temperature is estimated by electrical resistivity and not direct measurement. A plot of the calibration curve measured for the pyroprobe instrument taken approximately one year apart is shown in figure 3.5. The calibration curve is stable and has not changed appreciably with use over the course of the studies presented in this work.

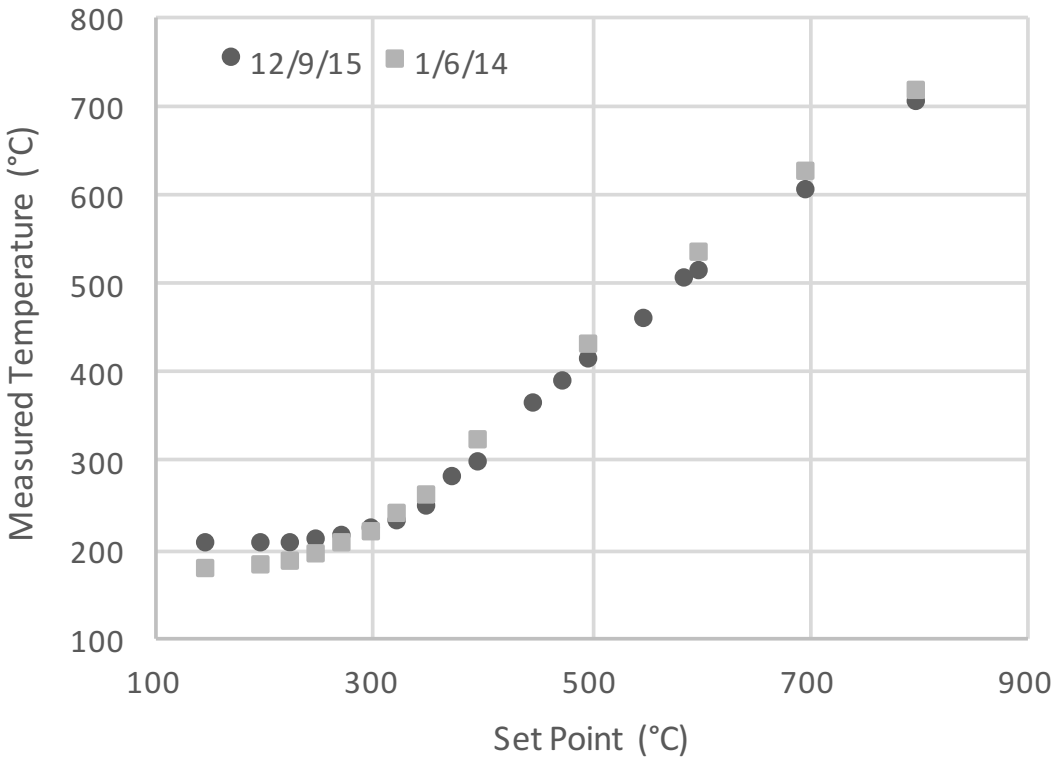


Figure 3.5. Pyroprobe filament temperature calibration curve.

All pyrolysis and torrefaction reactions within the pyroprobe system were carried out in a helium carrier gas at one atmosphere. Flow was controlled by the gas chromatograph attached to the pyroprobe instrument. For experimental temperatures below 450°C, the helium flow rate was 14 mL/min total flow; above 450°C, the total helium flow rate was 94 mL/min. This was to ensure a sufficient amount of thermochemical products were able to be detected, and is determined by the GC flow controller split ratio. A split ratio of 10:1 was used for lower temperature experiments, and a split ratio of 90:1 was used for higher temperature experiments.

3.2.3 Gas chromatography

Evolved vapors from the pyroprobe apparatus were transported via 1/16" Silcosteel® (*Restek Corporation, U.S., Bellefonte, PA*) transfer lines heated to 300°C which was

connected to the injection port of a Shimadzu QP-2010+ GC/MS-FID system *(Shimadzu Corporation, Kyoto, Japan)* with a 60m semi-polar Restek *(Restek Corporation, U.S., Bellefonte, PA)* RTX-1701 column (250µm diameter, 0.25µm film thickness). The injection port temperature was held at 280°C. All analysis presented in this paper used the same column temperature program, mass spectrometer settings, and FID settings, with the difference in split ratio (as noted before) for different temperature regimes in the pyroprobe. The temperature program on the column began at 45°C for 2 minutes, then increased at a rate of 3K min^{-1} for 78.33 minutes to a final temperature of 280°C, where it was held for 20 minutes. The mass spectrometer scanned masses 35.00 to 250.00 at 0.5 seconds per scan. The resulting ion chromatogram was used to identify significant peaks in the chromatogram. The software used was Shimadzu GCMS Solutions version 4.11 (copyright © 1999-2013, Shimadzu corporation) and GCsolution Postrun version 2.41 (copyright © 2000-2010, Shimadzu corporation).

3.2.4 Compound Identification

Two publications by Faix et al (*8*, *9*) were used as the primary means of compound identification. As Faix et al. used a 15m RTX-1701 column, the retention order (but not absolute time) of the pyrolysis products they observed are the same as in this work, and additionally they list the base peak masses with relative abundances of observed masses for the products, facilitating easy identification. In the case that a peak was unable to be identified using these two publications, the peaks were either assigned identifications based on NIST library searches, assigned to compound lumps based on major ions, or left unidentified. Figure 3.6 shows a flowchart of the methodology used to identify compounds.

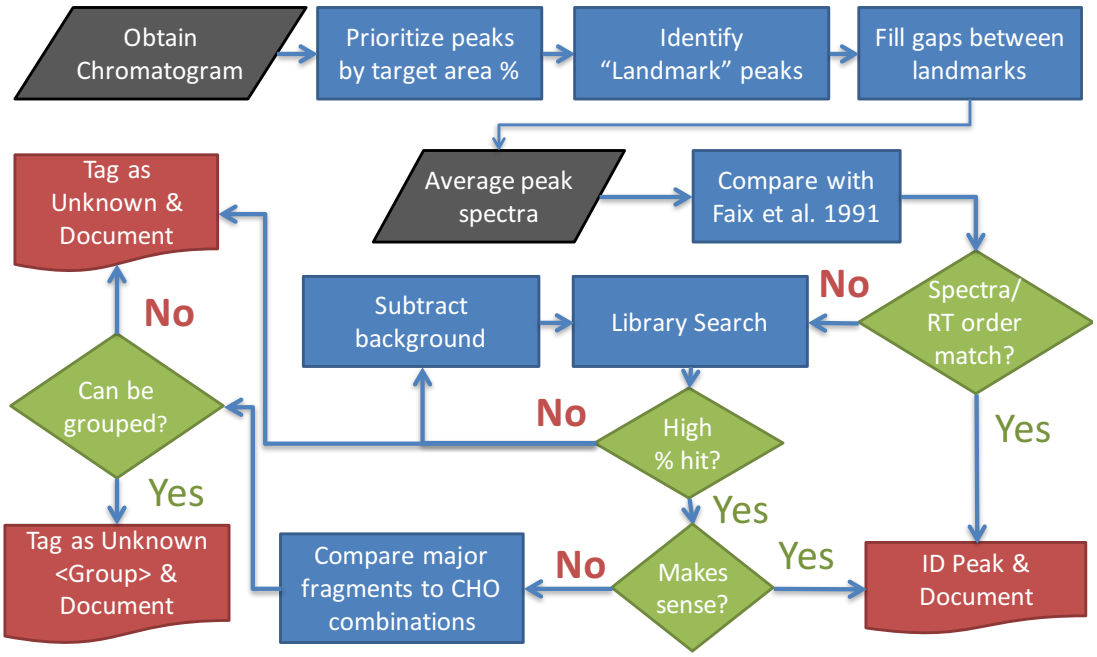


Figure 3.6. Flowchart showing compound identification methodology. "Landmark" peaks include but are not limited to: furfural, acetic acid, and levoglucosan.

The identified peaks from the ion chromatogram were then matched to the corresponding peaks from the FID chromatogram. The areas of these peaks were then determined by integration using the Shimadzu GCsolutions software. Each integrated peak area was then divided by the total mass of the material in the sample tube, thereby normalizing each experiment to the initial amount of sample fed into the pyroprobe.

3.2.5 Lumping approach

Identified compounds were assigned into lumps of compounds based on organic functionalities, in a similar manner as described by Dauenhauer et al. (*136*). The total peak area or carbon content (depending on what is reported) was summed for each of the compounds in the lump. The mean values of these compound lumps across the technical replicate experiments performed are reported throughout this work as 'yield'.

3.2.6 FID Calibration

Calibration of the FID area was performed by injection of varying concentrations of phenol in methanol at known quantities to develop a response curve. For each identified thermochemical product in the chromatograms, relative response factors in the literature were used (when available) in tandem with this calibration curve to calculate the carbon content of each compound.

When response factors were unavailable in the literature, a quantitative structure property relationship (QSPR) model was used to predict the effective carbon number (ECN) for each compound; from this, the unknown response factors were calculated. This model is described in detail in (*137*). Briefly, response factors for many organic molecules were gathered from literature sources, and each carbon atom was assigned a type (i.e., aliphatic, carbonyl, ether, primary/secondary/tertiary alcohol, etc.). A linear model was fit to predict the ECN from the number of each type of carbon atom, and this model was experimentally tested. Measured values were found to be in good agreement with predicted values.

3.3 Water content analysis

Water content analysis (reported as liquid weight percent) was carried out using a TitroLine Karl-Fischer analyzer (TitroLine KF, SCHOTT Instruments). Typically, liquid samples were injected through a septum into the titration cell using a one-way syringe. Before measurement, a calibration titration was performed using deionized water according to the instrument manual. Two drops of each liquid sample (0.1 – 0.2g) were drawn by syringe and injected into the analyzer, then titrated with KF reagent for

volumetric titration until all moisture in the analyzer was consumed. The water content percentage was determined from the KF reagent consumption.

3.4 Catalytic reactions

3.4.1 Catalytic reactor

A CDS pyroprobe model 5250 with autosampler (figures 3.3 & 3.4) was fitted with a separately heated packed bed flow reactor following the pyrolysis chamber (figure 3.7). The separate reactor consisted of 6" long stainless steel tube with ¼" O.D. and 6" long inserted into a block oven. Typically, less than 5mg catalyst would be loaded inside the reactor, mixed with 0.3g acid washed borosilicate glass beads obtained from Sigma Aldrich (*Sigma Aldrich, St. Louis, MO, Part no. G1145*). The zeolite catalyst was pelletized and sieved to a size of 150-212μm, within the same range as the glass beads, prior to mixing with the glass beads to prevent bypassing of the catalyst. The catalyst bed was sandwiched between 2 layers of 50mg quartz wool. The catalyst beds were positioned 2-4cm from the bottom of the reactor tube, which was measured to be the isothermal zone with an independent thermocouple. Product yields resulting from GCMS analysis are defined in terms of GCMS response, which has been defined as the integrated total ion count x 10^{-7} for ease of presentation.

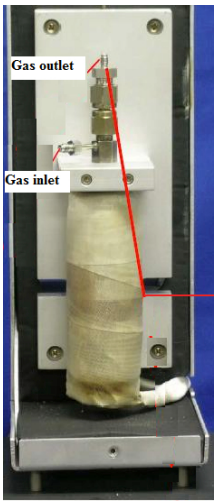


Figure 3.7. External catalytic reactor.

3.4.2 Catalyst characterization

The powder X-ray diffraction (XRD) pattern of the zeolite was recorded on a Rigaku automatic diffractor (Model D-MAX A) with a curved crystal monochromator and Cu Kα radiation operated at 40kV and 35mA. The morphology of the zeolite was evaluated by high resolution scanning electronic microscopy (SEM) using a Zeiss-NEON FEG-SEM system.

3.4.3 Coke on spent catalyst

The coke on the spent catalysts was estimated by temperature programmed oxidation using a Netzsch STA 449 F1 Jupiter attached to a Netzsch ZMS 403C Aeolos TGA system to analyze off gasses. Due to the presence of glass beads mixed with the spent catalyst bed, an extra layer of 150mg of alumina (γ-phase 99.97%, from Alfa Aesar) was placed in the crucible with the spent catalyst/glass beads mixture on top of the alumina such that at elevated temperatures, the glass would melt into the alumina and not adhere to the crucible. The instrument was heated from room temperature to 1000°C at a rate of 10 K min^{-1} under flow of 244.3 mL min^{-1} 50% air (grade zero) and

Argon (ultra high purity) from Airgas. The total CO_2 released during the burning was calculated based on calibration data and used to estimate the coke amount. Each spent catalyst was evaluated after contact with 10 pulses of biomass, corresponding to approximately 6.5mg.

3.4.4 Permanent gas analysis

Permanent gas analysis was conducted via online sampling of the vent line with a portable quadruple mass spectrometer (MKS Microvision Plus LM76). The MS was operated with a scan mode, with a range of mass/ion ratio from 1 to 60, and a speed of 2.5 second per scan period. A series of standard gas mixtures consisting of 5% CO, CO_2, and CH_4 respectively obtained from Airgas were used for calibration.

3.5 Biomass characterization

3.5.1 Oak cell wall composition

Cell wall composition was measured for the oak samples using methods that were essentially described previously (*120*). All measurements were made in triplicate. Briefly, matrix polysaccharides (hemicellulose and pectin) were depolymerized with 2 M trifluoracetate (TFA), diluted in weak base, and quantified relative to standards by high performance anion exchange chromatography. The TFA-insoluble pellet was dissolved with sulfuric acid and quantified via anthrone assay. Acetyl bromide soluble lignin determination utilized the extinction coefficient for aspen wood (*138*). (This information provided by Fan Lin who performed the characterization.)

3.5.2 Switchgrass NIRS Characterization

Dry switchgrass biomass was ground with a Thomas Wiley® Mini-Mill with 20-mesh. The samples were directly used for NIRS analysis. Two NIRS models previously

reported are used for NIRS analysis. One NIRS model is built on switchgrass samples and the other one is built on forage samples (*35*). The same switchgrass assays were used for high throughput assays including pyrolysis molecular-beam mass spectrometry (PyMBMS) (*99*) and enzymatic sugar release (*139*). Table 3.1 lists the variables measured.

Variable	**Description**	**Source**
ADF	Acid detergent fiber	(*140*)
ADF.forage	Acid detergent fiber (forage equation)	(*141*)
ADL	Acid detergent lignin	(*140*)
ARA	Arabinose	(*140*)
ASH	Minerals (total ash)	(*140*)
Ash.forage	Minerals (total ash) (forage equation)	(*141*)
AX	ARA + XYL	(*140*)
AXMG	ARA+XYL+Man+GAL	(*140*)
C	Carbon	(*140*)
Ca	Calcium	(*140*)
CA.forage	Calcium (Forage equation)	(*141*)
CAL	Calories	(*140*)
CP	Crude Protein	(*140*)
CP.forage	Crude Protein (Forage equation)	(*141*)
CWC	Cell wall concentration (KL+UA+RHA+FUC+ARA+XYL+MAN+GAL+GLC+ PCA + FEST + FETH)	(*140*)
CWE	Cell wall ethanol	(*140*)
CWEP	Theoretical ethanol conversion efficiency from cell wall hexosans	(*140*)
DM	Dry Matter	(*140*)
DM.forage	Dry Matter (Forage equation)	(*140*)
dNDF48	Digestible NDF 48	(*140*)
dNDF48.forage	Digestible NDF 48 (forage equation)	(*141*)
Dry Matter	Dry Matter	(*140*)

Table 3.1. Switchgrass compositional features.
Continued on following pages.

Variable	Description	Source
EE	Extracted fat	(*140*)
ETOH	Ethanol/g dry forage	(*140*)
FA	Ferulic acid	(*140*)
FEST	Esterified ferulates	(*140*)
FETH	Etherified ferulates	(*140*)
FRU	Fructose	(*140*)
FUC	Fucose	(*140*)
GAL	Galactose	(*140*)
GLC	Glucose	(*140*)
GLCS	Soluble glucose	(*140*)
glucose.nrel	Glucose (NREL Measurement)	(*139*)
HCA	Hydroxycinnamates	(*140*)
HEX	Total hexoses ((MAN + GAL + GLC)(180/162)) + NSC)	(*140*)
HEXE	Theoretical ethanol from hexoses [MAN + GAL + GLC)*0.57] + [0.51(GLCS + FRU)] + [0.537*SUC]	(*140*)
HEXEP	Hexose ethanol extraction efficiency (ETOH/HEXE)/100	(*140*)
IVDMD	In-vitro dry matter digestibility	(*140*)
IVDMD.forage	In-vitro dry matter digestibility (forage equation)	(*141*)
K	Potassium	(*140*)
K.forage	Potassium (Forage equation)	(*141*)
KL	Klason lignin	(*140*)
lignin ABSL.AP13 %	Acetyl Bromide soluble lignin	Not published
Lignin.forage	Lignin (Forage equation)	(*141*)
lignin.nrel	Lignin (NREL Measurement)	(*115*)
MAN	Mannose	(*140*)
Mg	Magnesium	(*140*)
MG.forage	Magnesium (Forage equation)	(*141*)
N	Nitrogen	(*140*)
NDF	Neutral detergent fiber	(*140*)
NDF.forage	Neutral detergent fiber (Forage equation)	(*141*)
P	Phosphorus	(*140*)
P.forage	Phosphorus (Forage equation)	(*141*)
PCA	p-Coumarate esters	(*140*)

Variable	Description	Source
PENT	Pentose sugars released/g dry forage	(*140*)
PENTP	Pentoses extraction efficiency (0.88 PENT / (ARA + XYL)) * 100	(*140*)
PPEN	Pentose proportion of total carbohydrates (1 - HEX/SUG))*100	(*140*)
RHA	Rhamnose	(*140*)
SC	Total soluble carbohydrates (SUC + GLCS + FRU)	(*140*)
sgratio.nrel	S/G ratio in lignin	(*115*)
STA	Starch	(*140*)
SUC	Sucrose	(*140*)
TDN	Total Digestible Nutrients	(*140*)
TDN.forage	Total Digestible Nutrients (forage equation)	(*141*)
total_sugar.nrel	Total sugar	(*139*)
UA	Uronic acids	(*140*)
XYL	Xylose	(*140*)
xylose.nrel	Xylose (NREL measurement)	(*139*)

3.5.3 Switchgrass cell wall chemical assay

For cell wall chemical assay, samples were ground to fine powder and made into destarched cell wall material as previously described (*120*). The chemical assays and cell wall components measured are described previously (*120*) except the acetyl content measurement. The acetyl content in cell wall material was released by 0.5 M NaOH for 1 h at room temperature with shaking at 500 rpm and measured by an Acetic Acid Assay Kit (Megazyme, K-ACET). (This information provided by Fan Lin who performed the characterization.)

Chapter 4: Decoupling HZSM-5 Catalyst Activity from Deactivation During Upgrading of Pyrolysis Oil Vapors

A paper published in ChemSusChem

Shaolong Wan, Christopher Waters, Adam Stevens, Abhishek Gumidyala, Rolf Jentoft, Lance Lobban, Daniel Resasco, Richard Mallinson, and Steven Crossley

4.1 Abstract

The independent evaluation of a catalyst's activity and its stability during the catalytic pyrolysis of biomass is challenging because of the nature of the reaction system and rapid catalyst deactivation that force the use of excess catalyst. In this contribution we use a modified pyroprobe system in which pulses of pyrolysis vapors are converted over a series of HZSM-5 catalysts in a separate fixed-bed reactor controlled independently. Both the reactor-bed temperature and the Si/Al ratio of the zeolite are varied to evaluate catalyst activity and deactivation rates independently both on a constant surface area and constant acid site basis. Results show that there is an optimum catalyst-bed temperature for the production of aromatics, above which the production of light gases increases and that of aromatics decrease. Zeolites with lower Si/Al ratios give comparable initial rates for aromatics production, but far more rapid catalyst deactivation rates than those with higher Si/Al ratios.

4.2 Introduction

The main processes under advanced development for the upgrading of biomass pyrolysis products involve two catalytic approaches: treating the pyrolysis vapors with acidic zeolites and/or hydrotreating the liquid product. The configurations employed for the upgrading of the pyrolysis vapors are varied and include combinations of the addition of catalysts within the pyrolyzer, in situ catalytic pyrolysis, (*142*) (*143*) and,

less commonly, ex situ upgrading in the vapor phase immediately following the pyrolyzer (*135*, *144*). By contacting the vapors with a catalyst prior to condensation as a liquid, the primary pyrolysis products are stabilized before they condense and oligomerize in the liquid phase to produce heavy products that present challenges for subsequent upgrading.

One distinct advantage of the use of zeolite catalysts for the conversion of bio-oil vapors lies in the absence of an external H_2 requirement for the process. However, compared to processes that incorporate an external source of H_2, this benefit comes at the expense of sacrificing the yield to upgraded liquid products per gram of biomass fed as pyrolysis vapors are hydrogen deficient. The low effective H/C ratio of the pyrolysis vapors results inherently in coke formation and catalyst deactivation, which accompany any upgrading process in the absence of an additional H_2 source (*145*). As a result of this, the design of catalysts that are less prone to deactivation and coke formation is critical. Among the various types of zeolites used for the catalytic upgrading of biomass pyrolysis vapors, the most commonly studied catalyst is HZSM-5 (*146-148*). Compared to other catalysts, HZSM-5 produces the highest yields of the preferred products, alkylated aromatics, if bio-oil vapors are treated (*149*). However, HZSM-5, similar to all other catalysts in this environment, deactivates relatively rapidly.

For a variety of reactions over zeolites, catalyst deactivation is highly dependent on catalyst properties such as the density of acid sites. For example, upon studying propylene oligomerization, Bell et al. observed dramatic decreases in conversion and increases in heavy aromatic species on the catalyst over HZSM-5 as the Si/Al ratio of

the zeolite decreased (*150*). The tradeoff between activity and deactivation rates with acid site density has not been fully investigated in the area of catalytic pyrolysis. Rapid catalyst deactivation associated with catalytic pyrolysis has made researchers use excess catalyst with catalyst/biomass feed ratios much greater than unity. As the products are generally collected over a period of several minutes and excess catalyst is used, it is difficult to measure intrinsic catalytic activity and rates of deactivation. Instead, yields to desired products are typically reported, which are a combination of both phenomena (*151*). Distinguishing the role of reaction parameters and catalyst properties in catalytic pyrolysis is also clouded by the fact that the vapors are typically condensed as liquids and aged in a trap, in which unconverted oxygenates can undergo secondary reactions, often for a period of several minutes or hours before analysis. Pyrolysis GC units, in which a small amount of biomass is pyrolyzed with the vapors sent directly to a GC for analysis, allow the direct measurement of primary pyrolysis products with no liquid aging and minimal subsequent reactions. This is a valuable experimental technique for the evaluation of the pyrolysis process itself as well as the upgrading of the pyrolysis vapors. Such units have been used extensively to screen the effectiveness of catalysts for the catalytic pyrolysis of biomass. Unfortunately, a great disadvantage of such processes lies in the way they are traditionally used to evaluate catalyst activity. In most cases, the biomass is mixed physically with excess catalyst in the pyrolysis tube (*152*). As there is no carrier gas flow through the pyrolysis tube, the residence time of pyrolysis vapors and contact time with the catalyst are very difficult to control and are dependent on the catalyst particle size, thermal conductivity, etc. Huber et al. pointed out the differences in partial pressure and contact times between this

approach and a typical fluidized bed as well as the exaggerated secondary reactions to produce naphthalenes that occur if the catalyst is placed directly in the pyrolysis tube. A more straightforward approach is to place the catalyst bed outside of the pyrolysis chamber, such that the pyrolysis vapors pass over the catalyst bed before they enter the GC (*153*). This approach gives valuable information on the catalyst deactivation rates and independent control of the catalytic reaction conditions versus pyrolysis conditions. In this contribution, a custom configuration of a CDS Analytical pyroprobe pyrolysis unit coupled with a separate catalytic reactor was constructed to evaluate the performance and deactivation of the catalysts upon contact with pyrolysis vapors. The product yields and catalyst stability were measured at various temperatures and Si/Al ratios independently for HZSM-5 catalysts under conditions in which excess catalyst was not used to decouple the activity from the stability of each catalyst. Catalysts are compared based on a constant catalyst mass and constant acid sites, and the coke on spent catalysts was quantified to better understand the role of acid site density on catalyst deactivation.

4.3 Experimental

Biomass used was locally sourced red oak sawdust, with an estimated ash content of 2%, determined by calcination in a TGA unit after after oxidation at 800°C. The samples were ground to a size of 0.25–0.45mm and then dried in vacuum (0.02 MPa) at 60°C overnight before use. Typical red oak compositions can be found in the literature with lignin, hemicellulose, and cellulose wt % of 21.3, 46.9, and 27.2, respectively. Levoglucosan (≥99%, ACROS Organics™) was obtained from Fisher, and the particles used in this work were sieved between 0.16-0.25mm. The three HZSM-5 catalyst

samples (CBV8014, CBV5524, and CBV2314) were obtained from Zeolyst with a Si/Al ratio of 40, 25, and 11.5 respectively. The catalysts were first calcined in flowing air at 600°C for 5 hours starting with a heating rate of 3°C/min to get the proton form, and then pelletized to particles with a size ranging from 0.16 to 0.25 mm.

All catalytic reactions were performed in the reactor as described in section 3.4. The catalyst performance was evaluated by exposing the catalyst to the vapors produced by the pyrolysis of a succession of 0.6-0.8mg pulses of biomass with the catalytically treated products of each pulse analyzed in sections 3.2.3 – 3.2.6.

4.4 Results and Discussion

4.4.1 Catalyst characterization

XRD and SEM results (Figures 4.1 and 4.2) demonstrate that all of the HZSM-5 samples retain their crystallinity after calcination and have a crystallite size of 100–500 nm. The acid site density was calculated based on the theoretical Si/Al ratio from inductively coupled plasma (ICP) analysis provided by the manufacturer. It has been shown experimentally that for HZSM-5 catalysts, the number of Brønsted acid sites is equal to their theoretical value (*154*). This has also been demonstrated for commercial ZSM-5 samples obtained from Zeolyst (*155*).

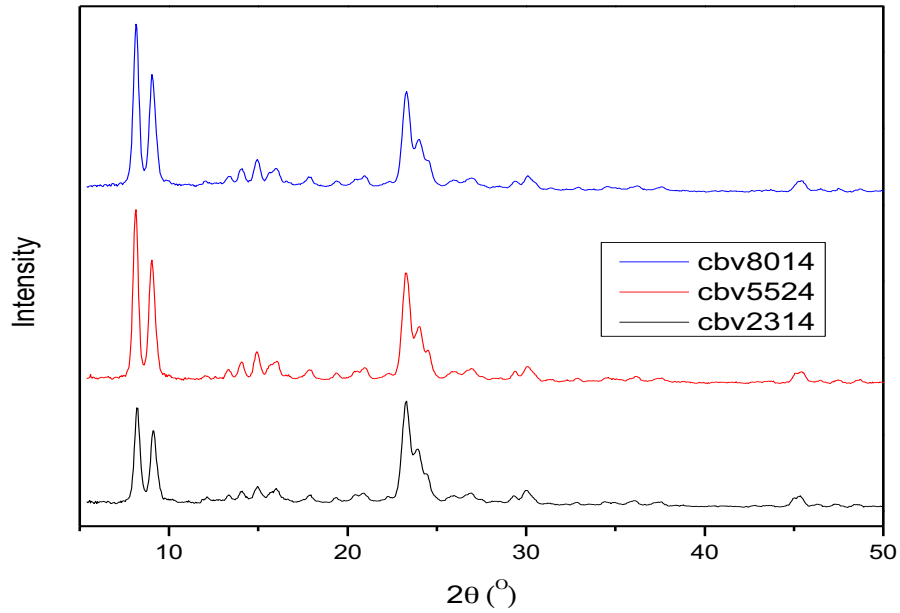
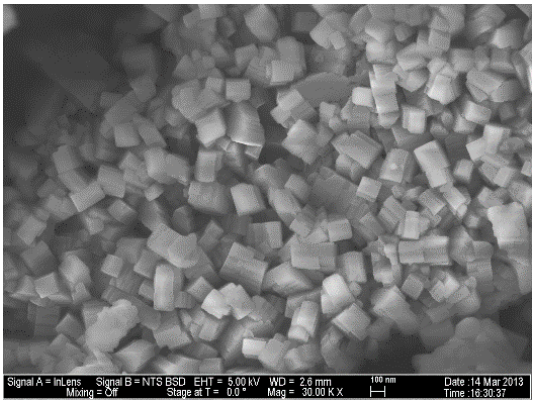
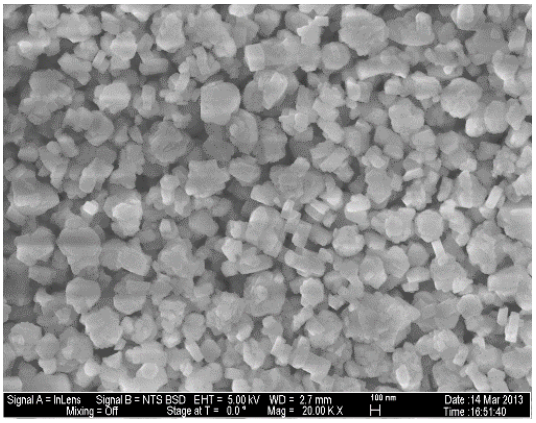


Figure 4.1. XRD of cbv2314, 5524 and 8014

a



b



c

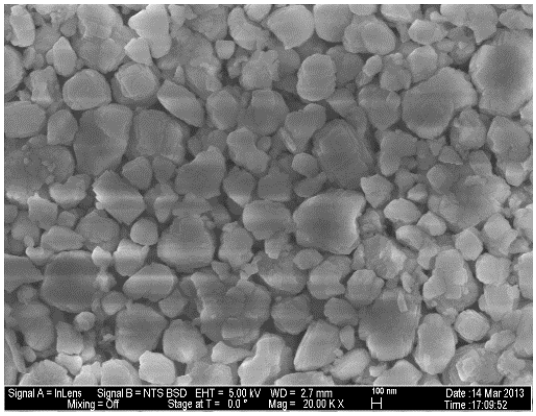


Figure 4.2. SEM image of cbv2314 (a), 5524 (b) and 8014 (c).

4.4.2 Role of catalyst temperature

As most catalytic pyrolysis experiments are conducted with the catalyst mixed physically with the biomass, modifying the temperature has an impact on both the pyrolysis and catalysis, which makes it difficult to decouple their respective roles. The formation of alkylbenzenes, which consist of single-ring aromatics (principally benzene, toluene, xylene, and mesitylene) over 4.46 mg of HZSM-5 in a separate reactor with Si/Al = 40 at temperatures of 400–600°C is shown in Figure 4.3. Pyrolysis conditions were maintained at 500°C in each case. No aromatics were produced in the blank (pyrolysis experiments with no catalyst). Of the three temperatures, the greatest aromatic yield is observed at 500°C, followed by 600°C, and the least at 400°C. The gas analysis of the primary light gas peaks quantified by MS is shown in Figure 4.4. An increased level of light gases are produced at 600°C compared with the lower temperatures, which likely explains the decrease in the yield of aromatics under these conditions. The amount of gases reported here is in general agreement with that found in the literature for fluidized-bed catalytic pyrolysis, in which increased gas yields are obtained at increased temperatures (*151*).

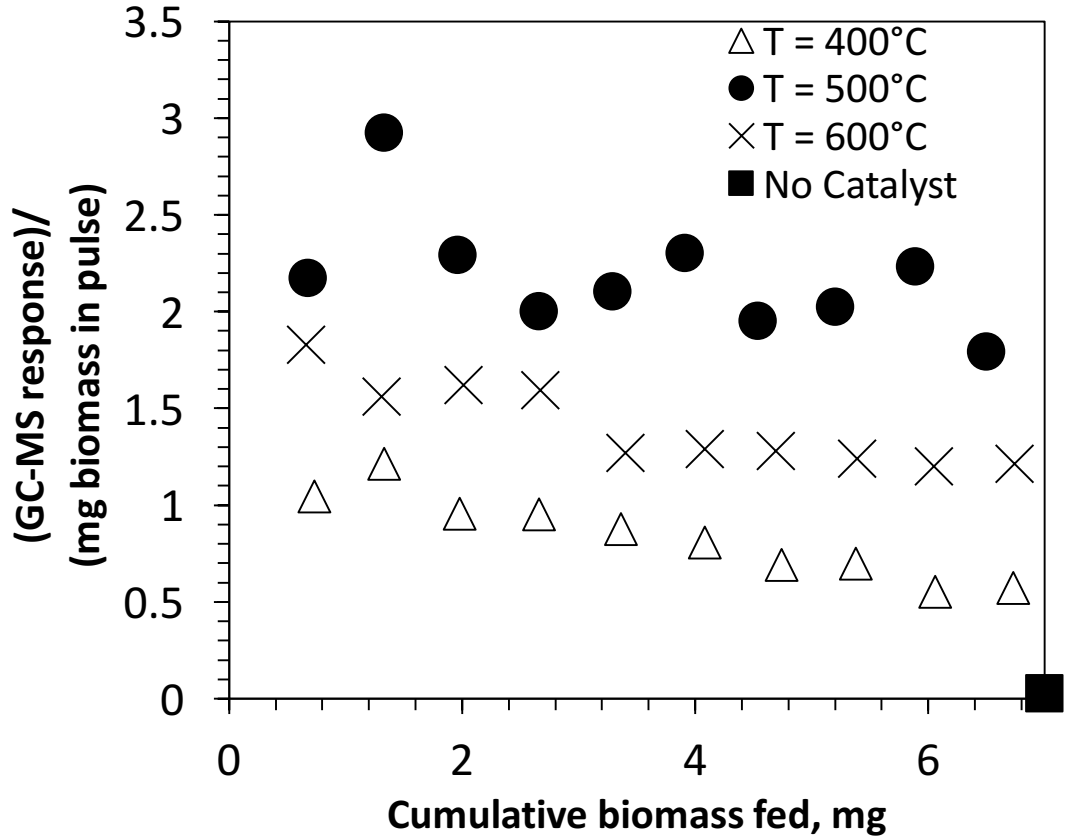


Figure 4.3. Comparison of alkyl benzenes production on HZSM-5 with Si/Al=40 from oak pyrolysis vapors at catalyst temperatures of 400, 500 and 600°C.

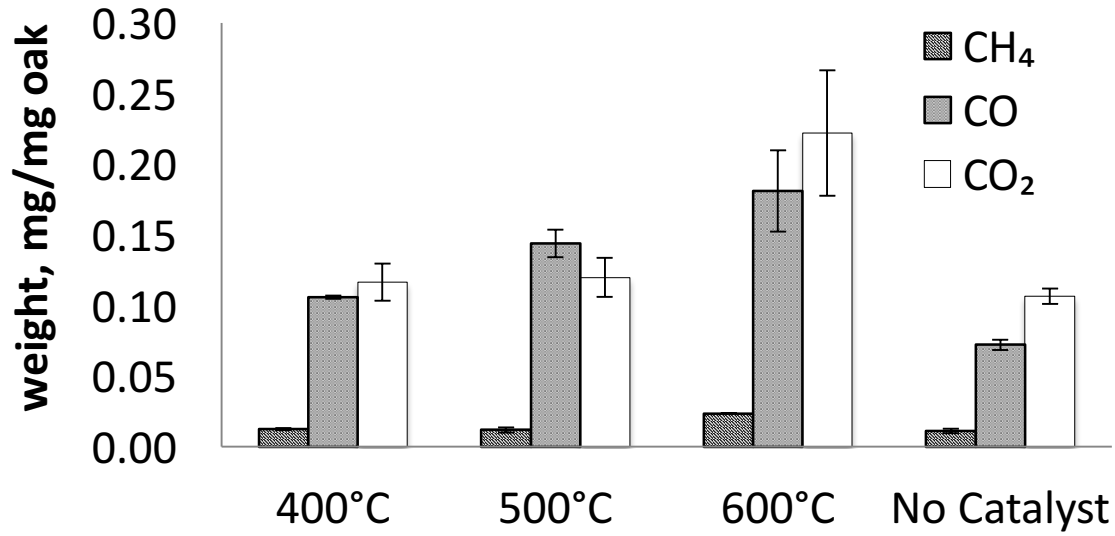


Figure 4.4. Gas yields from the initial pulse of pyrolysis with 4.46mg HZSM5 of Si/Al = 40 under three different temperatures.
Pyrolysis was conducted at 500˚C while the temperature of the catalyst was varied from 400 to 600˚C.

The role of the catalyst is decoupled from that of the pyrolysis process for the production of light gases by separating the catalyst bed from the pyrolysis chamber. Based on these results, it appears that CH_4 formation is not influenced by the presence of a zeolite catalyst at 400 or 500°C, which is to be expected because of the b-scission pathway promoted by the zeolite catalysts used. An increase in CH_4 yield occurs at 600°C, which could be because of thermal reactions, enhanced pyrolytic cracking at elevated temperatures, or a combination of both. In addition, the yield of CO and CO_2 are roughly double in the catalytic case because of decarboxylation and decarbonylation of the oxygenated pyrolysis products. This increase in CO and CO_2 at increased temperature agrees with the increased conversion of light pyrolysis compounds (such as acetic acid and other light oxygenates) as well as lignin-derived methoxyphenols, all of which show essentially full conversion at 600°C and a greater abundance exits the reactor as the reactor temperature is decreased (Figures 4.5-4.7). An increase in the conversion of pyrolysis vapors with catalyst temperature is also evident from the yield of alkylbenzenes. For example, at 400°C the catalyst is not active enough to produce a high yield of aromatics with the amount of catalyst provided, whereas at 600°C a greater production of light gases results in a decreased aromatics yield compared to that at 500°C. A comparable trend is observed with the larger two-ring aromatics (Figure 4.8). The yields of the additional intermediate compound groups, which include alkyl phenols, furans, and indanes, as well as the relatively stable product levoglucosan are shown in Figures 4.9 – 4.12

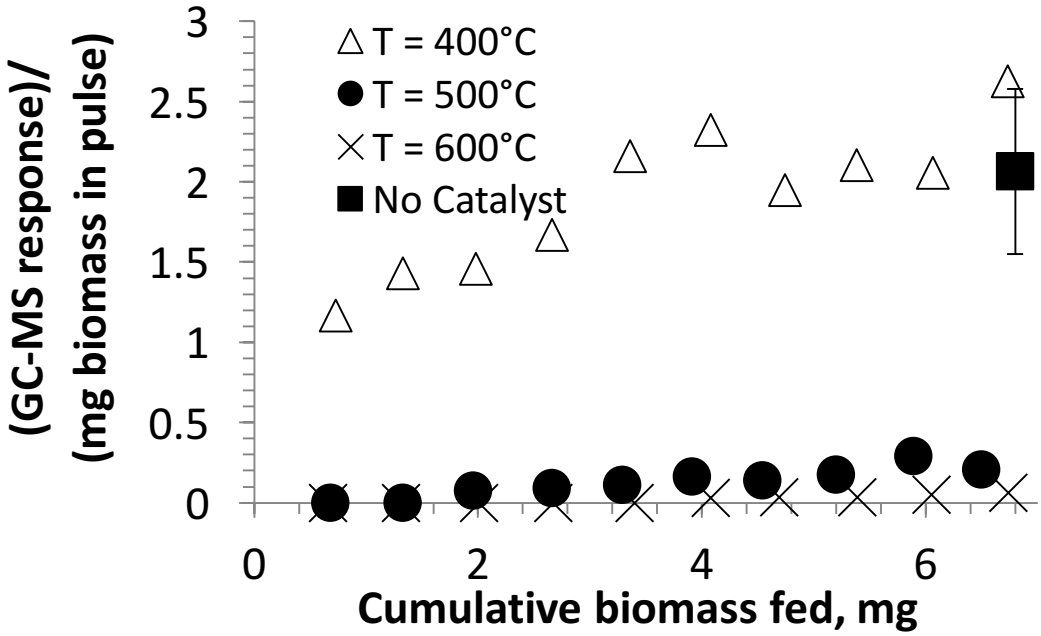


Figure 4.5. Comparison of acetic acid conversion on HZSM-5 with Si/Al=40 from oak pyrolysis vapors at catalyst temperatures of 400, 500 and 600°C.

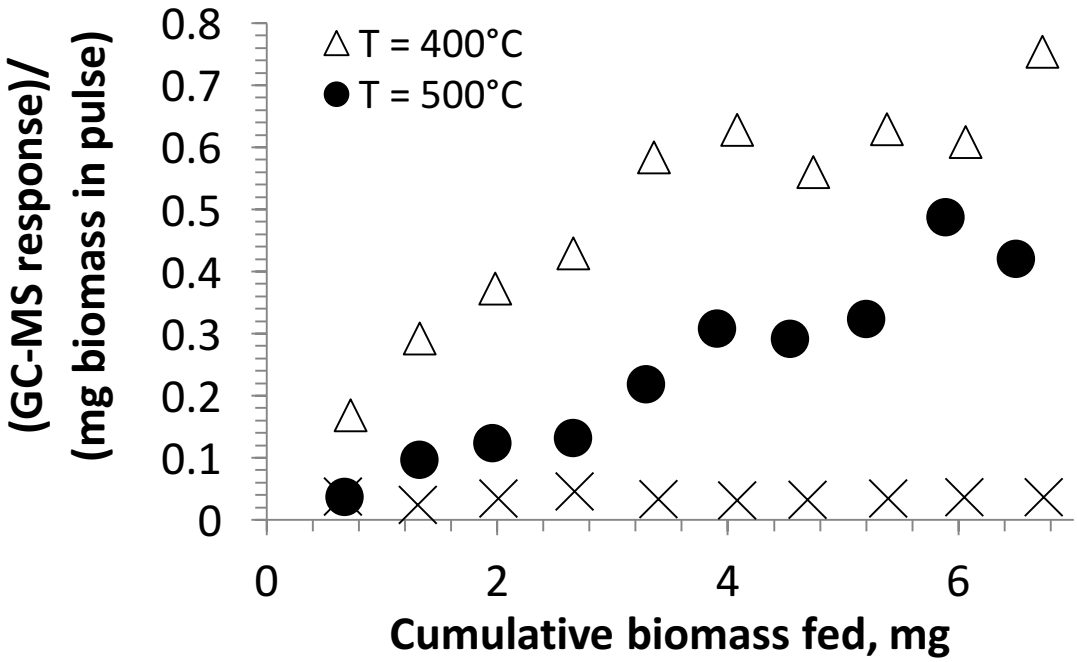


Figure 4.6. Comparison of methoxyphenolics conversion on HZSM-5 with Si/Al=40 from oak pyrolysis vapors at catalyst temperatures of 400, 500 and 600°C. Compounds in this group mainly include Creosol, Acetovanillone, Syringaldehyde, and Methoxyeugenol etc., which contain one, or more methoxy group attached in the benzene ring.

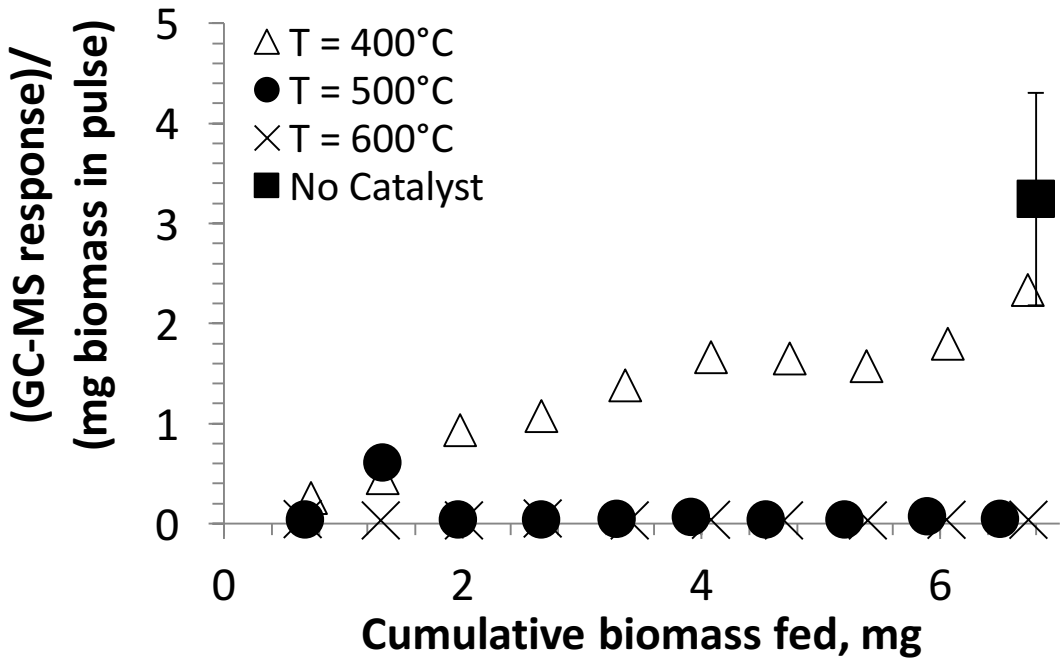


Figure 4.7. Comparison of light oxygenates conversion on HZSM-5 with Si/Al=40 from oak pyrolysis vapors at catalyst temperatures of 400, 500 and 600°C.

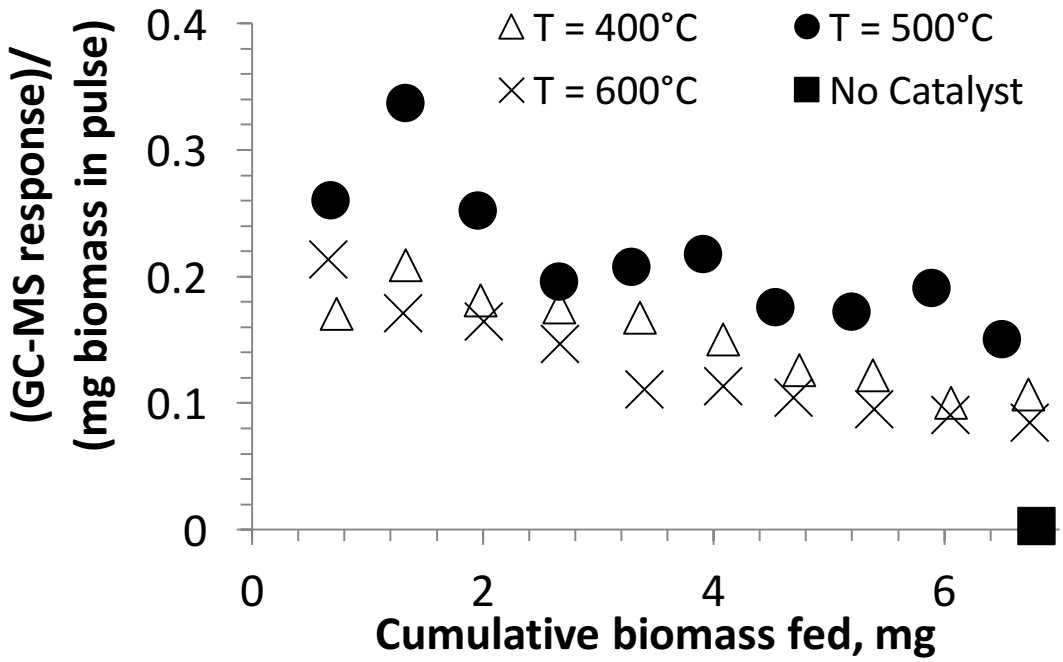


Figure 4.8. Comparison of Napthalenes production on HZSM-5 with Si/Al=40 from oak pyrolysis vapors at catalyst temperatures of 400, 500 and 600°C.

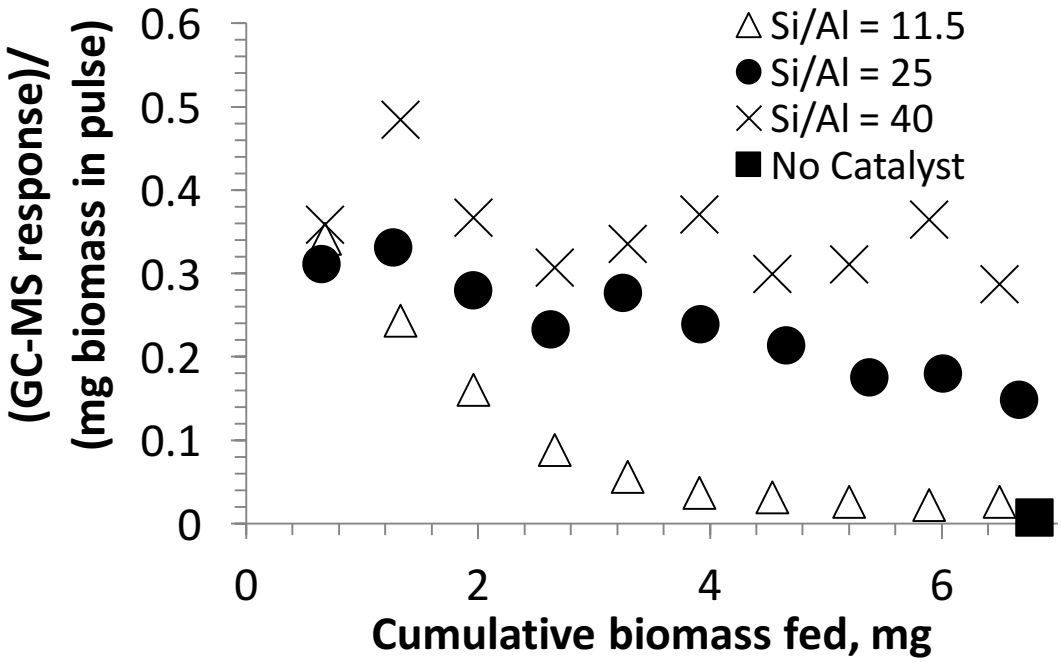


Figure 4.9. Comparison of indane/indene production on HZSM-5 for Si/Al=11.5, 25 and 40 from oak pyrolysis vapors with the fixed total sites at 500°C.
Compounds in this group include indane, Indene, Dimethyl Indene, and 2,3-dihydro-1H-Inden-1-one.

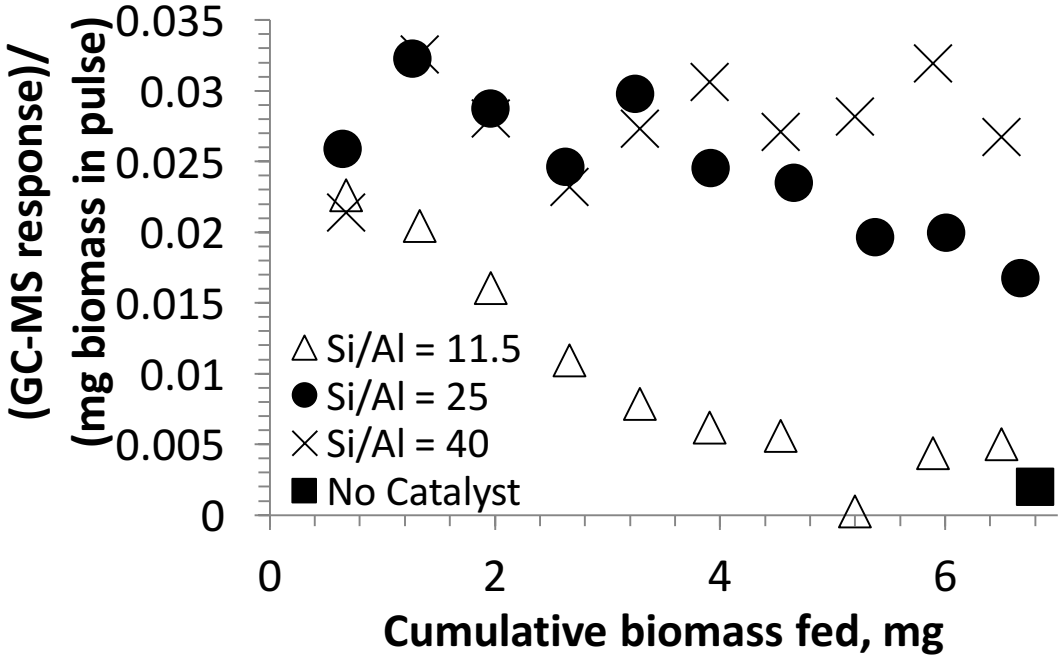


Figure 4.10. Comparison of Benzofuran production on HZSM-5 for Si/Al=11.5, 25 and 40 from oak pyrolysis vapors with the fixed total sites at 500°C.

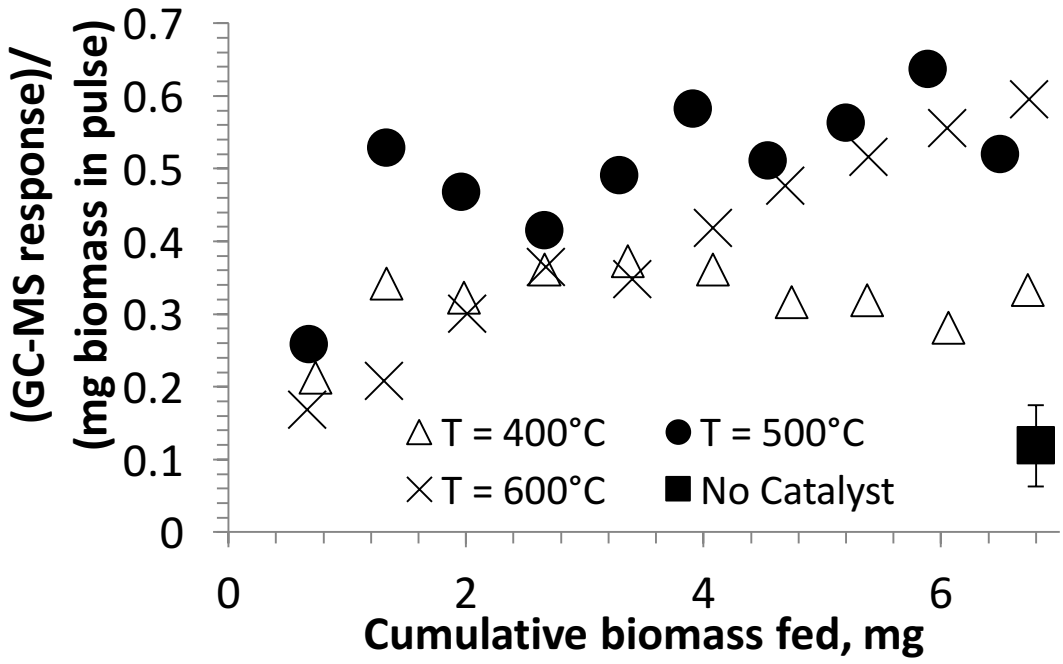


Figure 4.11. Comparison of Alkyl Phenols production on HZSM-5 with Si/Al=40 from oak pyrolysis vapors at catalyst temperatures of 400, 500 and 600°C.

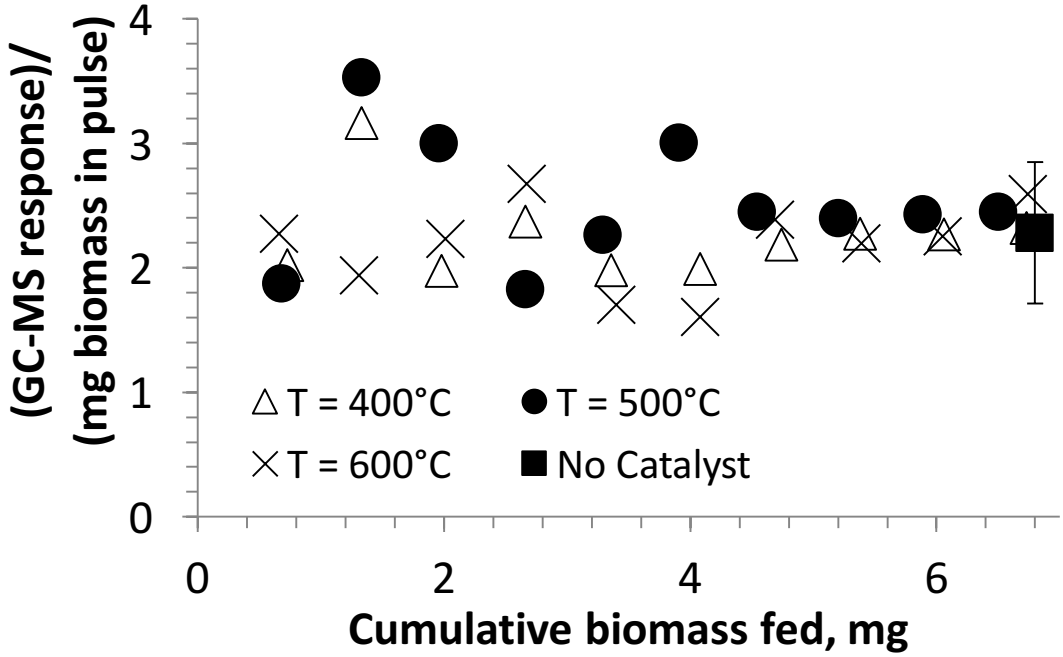


Figure 4.12. Comparison of Levoglucosan/sugars production on HZSM-5 with Si/Al=40 from oak pyrolysis vapors at catalyst temperatures of 400, 500 and 600°C.

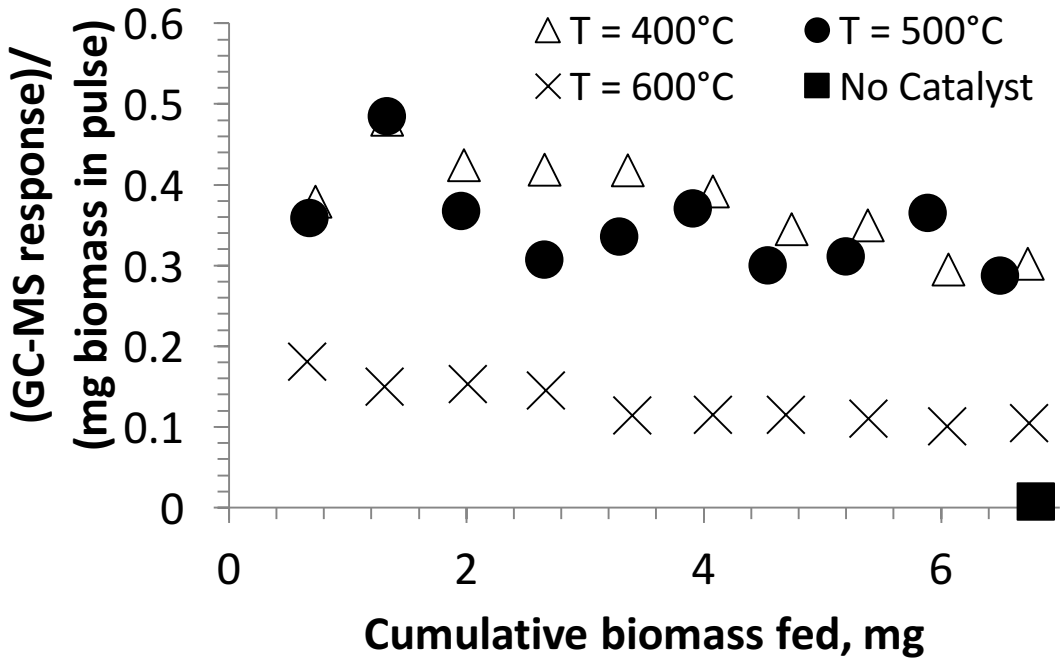


Figure 4.13. Comparison of Indanes production on HZSM-5 with Si/Al=40 from oak pyrolysis vapors at catalyst temperatures of 400, 500 and 600°C.

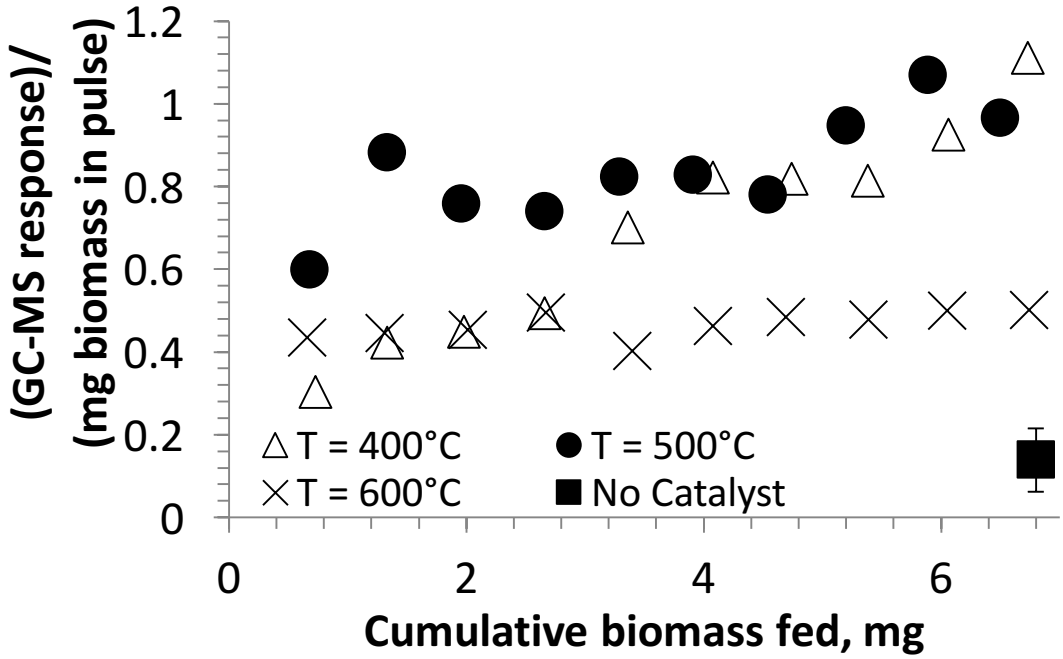


Figure 4.14. Comparison of Furans production on HZSM-5 with Si/Al=40 from oak pyrolysis vapors at catalyst temperatures of 400, 500 and 600°C.

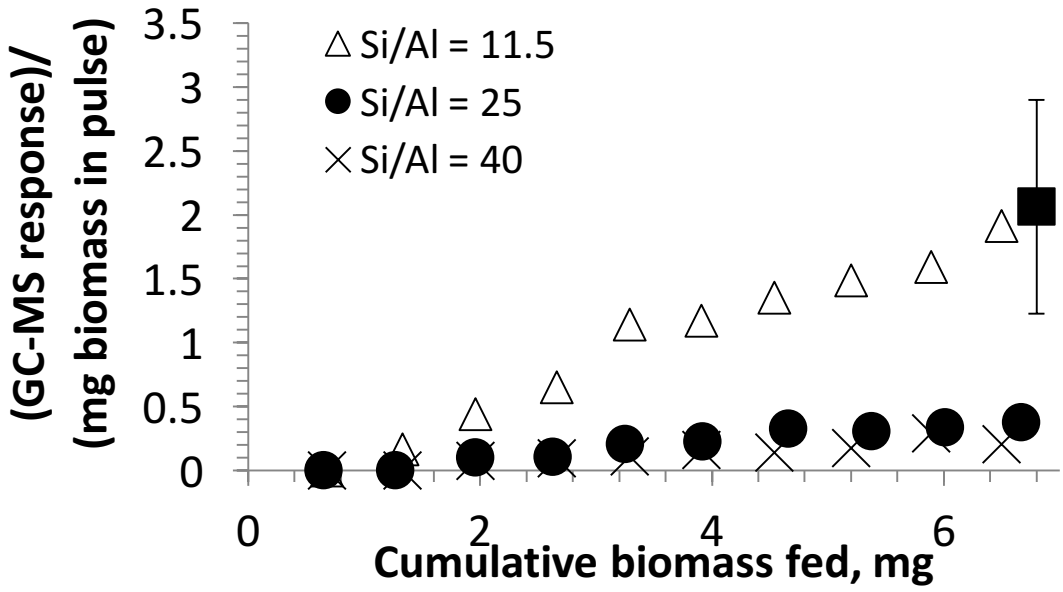


Figure 4.15. Comparison of acetic acid conversion on HZSM-5 for Si/Al=11.5, 25 and 40 from oak pyrolysis vapors with the fixed total sites at 500°C.

The amount of coke was analyzed by temperature-programmed oxidation, and the results are reported in Table 4.1. A trend of a decreased amount of coke on catalysts with increased reaction temperature is observed, in agreement with results reported in the literature (*151*). This significant decrease in coke at higher temperatures appears to correspond with the production of more light gases.

Catalyst Si/Al	Temp. [^{o}C]	Catalyst mass [mg]	Amount of coke [μg mg_{oak}^{-1}]	[μg $mg_{catalyst}^{-1}$ mg_{oak}^{-1}]
Temperature Series				
40	400	4.46	44.6	10
40	500	4.46	42.4	9.5
40	600	4.46	22.3	5
Constant Catalyst Mass Series				
40	500	2.83	21.2	7.5
25	500	2.83	36.8	13
11.5	500	2.83	18.4	6.5
Constant Acid Sites Series				
40	500	4.46	42.4	9.5
25	500	2.83	36.8	13
11.5	500	1.36	6.8	5

Table 4.1. Comparison of coke on spent catalysts after 10 pulses of biomass injected.

4.4.3 Varying acid site density

The alkylbenzene yields that result from the conversion of pyrolysis vapors over three different commercial zeolites with various Si/Al ratios are shown in Figure 4.16. Repeated pulses of biomass were pyrolyzed in which the vapors passed over the catalyst. The sum of the masses of these pulses is reported on the x axis, analogous to increasing time-on-stream. The pyrolysis temperature and the catalytic reaction temperature were maintained at 500°C in each case. A reaction temperature of 500°C was chosen because of the optimal alkylbenzene yields obtained at this temperature, as discussed in the previous section. So as not to introduce additional bias, the number of acid sites was maintained at a constant level of 1.81 mmol for each catalyst bed by appropriate variation of catalyst mass. Interestingly, the yield of alkylbenzenes is comparable for the initial pulse in each case. This implies that the initial activity of the HZSM-5 zeolites is comparable for the production of alkylbenzenes independent of the acid site density. This comparable initial activity is significant and would not be obtained through traditional experiments conducted in situ in which catalyst deactivation rates and catalyst activity contribute simultaneously to product yields.

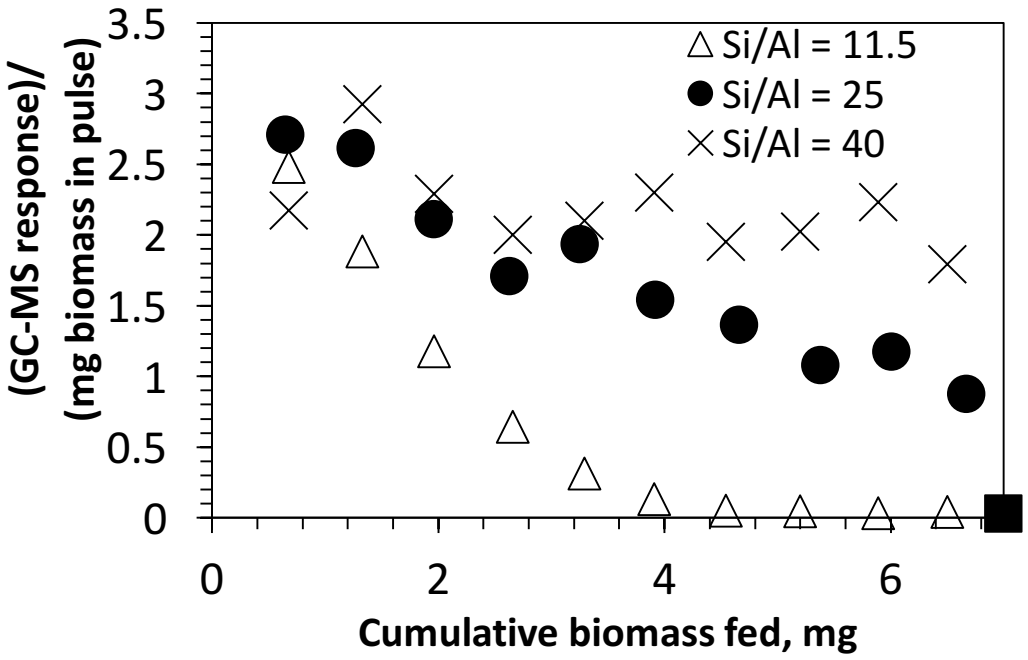


Figure 4.16. Comparison of alkyl benzenes production on HZSM-5 for Si/Al=11.5, 25 and 40 from oak pyrolysis vapors with the fixed total sites at 500˚C.

The deactivation rates of the three catalysts are markedly different, with increased deactivation rates for the production of alkylbenzenes that result from increased acid site density. These trends were similar for the other aromatic product groups: naphthalene, indane/indene, benzofuran, and their derivatives (Figures 4.8 - 4.10). A similar trend in deactivation can be observed for methoxyphenols (Figure 4.17) and light oxygenates (Figure 4.18). The trends in the deactivation rates for intermediate compounds alkylphenols, furans, and furan derivatives are shown in Figures 4.19, 4.20, and 4.21, respectively. Levoglucosan again demonstrates high stability across all catalysts (Figure 4.22). These results show that higher acid density, with a constant total number of acid sites (acid density mass), results in a severe deactivation of the catalyst. This result is not completely surprising as similar behavior has been observed by Bell et al. upon the conversion of propylene over HZSM-5 obtained from Zeolyst, the same

source used in these experiments (*150*). They observed decreased activity and a greater potential for aromatics formation over the catalysts with in-creased acid site density, which they attributed to a higher population of oligomers in close proximity that led ultimately to aromatics production. For the feedstocks considered here, a closer proximity of sites may influence the potential for condensation reactions such as aldol condensation (*156*) that could lead ultimately to pore plugging and catalyst deactivation.

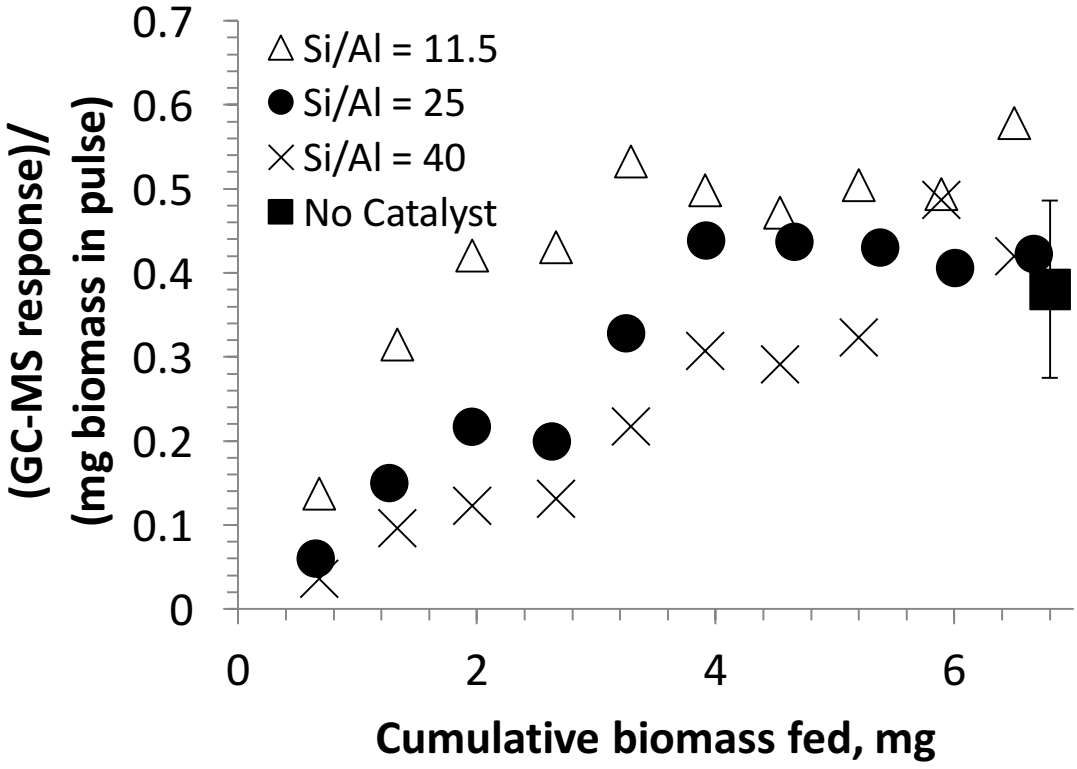


Figure 4.17. Comparison of methoxyphenols conversion on HZSM-5 for Si/Al=11.5, 25 and 40 from oak pyrolysis vapors with the fixed total sites at 500°C. Compounds in this group mainly include Creosol, Acetovanillone, Syringaldehyde, and Methoxyeugenol etc., which contain one, or more methoxy group attached in the benzene ring.

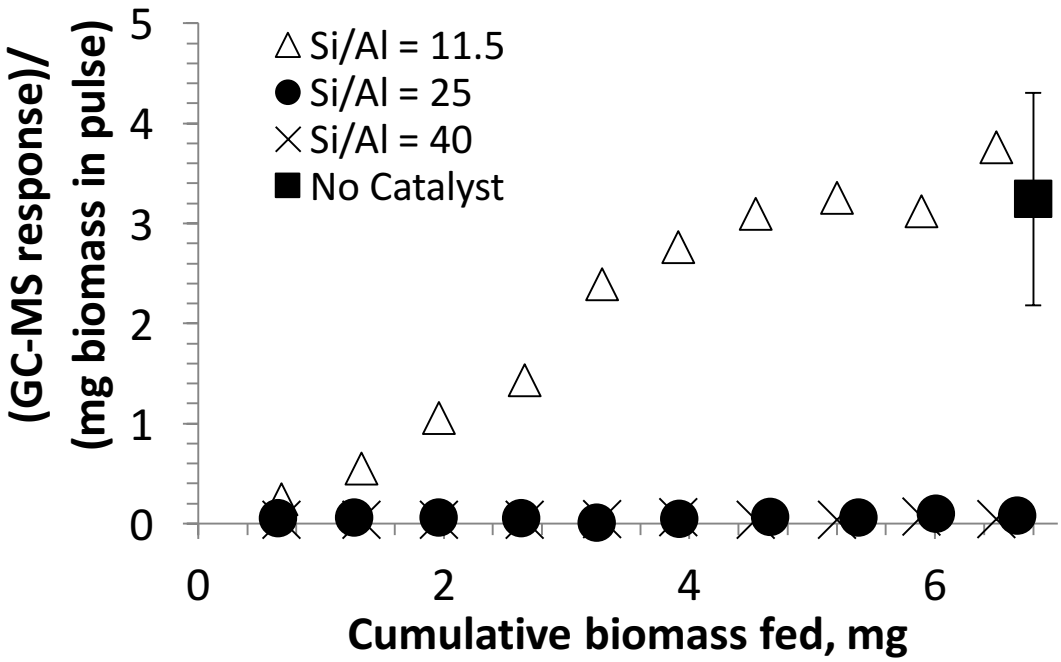


Figure 4.18. Comparison of light oxygenates conversion on HZSM-5 for Si/Al=11.5, 25 and 40 from oak pyrolysis vapors with the fixed total sites at 500°C.
Compounds in this group include 2,3-butanedione, acetol, Hydroxybutanal, 2,3-Pentanedione, etc. Acetic acid and intermediate oxygenate products like acetone, furan and methyl-furan were not listed in this group.

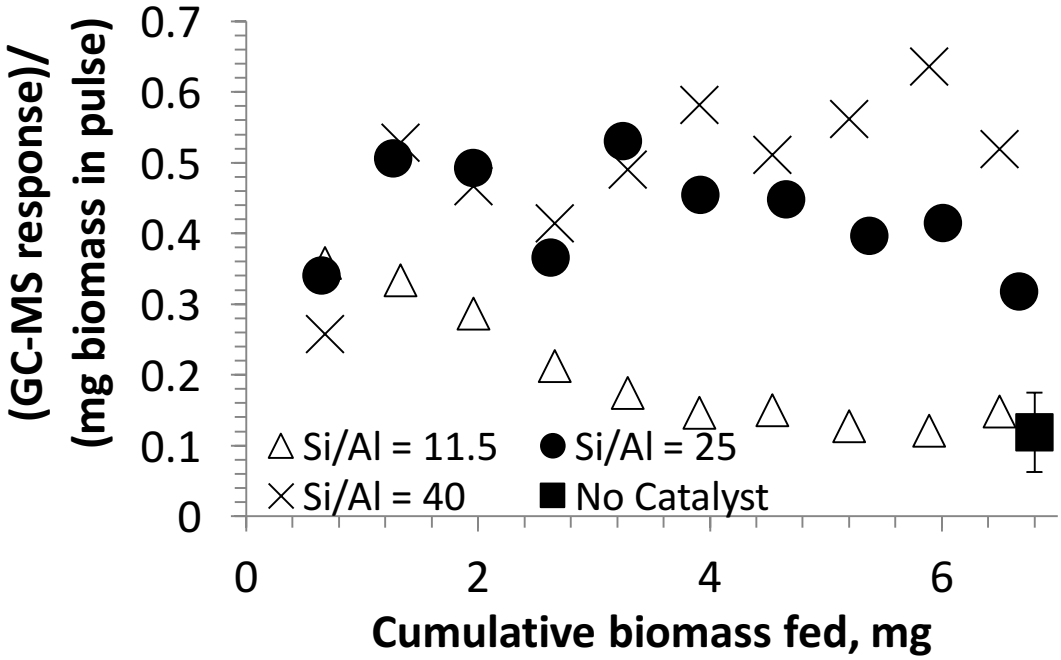


Figure 4.19. Comparison of alkyl phenols production on HZSM-5 for Si/Al=11.5, 25 and 40 from oak pyrolysis vapors with the fixed total sites at 500°C.
Compounds in this group mainly include Phenol, Methyl Phenol, and Dimethyl Phenol.

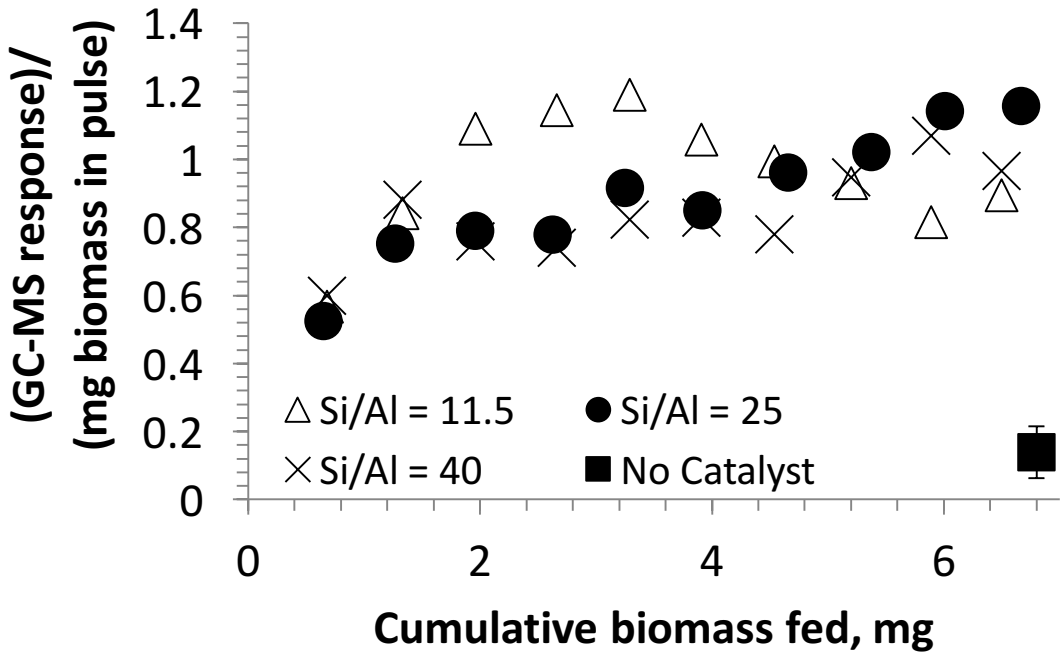


Figure 4.20. Comparison of furans production on HZSM-5 for Si/Al=11.5, 25 and 40 from oak pyrolysis vapors with the fixed total sites at 500°C.
Compounds in this group include Furan and Methyl Furan.Acetic acid and intermediate oxygenate products like acetone, furan and methyl-furan were not listed in this group.

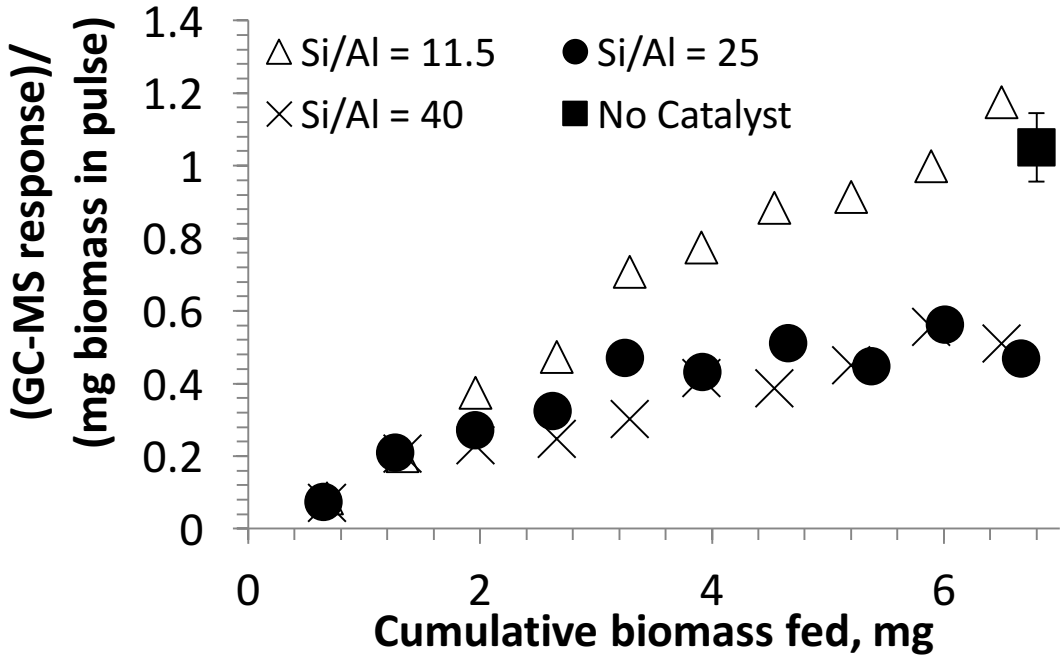


Figure 4.21. Comparison of furan derivatives production on HZSM-5 for Si/Al=11.5, 25 and 40 from oak pyrolysis vapors with the fixed total sites at 500°C.
Compounds in this group include 2-Cyclopenten-1-one, 2-Cyclopenten-1-one, 2,3-dimethyl-, and Furfural.

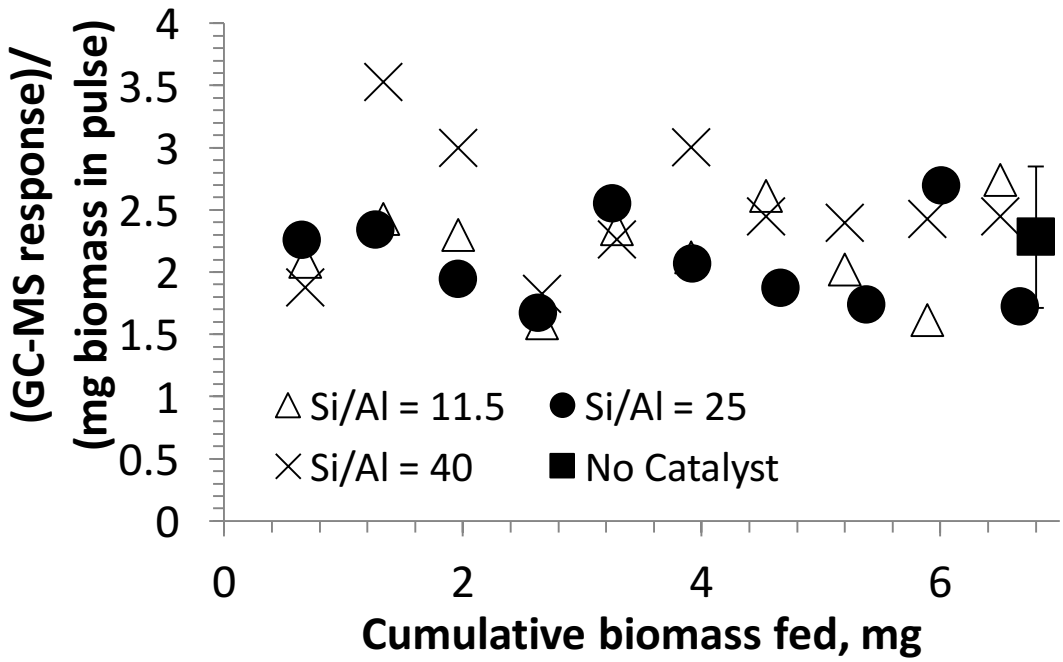


Figure 4.22. Comparison of levoglucosan conversion on HZSM-5 for Si/Al=11.5, 25 and 40 from oak pyrolysis vapors with the fixed total sites at 500°C.

One may suggest that this phenomenon could be a result of changes in acid site strength with different Si/Al ratios, which would impact product distribution. However, a number of studies have suggested that the strength of acid sites does not change over the range of Si/Al ratios used in this study (*157-159*). Another possible explanation could be that varying degrees of extraframework Al may exist in these samples, which could serve as nucleation sites for coke formation. It has been shown that internal defects have a strong correlation with deactivation rates for reactions with oxygenates over HZSM-5 such as methanol to olefins (*160*). Two studies that used MFI samples from Zeolyst led to conclusions that the fraction of extraframework Al per total Al in the sample is very low and does not change with the Si/Al ratio (*150*, *161*). As these catalysts are compared on a constant acid site basis, each catalyst should have a comparable amount of extraframework Al.

Notably, the three catalysts with different Si/Al ratios share a similar pore size distribution per unit mass as they have the same MFI structure. Thus, in the case of higher-acid-density catalysts, a lower mass results in a lower total surface area. This may make these catalysts more susceptible to pore plugging, which could cause faster deactivation.

To examine this further, another set of experiments was performed with the same mass, 2.83 mg, of each of the three catalysts, to give similar surface areas and pore volumes for each catalyst. The Si/Al = 40 catalyst has a lower initial yield of alkylbenzenes after the first pulse because of the lower number of acid sites but shows rather stable performance over the 10 pulses of exposure to vapors (Figure 4.23). Si/Al = 11.5 (which contains two times the number of total acid sites than Si/Al = 25 and 3.3 times that of Si/Al = 40) shows a stable aromatics yield at a high level in the first four pulses, after which deactivation proceeds at a faster rate than that in the other two catalysts. The Si/Al = 25 catalyst demonstrates an intermediate deactivation rate.

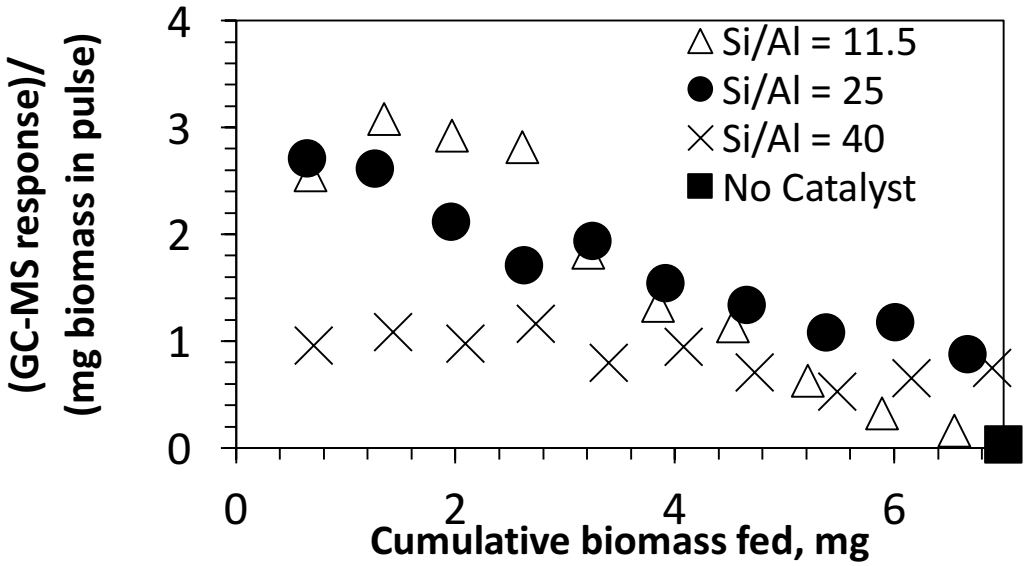


Figure 4.23. Comparison of alkyl benzenes production on HZSM-5 of Si/Al=11.5, 25 and 40 from oak pyrolysis vapors with fixed catalyst mass at 500°C.

Clearly, the initial stable yield for the Si/Al = 11.5 catalyst is caused by an excess of catalyst and is not an indication of the absence of deactivation. This trend is more evident if the Si/Al = 11.5 catalyst is compared at two different mass loadings (Figure 4.24), for which the initial yield of alkylbenzenes is constant between the two series. The catalyst deactivates continuously in both studies, but the excess catalyst results in a constant level of aromatics production. This result is significant, especially if we consider that most studies of catalytic pyrolysis use excess catalyst and far greater catalyst/biomass ratios than those used here. Attempts to quantify catalyst deactivation under these conditions may lead to the incorrect conclusion that the catalysts are more stable than they actually are.

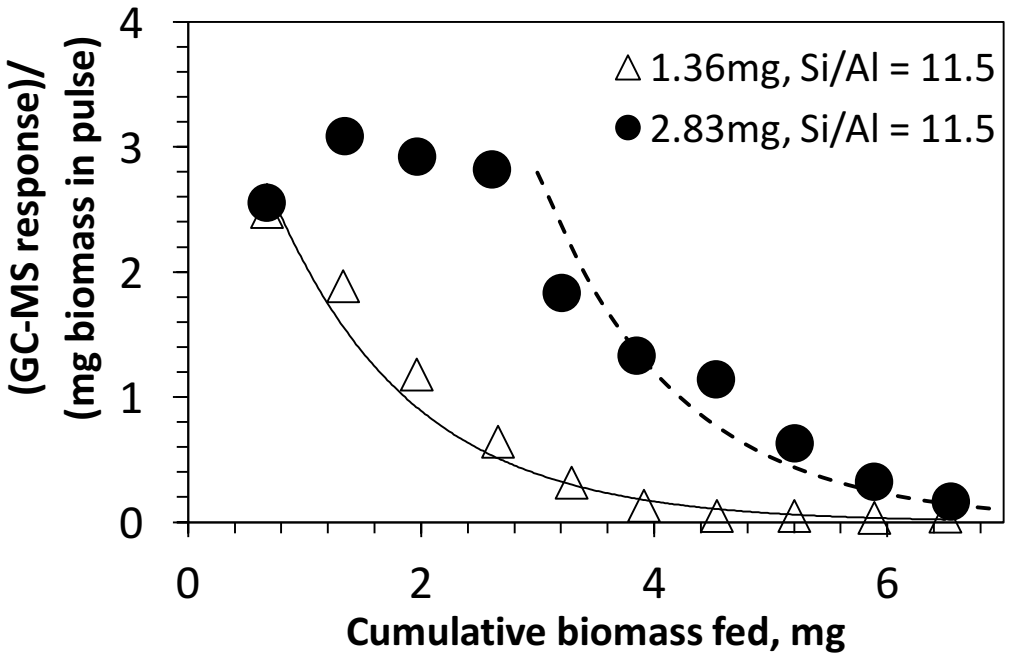


Figure 4.24. Alkyl benzenes exiting the reactor over HZSM-5 of Si/Al=11.5 from oak pyrolysis vapors with fixed catalyst mass at 500°C.
Catalyst mass was varied from 1.36mg to 2.83mg.

Once the excess catalyst is deactivated, after four pulses, the deactivation becomes observable as a decrease in product yields. These results demonstrate clearly that it is

the higher acid density that leads to a more rapid deactivation rate. The plateau in aromatics yield with excess catalyst is proposed to be caused by the complete conversion of the precursors for aromatics products. For example, the light oxygenates and acetic acid are converted fully for the initial pulses over 2.8 mg of HZSM-5 Si/Al = 11.5, and these pyrolysis compounds begin to appear after pulse 4, as shown in Figure 5.25 for acetic acid.

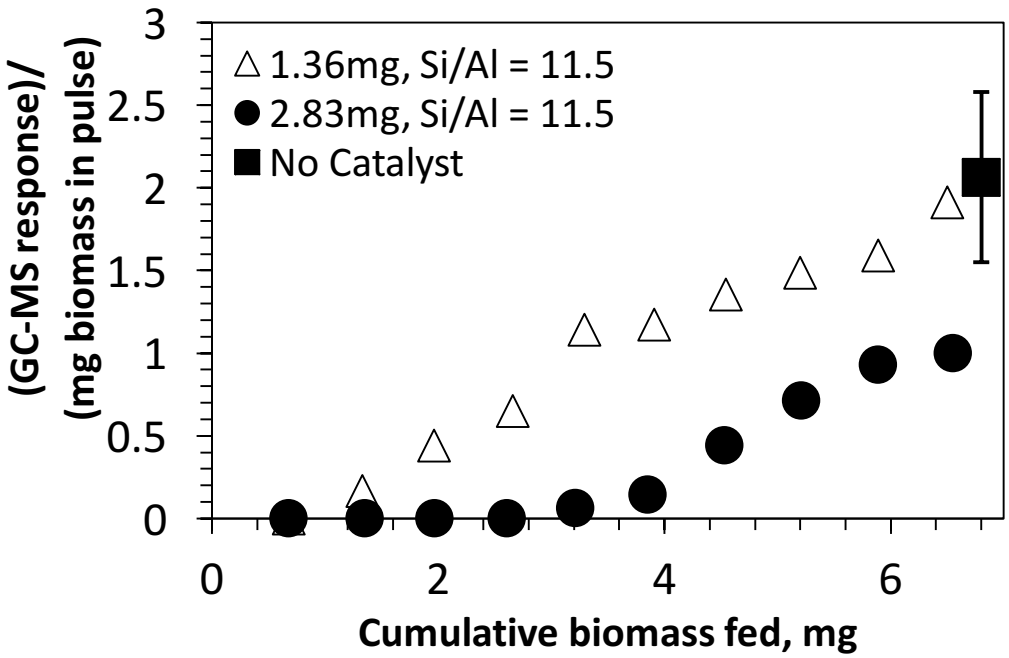


Figure 4.25. Acetic acid exiting the reactor over HZSM-5 of Si/Al=11.5 from oak pyrolysis vapors with fixed catalyst mass at 500°C.
Catalyst mass was varied from 1.36mg to 2.83mg.

A similar trend is observed upon analysis of the light gases for the case of Si/Al = 11.5 with 1.36 mg of catalyst. The noncondensable gases CH_4, CO, and CO_2 that result from conversion over the Si/Al = 11.5 zeolite as a function of biomass fed are shown in Figure 4.26. Interestingly, the CH_4 yield does not change considerably, which is in agreement with the results shown in 4.4 that show that CH_4 is not generated catalytically at this temperature. Both the CO and CO_2 yields decrease as the catalyst

begins to deactivate. The decreased light gas yield is in agreement with greater amounts of oxygenates observed exiting the reactor (less conversion) over the deactivated catalyst.

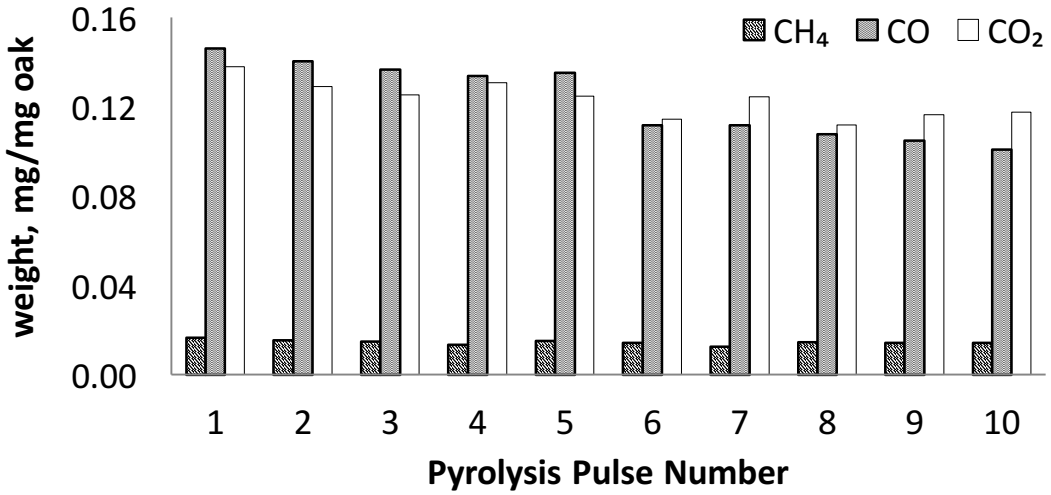


Figure 4.26. Gas yield from the 10 pulses of pyrolysis with HZSM5 of Si/Al=11.5 1.36mg under 500°C.

Analysis of the coke on the spent catalysts is shown in Table 4.1. For a constant catalyst mass, the most coke formation was observed with the medium acid density catalyst (Si/Al = 25). The lower acid density catalyst (Si/Al = 40) produced less coke, likely because of the lower amount of acid sites present compared with the Si/Al = 25 catalyst. The least coke formation occurred with the highest acid density catalyst (Si/Al = 11.5), which exhibited a very rapid deactivation rate in spite of its high initial reactivity. The close proximity of sites may have led to more coke initially, but this rapid coke formation could have resulted in pore plugging. The rapid deactivation of the active sites may reduce the rate of further coke formation. The same trend of decreased coke for the Si/Al = 11.5 catalyst was observed if comparable acid sites were used. If compared on an equivalent acid site basis, the total coke yield for the Si/Al = 40 catalyst was slightly higher than that of the Si/Al = 25 catalyst.

To further test the hypothesis of pore plugging for the low Si/Al ratio catalyst, experiments were conducted with both the Si/Al = 11.5 and Si/Al = 40 catalysts based on constant acid sites with increased catalyst volumes to enable characterization. If the amount of the Si/Al = 11.5 catalyst was increased to 5 mg, a micropore volume of 0.10 $cm^3\ g^{-1}$ is obtained for the fresh catalyst as shown in Table 4.2, which is in agreement with values obtained in the literature (*162*). After pulsing 46 mg of biomass over this catalyst, the larger catalyst bed volume exhibited a similar rapid deactivation based on a yield of alkylbenzenes of less than 15% of the initial yield obtained during the first pulses. A pore volume less than the detection limit of the instrument was obtained for the spent sample. We can claim accurately that the micropore volume is low, in line with the pore plugging hypothesis, and the surface area is below the detection limit of 100 $m^2\ g^{-1}$ for this sample size.

Catalyst Si/Al	Micropore Volume[a] [$cm^3\ g^{-1}$]	Brønsted site density[b] [$\mu mol\ g^{-1}$]
11.5 fresh	0.10	1093
11.5 spent	<0.02[c]	23
40 fresh	0.12	391
40 spent	0.09	123
[a] Measured using the t-plot method. [b] Measured by IPA-TPD. [c] Below the 100 m2 g-1 minimum detection limit of the instrument for this sample size.		

Table 4.2. Comparison of micropore volume and measured Brønsted sites on fresh and spent catalysts after exposure to 46 mg of biomass at 500°C.

Additional support for this conclusion of pore plugging over the Si/Al = 11.5 catalyst is found by the accessible acid sites post reaction. The acid sites were characterized by isopropylamine temperature-programmed desorption (IPA-TPD) to quantify the accessible Brønsted sites before and after reaction. The measured number of Brønsted sites post reaction for the Si/Al=11.5 catalyst decreased to 2% of the initial value (table

4.2; the corresponding propylene spectra are shown in figure 4.27). By contrast, for the Si/Al = 40 catalyst, the micropore volume decreases by less than 30% of the initial value, and the accessible acid sites decrease by 70% of the initial value. The greater relative reduction in accessible acid sites than the pore volume is an additional indication that the Si/Al = 40 catalyst is not de-activated solely by pore plugging under these conditions.

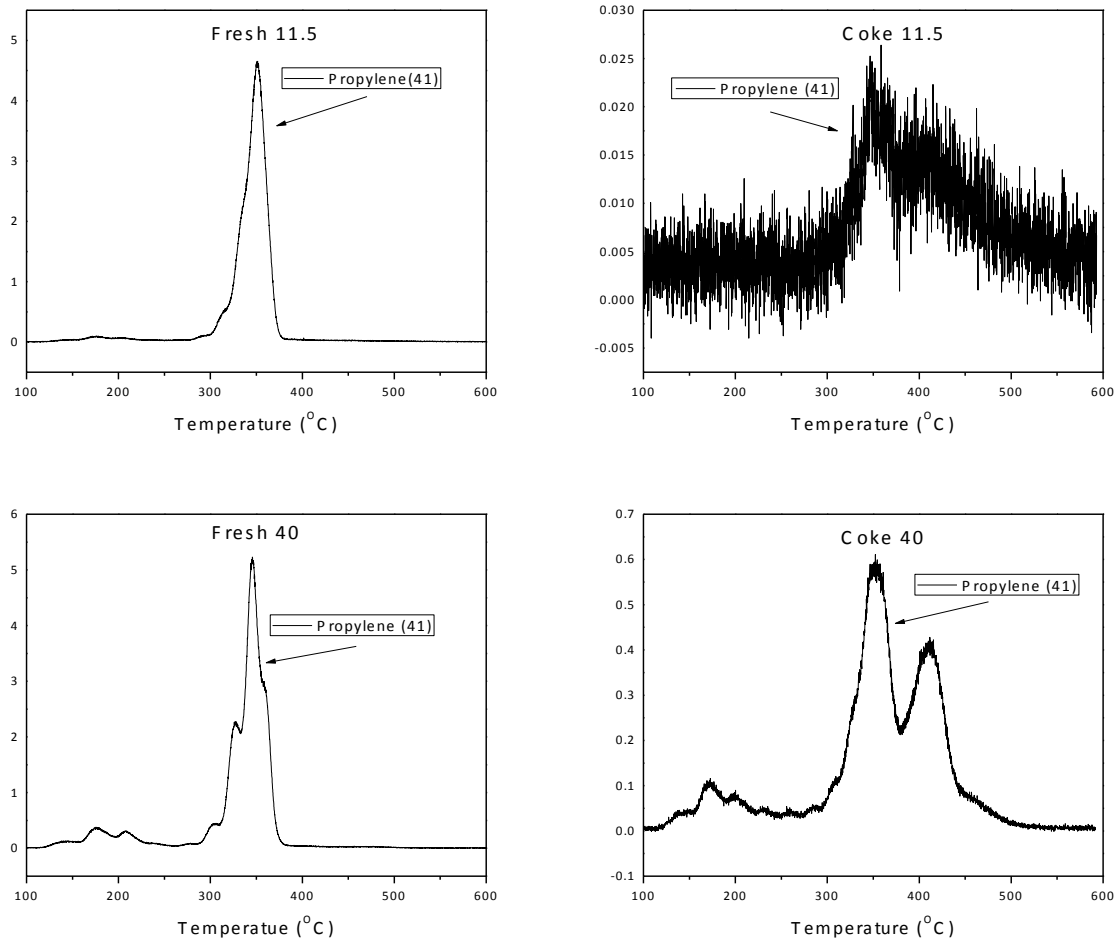


Figure 4.27. Gas yield from the 10 pulses of pyrolysis with HZSM5 of Si/Al=11.5 1.36mg under 500°C.

These results help to clarify others that occur in more complex systems in which several factors, such as rapid deactivation, excess catalyst, and the presence of an active

mesoporous support, make the identification of the role of the zeolite very challenging. As the catalyst with a high site density deactivates so rapidly, the net amount of catalytic coke may be lower and could easily lead to the wrong conclusions.

4.4.4 Product yields

The overall mass balance for the first pulse of six different catalytic and blank series is shown in Figure 4.28. As a result of the separation of the pyrolysis bed from the reactor bed, the char and ash can be quantified separately from the catalytic coke.

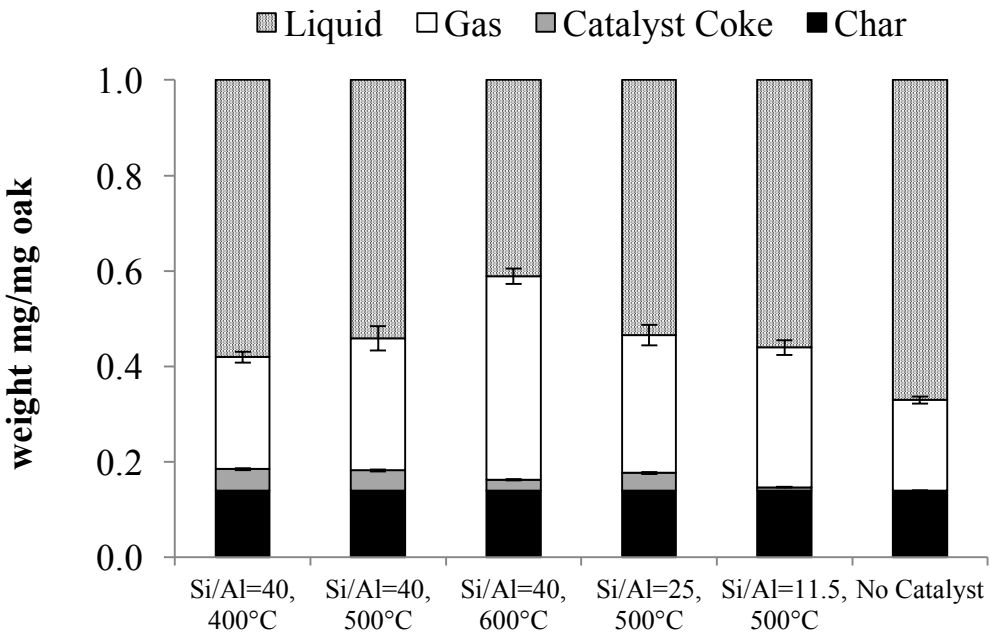


Figure 4.28. Overall mass balance from the first pulse of three different temperatures series and another two fixed acid sites series under 500˚C respectively.
The coke amount [mg $mg_{oak\ input}^{-1}$] in each catalytic case is averaged from the total coke in the spent catalyst after undertaking 10 pulses of pyrolysis vapors.

Quantification of the char, ash, coke, and noncondensable gases allows the estimation of an overall mass balance for the system. The coke yields reported are based on the amount of coke accumulated after 10 pulses of biomass injected (to enable a quantifiable amount measured by TGA as described in section 3.4.3). The cumulative

amount of coke after 10 pulses was divided by the amount of biomass injected to estimate the average amount of coke per pulse. The coke yield makes up a small portion of the biomass, so differences in coke yield that occur from the 1st to 10th pulse are not likely to have a great impact on the overall mass balance (Figure 4.28). The C_{2+} vapor and water yields reported in Figure 4.28 are taken by difference because of the small amounts of biomass vapors produced per pulse (*153*).

Interestingly, the coke values on the spent catalysts shown in Figure 4.28 are far less than coke values reported typically in in situ catalytic pyrolysis runs. This is because of the inability to distinguish catalytic coke from char during in situ catalytic pyrolysis experiments. The amount of char+coke reported here is in agreement with typical coke levels reported in the literature for in situ catalytic pyrolysis runs (*151*). The values of coke reported here that correspond to ~4.5% of the biomass fed are in agreement with other results reported for ex situ catalytic pyrolysis using a separate bed of HZSM-5 (*153*).

All the catalytic runs have a higher yield of CO, CO_2, and CH_4 and lower yield of C_{2+} vapors and water than the blank runs. As the temperature of the catalyst increases, the gas yield increases and the coke yield decreases as discussed previously. The aromatics yield passes through a maximum at 500°C (Figure 4.3), which highlights the significant difference in the composition of the vapor products as the temperature is increased. Variation of the acid site density results in nearly identical product yields for the first pulse if a comparable amount of acid sites are used. The lower coke yield over the Si/Al = 11.5 catalyst is because of the rapid deactivation of the catalyst, likely caused by pore plugging, and the fact that coke yields are measured after 10 pulses of biomass. These

results indicate that product yields that result from all three catalysts on a per acid site basis are initially very similar, but the major difference lies in the more rapid deactivation rates that result over the catalysts with acid sites in close proximity (Figure 4.16).

4.5 Conclusions

The use of a separate catalytic reactor in the custom pyroprobe allows independent evaluation of catalyst activity and catalyst deactivation. The conversion of biomass pyrolysis vapors to aromatics over HZSM-5 is favored at moderate temperatures around 500°C, with a tradeoff between low activity at lower temperature and the formation of excessive light gasses at higher temperatures. Initial yields of aromatics are comparable per acid site for the range of Si/Al tested here. It was found that the density of acid sites plays a critical role on catalyst stability under these conditions, and lower Si/Al ratios lead to increased catalyst deactivation rates.

4.6 Acknowledgements

The work has been financially supported by the NIFA-USDA BRDI (2012-10008-20271), the National Science Foundation (EPSCoR 0814361), and the US Department of Energy (DE-FG36GO88064). The assistance of Dr. Charles Mullen, of the Agricultural Research Service of the USDA, with the GC–MS analysis is gratefully acknowledged. We would also like to thank Nick Briggs for taking the SEM images and Zhenxi Liang for help with the char analysis.

Chapter 5: Staged Thermal Fractionation for Segregation of Lignin and Cellulose Pyrolysis Products: An Experimental Study of Residence Time and Temperature Effects

Extracted from a manuscript in preparation

Christopher Waters, Rajiv Janupala, Richard Mallinson, Lance Lobban

5.1 Abstract

Thermal conversion technologies may be the most efficient means of production of transportation fuels from lignocellulosic biomass. In order to increase the viability and improve the carbon emissions profile of pyrolysis biofuels, improvements must be made to the required catalytic upgrading to increase both hydrogen utilization efficiency and final liquid carbon yields. However, no current single catalytic valorization strategy can be optimized to convert the complex mixture of compounds produced upon fast pyrolysis of biomass. Staged thermal fractionation, which entails a series of sequentially increasing temperature steps to decompose biomass, has been proposed as a simple means to create vapor product streams of enhanced purity as compared to fast pyrolysis. In this work we use analytical pyrolysis to investigate the effects of time and temperature on a thermal step designed to segregate the lignin and cellulose pyrolysis products of a biomass which has been pre-torrefied to remove hemicellulose. At process conditions of 380°C and 180 second isothermal hold time, a stream containing less than 20% phenolics (carbon basis) was produced, and upon subsequent fast pyrolysis of the residual solid a stream of 81.5% levoglucosan (carbon basis) was produced. The thermal segregation comes at the expense of vapor product carbon yield, but the improvement in catalytic performance may offset these losses.

5.2 Introduction

Achieving a purely thermal separation of the cellulose- and lignin- derived components from biomass (e.g., levoglucosan and phenolic species) is of particular interest due to the inherent overlapping thermal stability regimes of these components and challenges of downstream separation as discussed in chapter 1. In this work, analytical pyrolysis is used to study the effects of time and temperature on an intermediate thermal treatment step designed to achieve separation of cellulose and lignin products from torrefied oak biomass. The objective of this study is to assess the efficacy of the staged thermal fractionation process in achieving enhanced purity of thermochemical vapor product streams. Time & temperature process parameters are varied and resulting products are characterized to better understand options for and potential of advanced catalytic valorization strategies.

5.3 Experimental

5.3.1 Biomass Feedstock

Red oak (*Q. Rubra*) sawdust was generated with a table saw from boards acquired at a local wood supplier. Sawdust was sieved to sizes between 250 and 425 μm and then dried in vacuum (0.02 MPa) at 60˚C overnight before use. Cell wall composition was measured for the oak samples using described in section 3.5.1. Elemental analysis was performed by Galbraith laboratories (Knoxville, TN). The composition is summarized in table 5.1 and 5.2.

Oak Component	*wt%*
Lignin	22.67%
Cellulose	27.70%
Hemicellulose	30.24%
Non-structural sugars	2.15%
Sum	*82.76%*

Table 5.1. Oak feedstock compositional data.

Oak Component	*wt%*
Carbon	48.22%
Hydrogen	6.03%
Oxygen	44.33%
Sum	*98.58%*

Table 5.2. Oak feedstock elemental analysis data.

5.3.2 Selection of initial conditions

With the goal of biopolymer thermal segregation in mind, biopolymer thermal stability regimes in the literature in tandem with a kinetic weight loss model developed specifically for oak biomass were employed to determine initial experimental conditions. As previously stated, hemicellulose decomposes at a lower temperature than cellulose while lignin decomposes over a broad range of temperatures. Cellulose has been shown to not undergo significant mass loss at temperatures below 275°C (*163*, *164*). As the hemicelluloses are much less thermally stable, a stage 1 temperature not exceeding 275°C should decompose hemicellulose while leaving the cellulose unconverted. Therefore, a temperature of 270°C was selected for stage 1. Additionally, temperatures between 500°C – 550°C has been shown to be an optimal temperature for biomass fast pyrolysis to optimize overall liquid yield (*106*, *165*). Therefore, stage 3 was carried out at 500°C.

The weight loss model described and parameterized for oak by Di Blasi (*166*) was used to choose the initial conditions for stage 2 and the hold time for stage 1, given the

conditions selected for the other two based on TGA studies in the literature. The source code for the implemented model can be found in Appendix A.

The model results for the conditions selected are presented in table 5.3 and figures 5.1 – 5.3. For the 270°C stage 1, 20 minutes was selected as the process time to achieve conversion of most of the hemicellulose. 350°C for 3 minutes was chosen as the initial stage 2 condition, as the kinetic model predicts near total conversion of the cellulose while leaving some lignin still available for conversion at stage 3. The prediction offered by this kinetic model suggests that the goal of separating the two polysaccharides via thermal degradation is achievable. The significant overlap of lignin is noteworthy, as the results of this work do not reflect high degrees of lignin degradation at the lowest temperature. The source code for this model can be found in Appendix A.

Temperature/Time	Predicted Cumulative Mass conversion	Predicted Cumulative Hemicellulose conversion	Predicted Cumulative Cellulose conversion	Predicted Cumulative Lignin conversion
270°C, 20 min	30.9%	87.3%	2.3%	17.0%
350°C, 3 min	61.1%	100.0%	99.0%	15.2%
500°C, 1 min	61.8%	100.0%	100.0%	17.0%

Table 5.3. Kinetic model weight loss predictions for oak biomass.

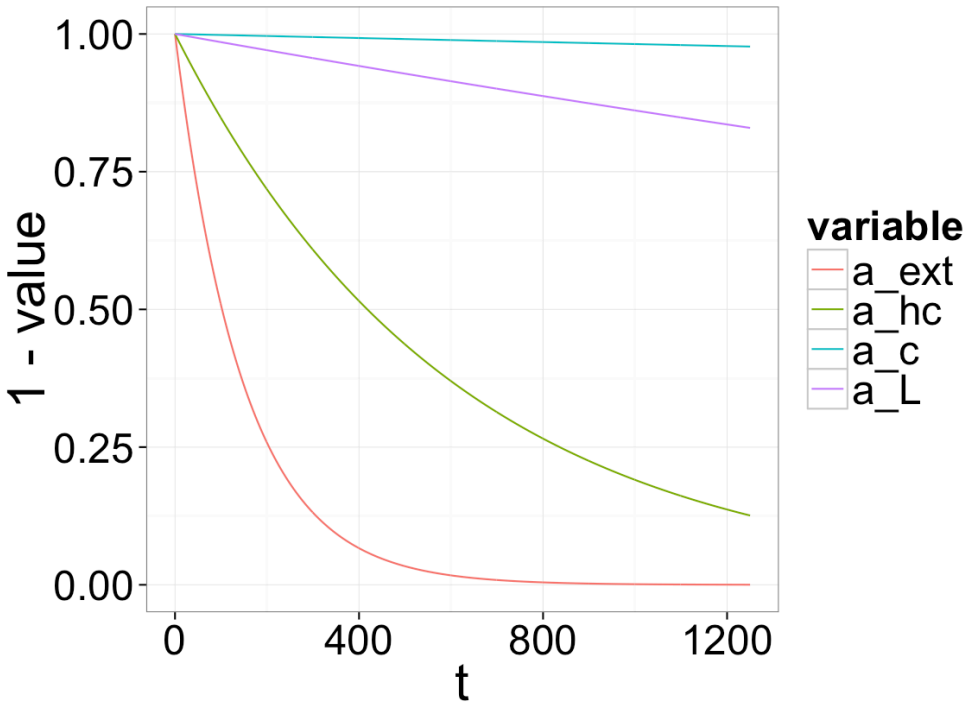


Figure 5.1. Model weight loss prediction, 270°C.
a_ext is the wt% of extractives volatilized, a_hc is the wt% of hemicellulose volatilized, a_c is the wt% of cellulose volatilized, and a_L is the wt% of lignin volatilized.

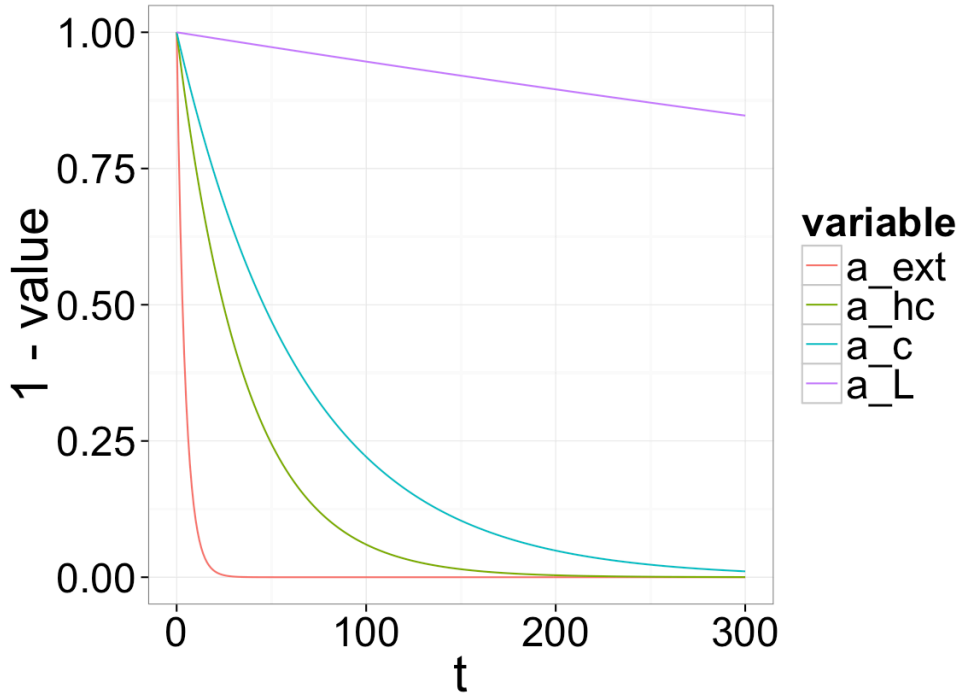


Figure 5.2. Model weight loss prediction, 350°C.

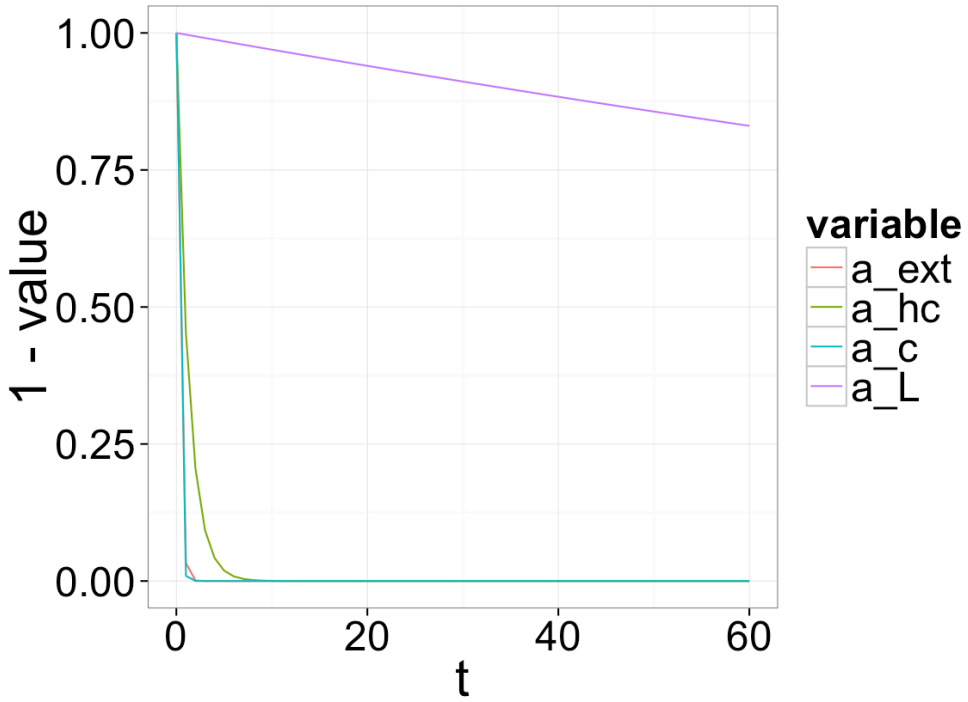


Figure 5.3. Model weight loss prediction, 500°C.

5.3.3 Conversion of stage 1 solid samples at stage 2 conditions

In preparation for the analytical pyrolysis investigation of the stage 2 conditions, sufficient quantities of stage 1 residual solids were generated from raw oak biomass using a bench scale system as described in section 3.1. Water content analysis of the resulting liquid (reported as liquid weight percent) was performed as described in section 3.3. Samples from the stage 1 solid residue were then prepared for analytical pyrolysis as described in section 3.2. Approximately 0.75 mg of stage 1 solid residue was used in each analytical pyrolysis experiment. Table 5.4 lists compounds and their lumps as used in this work.

Acetic acid	**Light Oxygenates**	**Furans**
Acetic acid	1-Acetyloxypropane-2-one	Furan
	1-Penten-3-ol	Furan, 2-methyl-
		Furan, 2,5-dimethyl-
Alkyl Phenols	1,2-Dihydroxyethane	Furan, 3-methyl-
Phenol	1,2-Dihydroxyethene	
Phenol, 3-methyl-	2-Butanone	
Phenol, 3-ethyl-	2-Butenal	**Sugars**
Phenol, 4-vinyl-	2-Hydroxy-3-oxobutanal	1,4:3,6-Dianhydro-α-D-glucopyranose
	2-Propanone, 1-(acetyloxy)-	
Furfurals	2-Propanone, 1-hydroxy-	1,5-Anhydro-arabinofuranose
(2H)-Furan-2-one	2-Propenal	
(2H)-Furan-3-one	2-Propenoic acid, methyl ester	ß-D-Glucopyranose, 1,6-anhydro-
(3H)-Furan-2-one	2,3-Butanedione	
(5H)-Furan-2-one	3-Buten-2-one, 3-methyl-	Levoglucosanone
1,3-Dioxol-2-one, 4,5-dimethyl-	3-Hydroxypropanal	
	Acetaldehyde	**Methoxyphenols**
2-Furanmethanol	Formic acid	2-Methoxy-4-vinylphenol
2-Furfuryl alcohol	Hydroxyacetaldehyde	Creosol
2,4(3H,5H)-Furandione, 3-methyl-	Propanal-2-one	Guaiacylacetone
	Propanal, 2-oxo-	Isoeugenol
2,5-Furandione, dihydro-3-methylene-	Propanoic acid, 2-oxo-, methyl ester	Methoxyeugenol
		Phenol, 2-methoxy-
2(3H)-Furanone, 5-methyl-	Unknown with Aldehyde fragment	Phenol, 2,6-dimethoxy-4-(2-propenyl)-
5-Hydroxymethylfurfural		
Dihydro-methyl-furanone		Syringol
2-Cyclopenten-1-one, 2-hydroxy-3-methyl-	**Levoglucosan**	Vanillin
	Levoglucosan	

Table 5.4. Identified compounds and lumps in this study.

5.4 Results & Discussion

5.4.1 First temperature step products characterization

Figure 5.4 shows the composition of the liquid collected from stage 1. The water content was determined by Karl-Fischer titration to be 64.9 weight percent (dry basis). A mass balance from this reaction is shown in table 5.5. As stated above, the stage 1 recovered solid was used as the starting material for the pyroprobe screenings to produce the intermediate and final temperature step results that follow. Levoglucosan is the primary product of cellulose pyrolysis (*31*, *98*); the low abundance in the liquid

indicates that the cellulose was mostly unconverted by this stage 1 treatment, in agreement with the results predicted by the kinetic model. The relatively low yields of phenolic species and levoglucosan confirm that at these mild conditions, hemicellulose breakdown is the primary source of the volatile products formed.

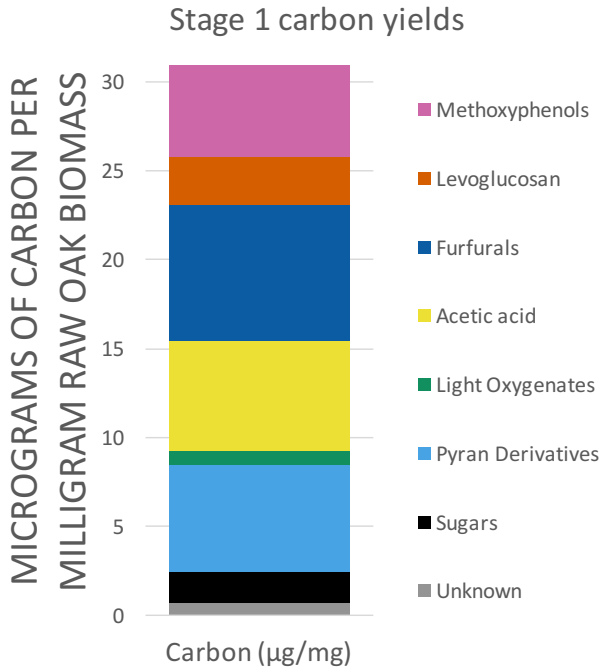


Figure 5.4. Stage 1 collected liquid composition.

Total oak sawdust fed	15.8 gm	100.0%
Total liquid collected	3.3 gm	20.9%
Total solid collected	11.9 gm	75.3%
Non-condensable gases*	0.6 gm	3.8%

Table 5.5. Mass balance of Stage 1 Thermal Treatment.

5.4.2 Total Yields

The measured carbon content in the organic vapor product for the sequential stages 2 and 3 in the pyroprobe were summed and are presented in table 5.6 and figure 5.5 with the stage 1 carbon yield as functions of the stage 2 independent process condition variables (i.e., time & temperature).

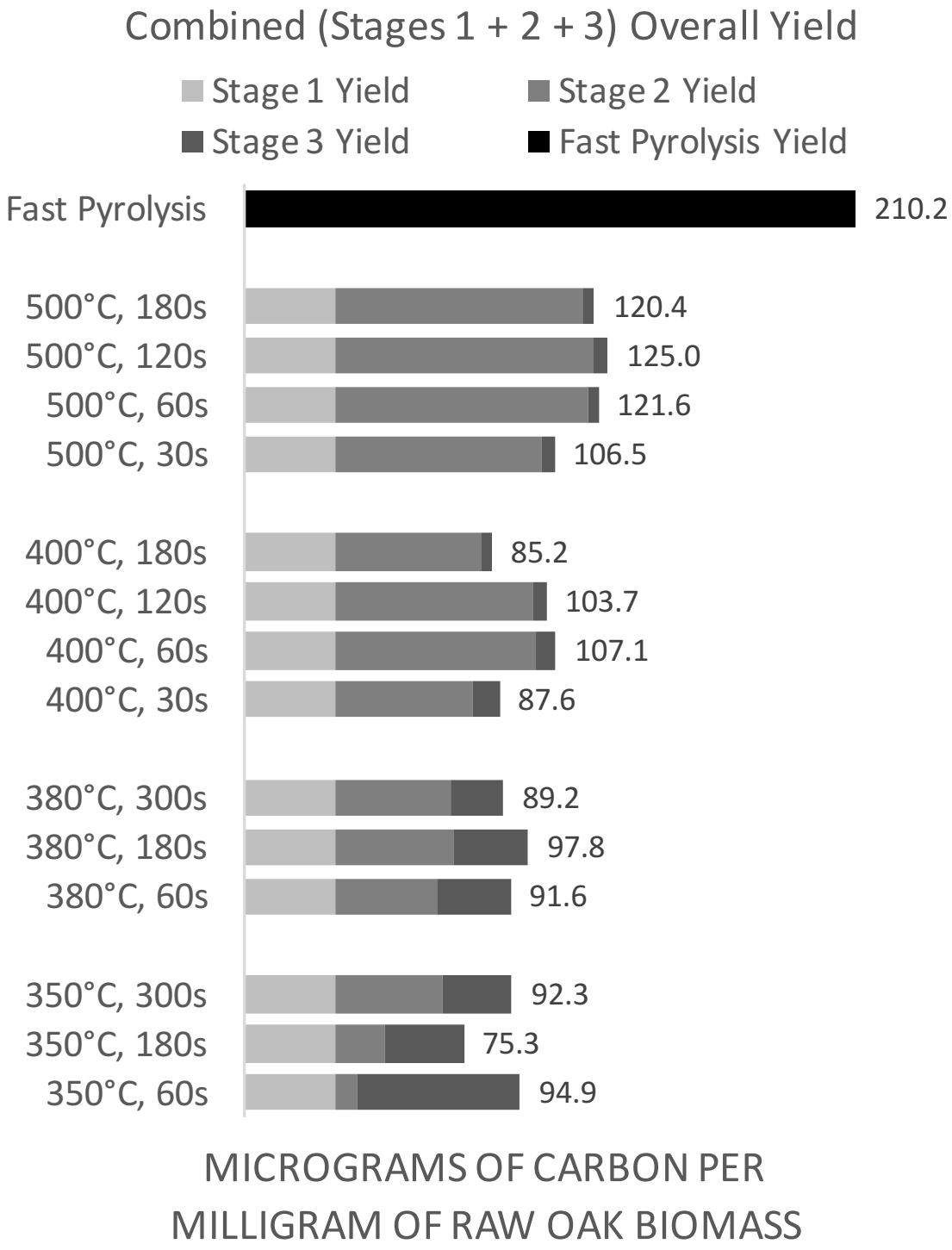


Figure 5.5. Cumulative stages 1, 2 & 3 carbon yields compared with single-step fast pyrolysis carbon yields. *(Does not include CO/CO_2).*

Temperature	Time	Stage 1 Carbon Yield	Stage 2 Carbon Yield	Stage 3 Carbon Yield	Combined Carbon Yield	% of fast pyrolysis
350°C	60s	30.9	7.59	56.45	94.94	45.2%
	180s	30.9	17.28	27.15	† 75.33	35.8%
	300s	30.9	37.22	24.15	92.27	43.9%
380°C	60s	30.9	34.97	25.70	91.56	43.6%
	180s	30.9	41.00	25.87	97.77	46.5%
	300s	30.9	40.25	18.01	89.16	42.4%
400°C	30s	30.9	47.36	9.38	87.64	41.7%
	60s	30.9	69.69	6.54	107.13	51.0%
	120s	30.9	68.62	4.20	103.72	49.3%
	180s	30.9	50.61	3.65	85.16	40.5%
500°C	30s	30.9	71.29	4.32	106.51	50.7%
	60s	30.9	87.02	3.73	121.65	57.9%
	120s	30.9	89.02	5.07	*124.99	59.4%
	180s	30.9	85.86	3.67	120.43	57.3%
Single-Step Fast Pyrolysis					210.24	100.0%

Table 5.6. Carbon yields for stages 1, 2 & 3.
All carbon yield values are reported as (measured μg carbon) / (mg raw oak biomass).
† denotes the lowest yield multi-stage case; * denotes the highest yield multi-stage case.

These results indicate that the more complete conversion of the solid with progressively harsher stage 2 conditions limits the amount of material available for subsequent volatilization in stage 3. As expected, the stage 2 carbon yields generally increase with increasing temperatures, and the stage 3 carbon yields drop with increasing stage 2 carbon yields (as seen in table 5.6 & figure 5.5). At 350°C, the the stage 2 carbon yields increase with increasing stage 2 treatment time, and the stage 3 yields decrease with increasing stage 2 treatment time; for stage 2 treatment times of 60 seconds, the stage 2 carbon yields increase with increasing temperatures, and the stage 3 yields decrease with increasing stage 2 temperatures. However, at the intermediate temperatures (380°C and 400°C), decreasing carbon yields are observed with increased treatment time. These decreasing total carbon yields are observed across the whole product spectrum and thus are not indicative of dramatically changing product selectivities (figure 5.11), and the MS chromatograms from these experiments do not show any pronounced increases in CO_2.

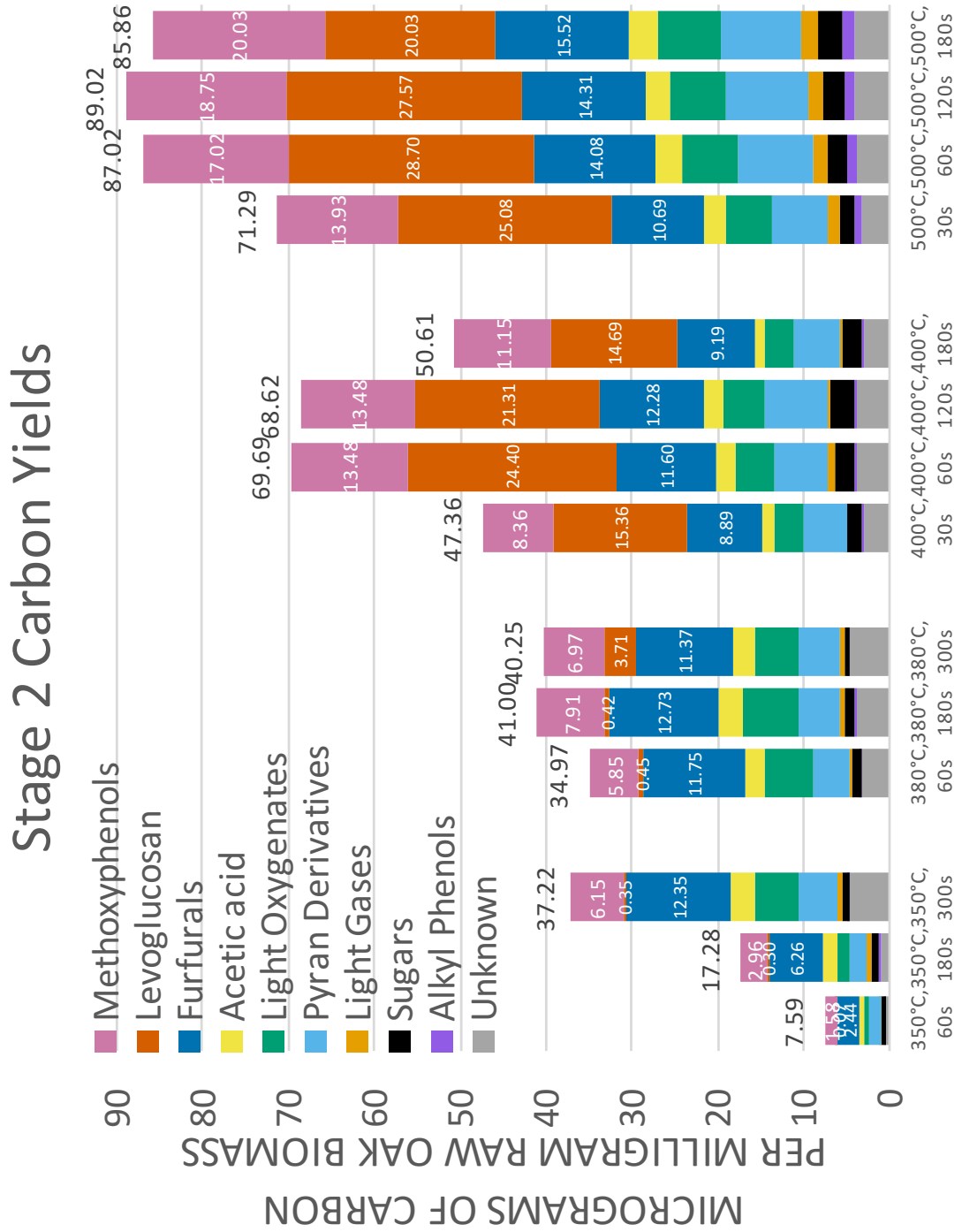


Figure 5.6. Stage 2 yields, separated by compound lumps, for the stage 2 conditions labeled.

The number above the stacked bars is the sum of the colored bars, indicating the total carbon yield of each stage 2 condition. Numbers in the pink, orange, and blue segments (corresponding with methoxyphenols, levoglucosan, and furfurals respectively) are the yields of those specific lumps. Yield values are (measured µg carbon in vapors) / (mg of raw biomass).

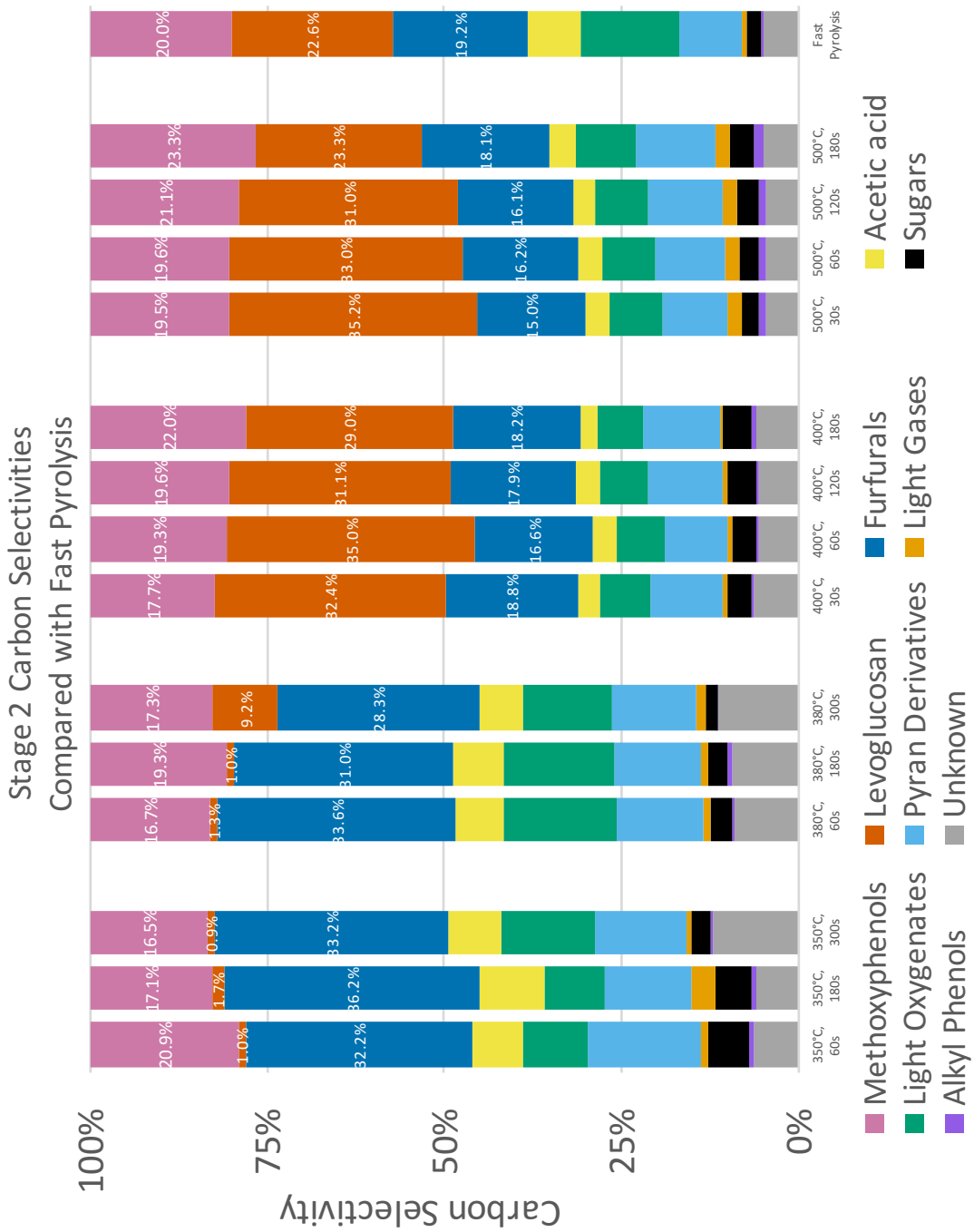


Figure 5.7. Stage 2 carbon selectivities, separated by compound lumps, for the stage 2 conditions labeled, compared with single-step fast pyrolysis of raw oak biomass (far right).
Numbers in the pink, orange, and blue segments (corresponding with methoxyphenols, levoglucosan, and furfurals respectively) are the carbon selectivities of those specific lumps.

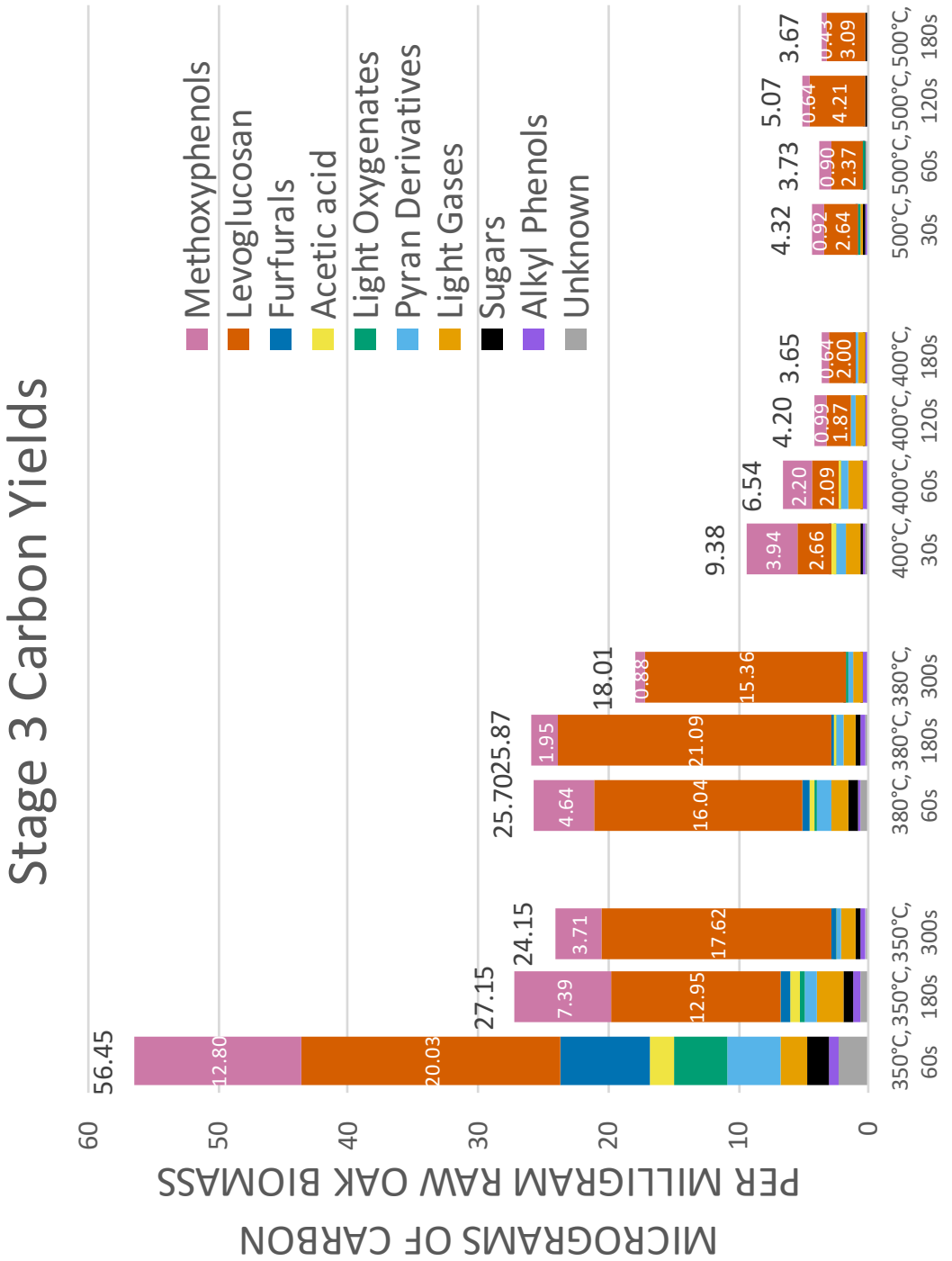


Figure 5.8. Stage 3 carbon yields, separated by compound lumps, for the 2nd stage conditions labeled.
The number above the stacked bars is the sum of the colored bars, indicative of the total carbon yield of the 3rd stage thermal treatment. Numbers in the pink & orange segments (corresponding with methoxyphenols and levoglucosan respectively) are the carbon yields of those specific lumps. Yield values are (measured µg carbon in vapors) / (mg of raw biomass).

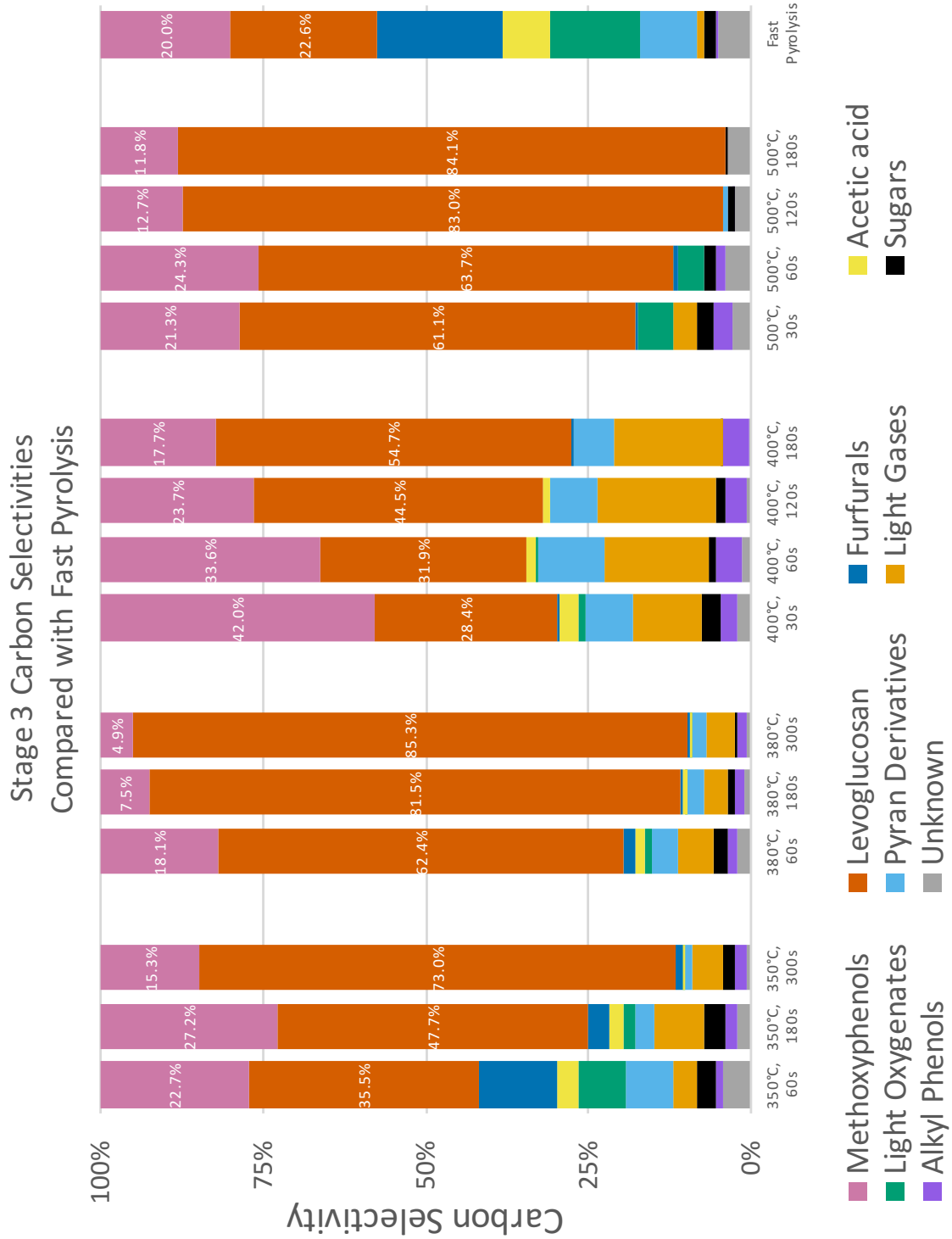


Figure 5.9. Stage 3 carbon selectivities, separated by compound lumps, for the stage 2 conditions labeled, compared with single-step fast pyrolysis of raw oak biomass (far right).
Numbers in the pink & orange segments (corresponding with methoxyphenols and levoglucosan respectively) are the selectivities of those specific lumps.

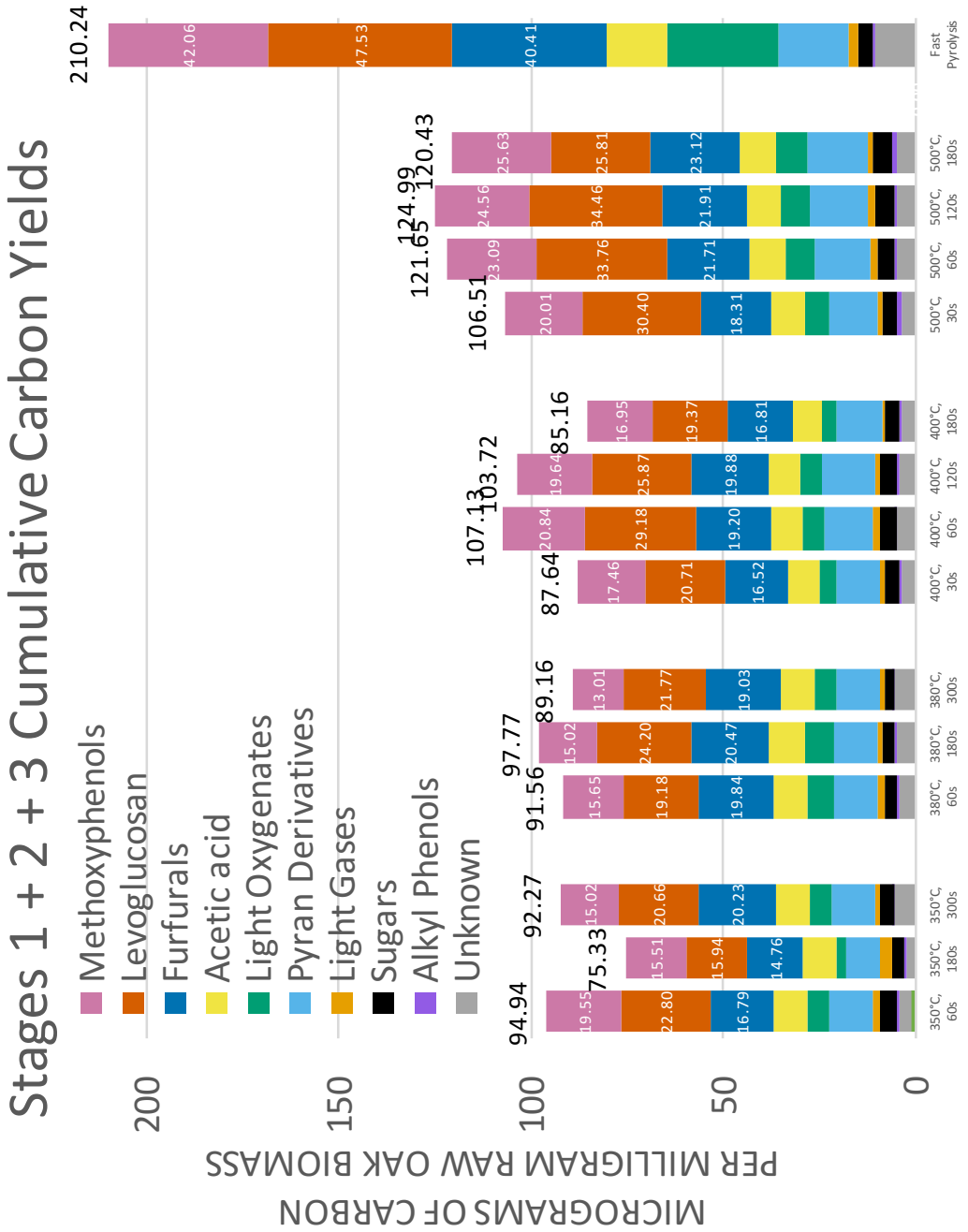


Figure 5.10. Stages 1, 2 & 3 cumulative carbon yields, separated by compound lumps, for the stage 2 conditions labeled, compared with single-step fast pyrolysis of raw oak biomass (far right).
The number above the stacked bars is the cumulative carbon yield of the three stages. Numbers in the pink, orange, and blue segments (corresponding with methoxyphenols, levoglucosan, and furfurals respectively) are the carbon yields of those specific lumps. Much of the decrease in the overall carbon yield at the lower temperatures is due to the decrease in the yield of these compound lumps. Yield values are (measured µg carbon in vapors) / (mg of raw biomass).

Figure 5.11. Stages 1, 2 and 3 cumulative carbon selectivities, separated by compound lumps, for the stage 2 conditions labeled, compared with single-step fast pyrolysis of raw oak biomass (far right).
The number above the stacked bars is the total normalized peak area, indicative of the cumulative carbon yield of the three stages. Numbers in the pink, orange, and blue segments (corresponding with methoxyphenols, levoglucosan, and furfurals respectively) are the carbon yields of those specific lumps.

The cumulative staged thermal fractionation acetic acid yields compared with single-step fast pyrolysis yields (figure 5.12) indicate that at stage 1 conditions the hemicellulose in the biomass is nearly fully decomposed. The maximum acetic acid yield from stages 2 and 3 (cumulative) as a percentage of single-step fast pyrolysis acetic acid yield is 27.6%, occurring at stage 2 conditions of 500°C for 180s. As cellulose pyrolysis is not a major producer of acetic acid (*31*), any acetic acid formed in stages 2 and 3 must come from decomposition of any remaining unreacted hemicelluloses and from lignin decomposition (*12*). It is possible that the acetic acid seen in stage 2 is the result of incomplete stage 1 decomposition of hemicellulose. However, if that were the case, the successively increasing yields with time in the 350°C cases would not be expected - the yields at 180s and 300s would not be expected to be dramatically different than the yield at 60s. These results suggest that the acetic acid may be a result of reactions with lower activation energies than the reactions that produce larger, lignin monomer unit-like products such as 4-substituted syringols and guaiacols.

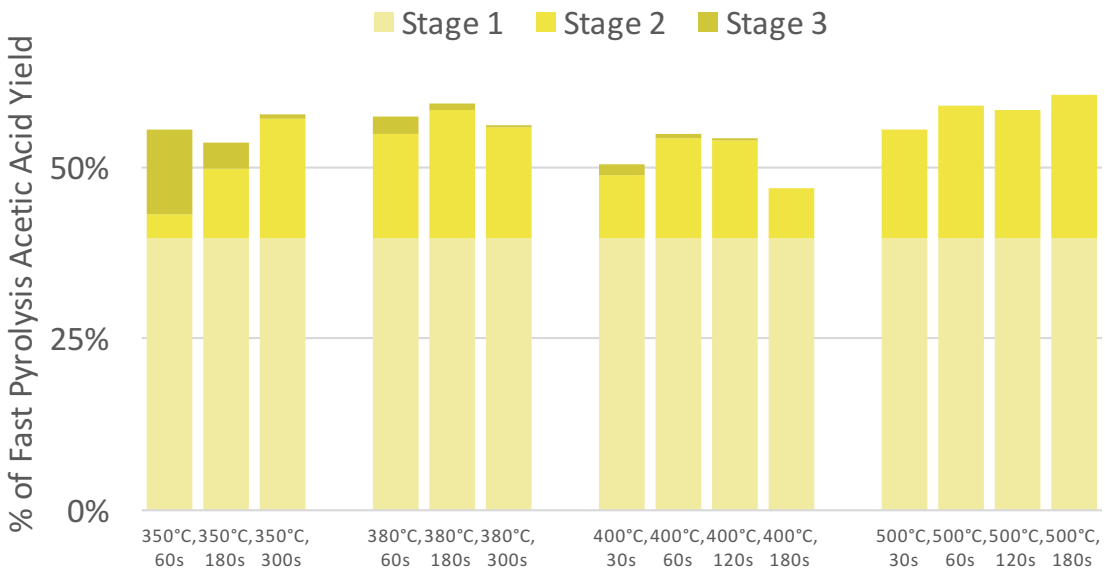


Figure 5.12. Acetic acid yields for stages 1, 2 & 3, as a percentage of the fast pyrolysis acetic acid yield.
The low yields of acetic acid indicate that the hemicellulose is decomposed at stage 1 conditions.

Much of the difference in the stages 2 & 3 total carbon yields is due to the combined behavior of the levoglucosan, methoxyphenols, and furfural lumps, as shown in figures 5.6, 5.8, and 5.10. It is well established that phenolic pyrolysis products originate from lignin, while furfurals and levoglucosan result from polysaccharide decomposition (levoglucosan exclusively from cellulose) (*12*, *30*, *31*). Since all or almost all of the hemicellulose has decomposed before screening at these conditions, it is expected that the furfurals measured here are cellulose decomposition products as well. Beyond the differences due to hemicellulose removal discussed above, these observed lower total process carbon yields following intermediate thermal treatments (especially in the methoxphenol, levoglucosan, and pyran lumps) are ascribed to changes in the thermal stability of the biomass caused by (primarily) condensation and repolymerization reactions that may not be kinetically favored in fast pyrolysis. Early work in the field

established that organic yields from cellulose in general are maximal at 500°C (*98*, *167*). More recent NMR investigations (*168*, *169*) have demonstrated that new aliphatic and aromatic structures are formed in polysaccharides treated at torrefaction temperatures. The mechanism for cellulose pyrolysis advanced by Huber (*170*) suggests that cellulose begins to depolymerize at temperatures as low as 100°C, with breaks in the polymer chain occurring until the anhydrosugar monomer (levoglucosan) is obtained. At temperatures low enough to prevent volatilization of levoglucosan, char formation (proposed to be the result of repolymerization of levoglucosan or rearrangement/decomposition products of levoglucosan) may be enhanced. Similar condensation and repolymerization behavior also has been shown to occur in lignin; Wen et. al (*171*) observed that β-O-4, β-β, and β-5 lignin bonds disappear under torrefaction (275°C – 300°C) conditions. The β-O-4 linkages (the most abundant type) cleave and reform aromatic C-C bonds within the lignin. These new refractory bonds have much higher dissociation enthalpies, and may not dissociate at typical fast pyrolysis temperatures, leading to enhanced char formation and corresponding loss of organic vapor yield.

While the pyroprobe/GCMS-FID system is a useful analytical tool that provides rapid assessments of pyrolysis vapor products, it is also limited in that it cannot provide good mass balance information that a larger-scale system can (such as the one used to generate the stage 1 products). Successive pyroprobe treatments on the same sample leaves that sample in the (unheated) pyrolysis chamber during the chromatography run, preventing measurement of the residual solid before the next pyrolysis step. Furthermore, other studies have noted that the pyroprobe apparatus appears to generate

much more char than other systems such as TGA, ascribed to the lack of sweep gas through the pyrolyzing solid due to the required tube loading configuration (*170*). This can lead to high concentrations of volatiles in the heated zone and increased coke formation. Additionally, the pyroprobe system only provides vapor product compositional data, and cannot be used to generate liquids for analysis of properties like water content, viscosity, elemental analysis, etc. – information which is necessary for techno-economic analyses and life cycle analyses. Further, larger-scale studies are needed to generate these kinds of data for the optimized conditions discussed here.

5.4.3 Segregation Ability

The ability of the staged thermal fractionation process to achieve enhanced purity of thermochemical vapor product streams that will facilitate improved catalytic processing is of specific interest and will be the focus of the remainder of the discussion. First, a case of relatively poor thermal separation is examined. Figure 5.13 shows the carbon yield & selectivity of stage 2 pyrans plotted against stage 2 furfurals. These two compound lumps show relatively poor thermal segregation ability; as seen in figure 5.13a, there is no condition that favors production of one of these lumps over the other. While the selectivity to furfurals is higher at 350°C and 380°C, this is mostly due to the production of levoglucosan at 400°C and 500°C. Both furfurals and pyrans are known pyrolysis products of cellulose, and cannot be readily separated using staged thermal fractionation.

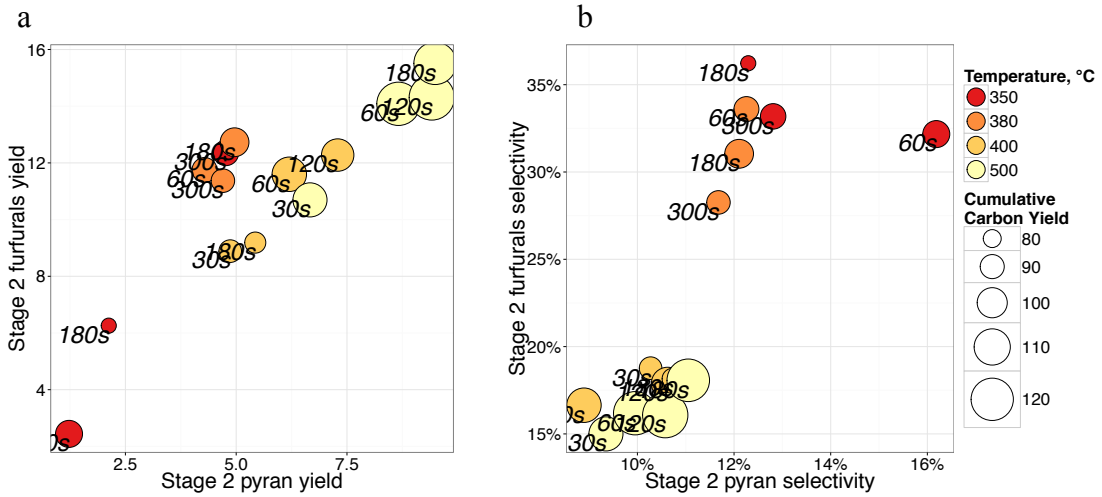


Figure 5.13. Stage 2 pyran vs. Stage 2 furfurals carbon yield (a) and selectivity (b) for the stage 2 conditions indicated.
Cumulative carbon yield is the sum of the carbon in all compound lumps at both stage 2 & stage 3 conditions. Yield values are (measured µg carbon in vapors) / (mg of raw biomass).

A separation of particular interest for this work was of the levoglucosan and the combined phenolic lumps (alkyl phenols + methoxyphenols) between stages 2 and 3. Figures 5.14 & 5.15 show the carbon yields and carbon selectivity of these two lumps plotted against each other for stages 2 and 3. As discussed in the introduction, the specific catalytic upgrading approaches for these kinds of compounds is different and non-complementary, and avoiding downstream separations methods is desirable. Levoglucosan may be better converted selectively to gluconic acid, while phenolics should be transalkylated and ultimately hydrotreated. From these two figures, 380°C, 180s is likely the best stage 2 process condition to achieve segregation of these two lumps into separate product streams while minimizing total yield loss. The stage 2 products at these conditions contains almost no levoglucosan, and the stage 3 products have very high levoglucosan carbon selectivity (81.5%) with little phenolic content.

However, this separation comes at the expense of the cumulative carbon yield as discussed in the section above.

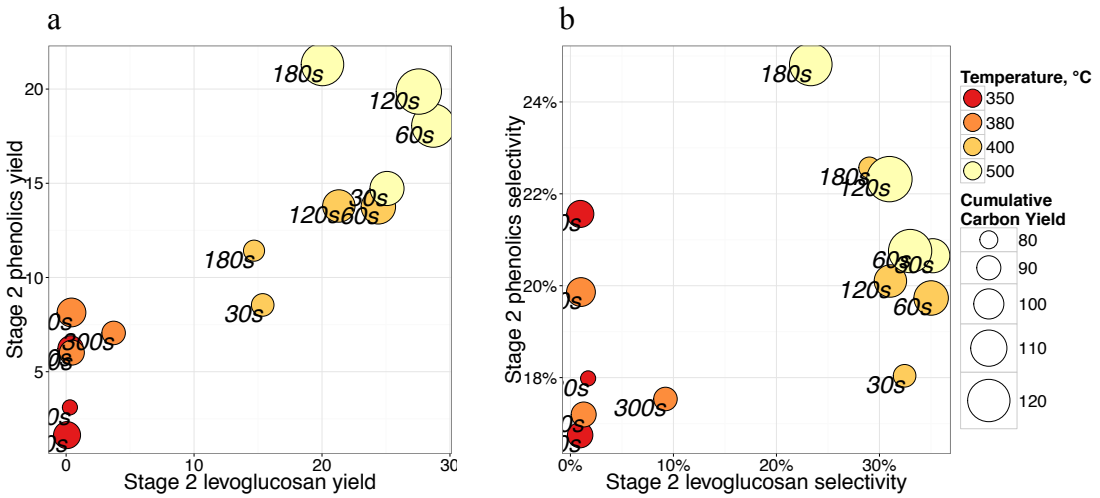


Figure 5.14. Stage 2 Levoglucosan vs. Stage 2 Phenolics carbon yield (a) & carbon selectivity (b) for the stage 2 conditions indicated.
Cumulative carbon yield is the sum of the carbon in all compound lumps at both stage 2 & stage 3 conditions. Yield values are (measured µg carbon in vapors) / (mg of raw biomass).

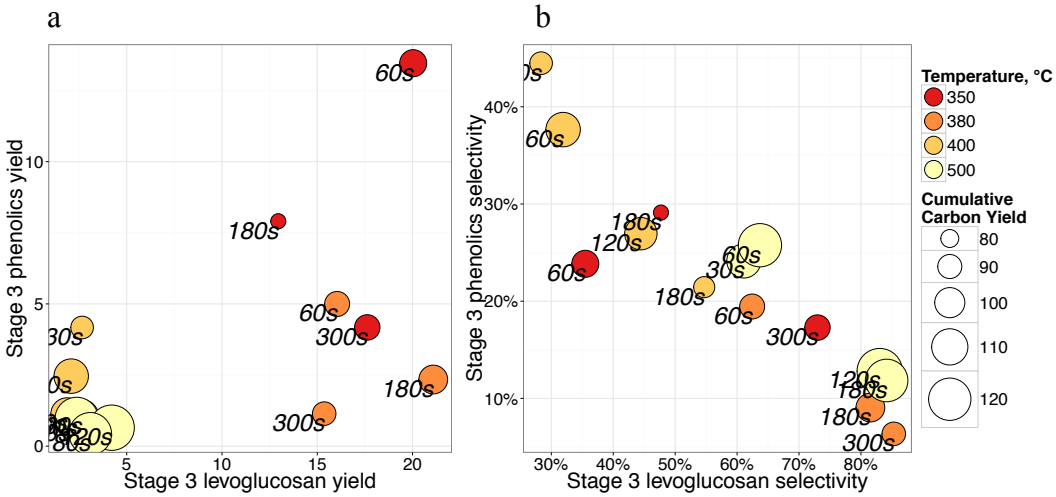


Figure 5.15. Stage 3 Levoglucosan vs. Stage 3 Phenolics carbon yield (a) & carbon selectivity (b) for the stage 2 conditions indicated.
Cumulative carbon yield is the sum of the carbon in all compound lumps at both stage 2 & stage 3 conditions. Yield values are (measured µg carbon in vapors) / (mg of raw biomass).

Figures 5.16 – 5.18 shows several other examples of compound lump segregations that could be possible using staged thermal fractionation. From figures 5.16 and 5.17, a

moderate segregation between the light oxygenates and phenolics into stages 2 & 3, respectively, is possible at 380°C. Nearly all of the light oxygenates are segregated into stage 2; very little are produced in stage 3 across nearly all conditions observed (Figure 5.17b). The light oxygenates lump consists of small molecule fragments and is produced by all major biomass components; the ability to segregate these from phenolics at lower temperatures likely is due to the minimal activation and decomposition of cellulose. In this case, most of the light oxygenates are expected to be formed primarily from lignin decomposition.

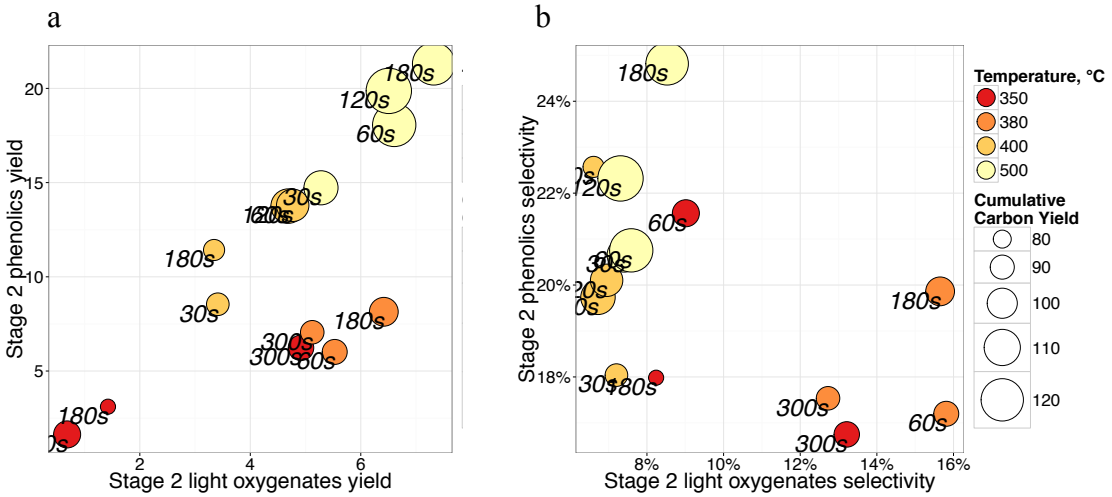


Figure 5.16. Stage 2 light oxygenates vs. Stage 2 phenolics carbon yield (a) and carbon selectivity (b) for the stage 2 conditions indicated.
Cumulative carbon yield is the sum of all compound lumps at both stage 2 & stage 3 conditions. Yield values are (measured µg carbon in vapors) / (mg of raw biomass).

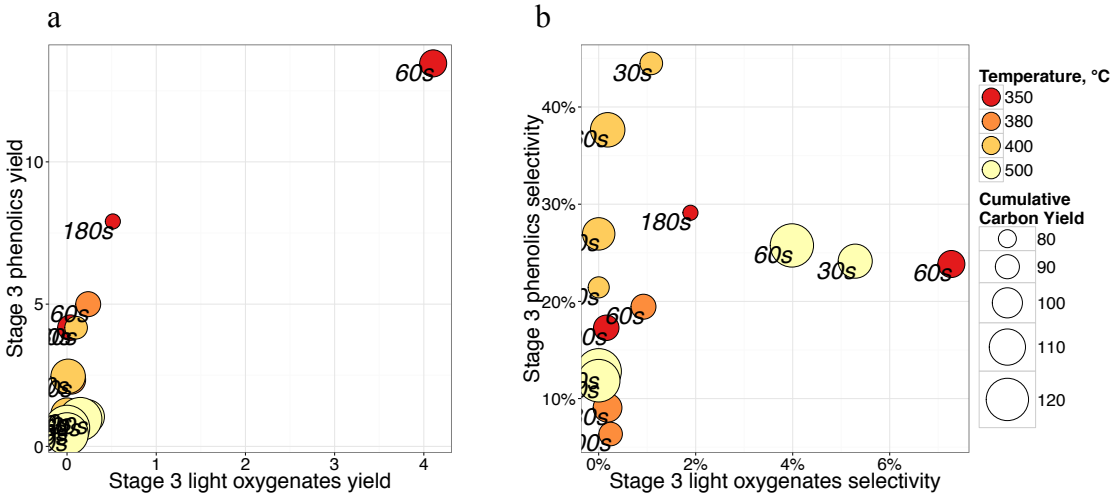


Figure 5.17. Stage 3 light oxygenates vs. Stage 3 phenolics carbon yield (a) and carbon selectivity (b) for the stage 2 conditions indicated.
Cumulative carbon yield is the sum of all compound lumps at both stage 2 & stage 3 conditions. Yield values are (measured µg carbon in vapors) / (mg of raw biomass).

The acetic acid and the phenolics also appear to be thermally segregable at lower temperature stage 2 conditions. Stage 2 conditions of 380°C or 350°C for 300s achieve good segregation of acetic acid into stage 2 while minimizing stage 2 phenolics yields, as seen in figure 5.18. As discussed above, acetic acid is not a major product of cellulose pyrolysis, and as such is expected to be derived primarily from unreacted hemicelluloses and from lignin decomposition (as are the phenolics) (*12*). Since the hemicellulose has been removed, the higher temperature cases (400°C-500°C) show little difference in the acetic acid as compared to the lower temperature cases, but the phenolics yield is greater due to enhanced lignin decomposition at higher temperatures.

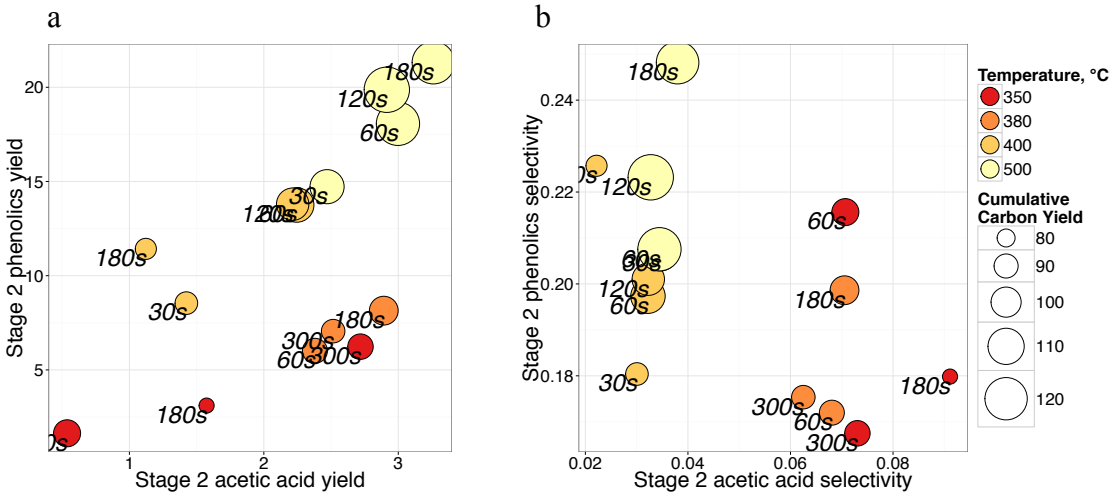


Figure 5.18. Stage 2 acetic acid yield vs. Stage 2 phenolics carbon yield (a) and carbon selectivity (b) for the stage 2 conditions indicated.
Cumulative carbon yield is the sum of all compound lumps at both stage 2 & stage 3 conditions. Yield values are (measured µg carbon in vapors) / (mg of raw biomass).

5.5 Conclusions

Rapid screening in the pyroprobe has demonstrated that the production of thermochemical vapor product streams of enhanced purity and decreased complexity as compared to fast pyrolysis is achievable via staged thermal fractionation by manipulating the exposure time and temperature at which the conversion is conducted.

The ability of this technique to segregate the compound lumps is primarily derived from the thermal stability of the biopolymers from which the lumps are derived, with the exception of lignin, which may exhibit selectivity changes over a temperature range not observed with polysaccharides. However, some of the more effective thermal segregation conditions (namely, 350°C and 380°C) carry with them a penalty to the cumulative carbon yield of organics from the process due to condensation and repolymerization reactions that form refractory C-C bonds which do not exist in untreated biomass.

The stage 2 conditions of 380°C and 180 seconds seem to be best of those tested for creating both stage 2 and 3 product streams of enhanced purity as compared to fast pyrolysis. At these conditions, less than 10% of the carbon in stage 2 comes from levoglucosan, while over 85% of the carbon in the stage 3 product stream is in the levoglucosan. Only 6.3% of the carbon in stage 3 is in the phenolic lumps, and there is also minimal light gas production (4%). For these process parameters, an appropriate catalytic strategy for the best stage 2 product stream might involve a simple condensation separation for the heavy phenolics followed by a sequential ketonization/aldol condensation, ending with transalkylation of the phenolics and subsequent hydrotreating (*22*). Stage 3 could be treated as pure levoglucosan, with the aforementioned gluconic acid pathway. Figure 5.19 is an updated version of figure 1.1, showing a revision to the hypothetical process schematic based on the results of this work.

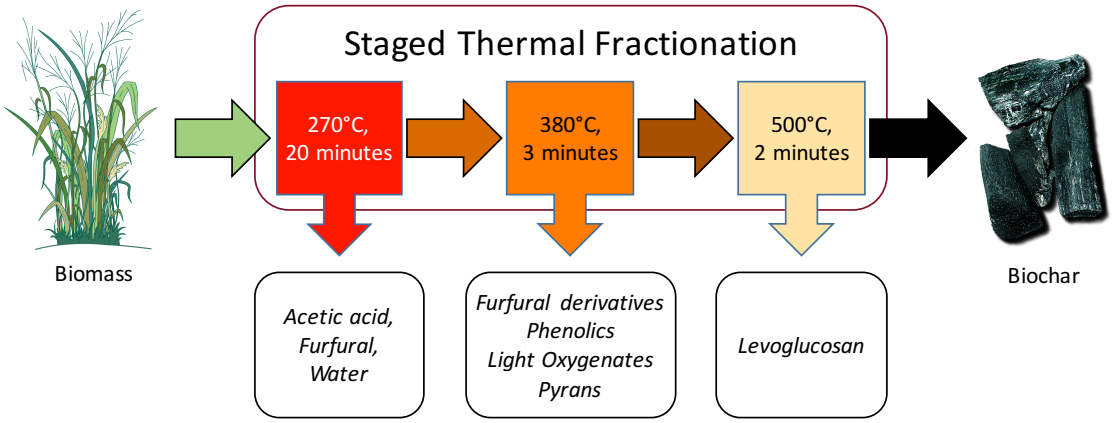


Figure 5.19. Revised configuration for thermal fractionation of lignocellulosic biomass with resultant purified product streams based on this work.

Although staged thermal fractionation at stage 2 conditions of 380°C and 180 seconds does indeed yield product streams of greatly enhanced compositional purity as compared to fast pyrolysis, the separation achieved does come at the expense of total carbon yield. Full investigation and understanding of the impacts of this tradeoff of yield for better selectivity on process economics are beyond the scope of this work, but with this effort we hope to guide inquiry into the acceptable limits of catalytic performance and the associated process costs. From an emissions perspective, however, this is not necessarily an undesirable outcome. Production of solid char from biomass has been shown to be an effective form of carbon sequestration. If the increased formation of char due to the intermediate thermal treatment allows for lower catalyst deactivation (and subsequent CO_2 formation upon regeneration), the carbon cycle efficiencies of staged thermal fractionation as compared to fast pyrolysis could be a net positive.

To hedge against the possibility that the carbon yield loss may be too severe of a penalty, it is necessary to consider a two-step fractionation strategy, of which the 500°C stage 2 conditions are representative. The 500°C stage 2 conditions represent the

highest overall process carbon yields of the stage 2 conditions tested. In this scheme, the stage 1 treatment achieves segregation of the acetic acid and furfural with minimal production of lignin-derived phenolics; however, the stage 1 treatment is mild enough to limit the degree to which the lignin & cellulose condenses/repolymerizes, providing comparable carbon yields in the resulting vapor product to single-step fast pyrolysis. The segregation of the acetic acid into stage 1 is probably the most notable product stream purity enhancement, as the acid-catalyzed reactivity of the high-temperature liquid product would be expected to be greatly reduced. Further investigation of the storage and handling stability, separation strategies, and catalytic upgrading performance parameters for a liquid having this composition are underway.

The differences between the empirical kinetic model used to determine initial experimental conditions and the observed vapor products across the stages are not trivial. The kinetic model, developed using isothermal techniques, was far off the mark when compared to the observed stage 2 temperatures at which production of levoglucosan occurred. Almost no levoglucosan was produced at temperatures below 400°C in these experiments, with the exception being the 380°C, 300s case. The most likely explanation is offered by Lin et al (*170*), where they observe a significant thermal lag effect in temperature-ramp thermogravimetric analysis of cellulose pyrolysis. At slow (1K min^{-1}) heating rates, rapid volatilization of cellulose is observed to begin at roughly 275°C; at fast (150K min^{-1}) heating rates, cellulose volatilization does not begin to occur until nearly 375°C. The kinetic model used to select initial conditions in this work was developed using a 5K min^{-1} temperature ramp, much slower than the 1000K s^{-1} heating rate in the pyroprobe apparatus used. This thermal lag phenomenon would

account for the discrepancy between the model and our results, including the slight levoglucosan production at the 380°C, 300s case.

Beyond the specific results of this investigation with red oak biomass, the methodology developed in this work can be applied to investigate the segregation ability of other feedstocks. Oak is unlikely to ever be a serious candidate as an at-scale biofuel feedstock but perhaps provides some unique challenges which other feedstocks do not. Of particular interest are any differences in behavior between lignin-derived product lumps between feedstocks that have evolutionarily divergent lignin biosynthesis pathways. It has been established that lignin pyrolysis products generally retain the characteristic ring substituents of the monolignols from which they originate (*81*); this forms the basis of py-MBMS analysis for lignin composition analysis (*99*). The impacts on the staged compositions due to this key biological factor are unknown, but some differences in the behavior of the phenolic species is to be expected.

5.6 Acknowledgements

The Department of Energy is gratefully acknowledged for support under Grant DE-EE0006287 of the Bioenergy Technology Office CHASE program, as well as the National Science Foundation EPSCoR program under Grant No. EPS-0814361.

Chapter 6: Predictive modeling of switchgrass thermochemical products from biomass compositional features

Extracted from a manuscript in preparation

Christopher Waters, Fan Lin, Laura Bartley, Richard Mallinson, Lance Lobban

6.1 Abstract

Switchgrass has been proposed to be a good candidate for a dedicated biomass feedstock for production of thermochemical biofuels (e.g., from fast pyrolysis technologies). Staged thermal fractionation of biomass has been proposed as an alternative to fast pyrolysis as a way to improve catalytic processing of bio-oil to transportation fuels, increasing overall carbon yields and reducing external hydrogen input requirements. As the composition of the biomass dictates the products yielded upon thermal decomposition, and switchgrass has been shown to have significant compositional diversity, we hypothesize that some switchgrasses may have more favorable compositions for staged thermal fractionation than others. To test this, we used analytical pyrolysis to screen characterized switchgrass samples at pyrolysis (500°C) and torrefaction (350°C) conditions and developed predictive linear models of the products at each of those conditions as functions of the composition. The models were then used to predict the thermochemical product yields of an approximately 1,300 member switchgrass association panel. Several of the developed models demonstrate good re-predictive ability, and the relationships between the dependent and independent variables reflect other published literature relating biomass composition and thermochemical products, and provide insight into the most important compositional variables for in vivo biomasses. We conclude that it is possible to select switchgrasses

which are more favorable for the thermal fractionation process, leading to improved catalytic processing and improved process viability.

6.2 Introduction

While this staged thermal fractionation process as described in chapter 5 suggests that separation should be possible based on the biopolymer stability regimes, it is well established that the product distribution from thermal decomposition is widely variable between biomass types and is dependent on biomass compositional features. In chapter 2, three possible schema* of how biomass components may influence the yield or composition of thermal products were introduced based on evidence in the literature. Briefly, schema 1 considers biomass components as the direct sources of thermochemical products (e.g., cellulose decomposition forming levoglucosan); schema 2 considers biomass components to act as catalysts, altering the product yields and selectivities of thermochemical decomposition (e.g., levoglucosan reaction catalyzed by minerals to form smaller products); and schema 3 considers biomass components which interact with other biomass components to alter the product distribution (e.g., cellulose-lignin interaction leading to decreased levoglucosan yield).

In this work, we use analytical pyrolysis to study the effects of composition on the product distribution of the vapors formed upon the pyrolysis and torrefaction of untreated switchgrass biomass. The objective of this study is to identify which compositional features are most important to the yields of the thermochemical products

* In chapter 2 (and the review it is adapted from), the three types of compositional influence are referred to as "Models"; however, in the context of this paper, "schema" is used to avoid ambiguity.

formed, giving us insight into how to best achieve enhanced purity of thermochemical vapor product streams and high yields of favorable compounds for further processing to fuels and chemicals. By torrefying and pyrolyzing biomass of the same species but of known variable composition, we can model the thermochemical product distribution as a function of the composition. The developed models can then be used to predict the behavior of a larger population of switchgrass with known composition.

6.3 Experimental

Switchgrass (*Panicum virgatum*) has been identified as a good candidate for a dedicated bioenergy crop by the United States Department of Energy (*172*). Switchgrass has been studied as a high-value forage for decades (*173*, *174*) and exhibits high mass yields per acre as compared with other herbaceous graminoids (grasses). Because switchgrass is a perennial, it can produce yields year after year with no requirement for seasonal planting and little decrease in yield per acre across many environments (*175*). It exhibits a large degree of variability in the wild and can adapt to the point of dominance across a variety of different environments (*176*). All of these qualities make it an ideal candidate for further study and investigation into suitability for fuels production.

Twenty-two switchgrass samples were chosen from a panel of approximately 1300 biochemically characterized switchgrasses (the 'association panel') provided by the Samuel Roberts Noble foundation. Switchgrass composition was characterized as described in section 3.5. Of the approximately 1300 switchgrasses which were characterized in this manner, 22 were selected for further biochemical and thermochemical characterization.

6.3.1 Thermochemical Characterization

As previously stated, hemicellulose decomposes at a lower temperature range than cellulose while lignin decomposes over a broad range of temperatures. While TGA experiments with cellulose indicate mass loss above 275°C (*164*), high heating rates have been shown to delay thermal breakdown onset to over 350°C (*170*). Therefore, experimental conditions of 350°C for 60 seconds were chosen as the torrefaction conditions to minimize cellulose decomposition while facilitating the breakdown of most of the hemicellulose and the more thermally unstable lignin. For pyrolysis conditions, 500°C for 60 seconds was chosen to achieve complete deconstruction of the biomass (*106*, *165*). The 22 switchgrasses selected for thermochemical characterization were analyzed at these conditions, in the manner described in section 3.2. Tables 6.1 and 6.2 list the identified compounds and their lumps as used in this work. For torrefaction experiments, analysis of variance (ANOVA) indicated that all lumps with the exception of the light oxygenates show significant differences between the 22 samples characterized ($p < 0.05$); for pyrolysis, all lumps showed significant differences between the 22 samples characterized.

Acetic Acid	**Light Oxygenates**	**Methoxyphenols**
Acetic acid (18)	Propanal-2-one (7)	Benzaldehyde, 4-hydroxy- (L23)
	Unknown Light Oxygenate (OU)	Guaiacol (L24)
Alkyl Phenols	2,3-Butanedione (11)	Guaiacol, 4-methyl- (L34)
Phenol (L2)	Unknown: similar to 3-Pentanone (12)	Guaiacol, 4-ethyl- (L40)
Phenol, 4-ethyl- (L22)	Formic acid (15)	Guaiacol, 4-vinyl- (L38)
Phenol, 4-vinyl- (L14)	Hydroxypropanone (19)	Eugenol (L45)
	2-Propenoic acid methyl ester (23)	Syringol (L42)
Furfurals	3-Hydroxypropanal (isomer of	Isoeugenol (cis) (L46)
3-Methylfuran (22)	compound no. 19) (26)	Isoeugenol (trans) (L47)
(2H)-Furan-3-one (29)	2-Hydroxy-3-oxobutanal (33)	Phenol, 4-propenyl- (trans) (L29)
2-Furaldehyde (34)	1-Acetyloxypropane-2-one (38)	Syringol, 4-methyl- (L51)
2-Furfuryl alcohol (37)	2,3-Dihydroxyhex-1-ene-4-one (52)	Vanillin (L41)
Tetrahydro-4-methyl-3-furanone (39)	Methyl-butyraldehyde derivative (67)	C10H10O2 (L43)
Dihydro-methyl-furanone (40)	2-Hydroxy-butanedial (89)	Homovanillin (L49)
4-Cyclopentene-1,3-dione (44)		Syringol, 4-ethyl- (L62)
Dihydro-methyl-furanone (47)	**Pyran Deriv.**	Acetoguaiacone (L50)
Dihydro-methyl-furanone (48)	4-Hydroxy-5,6-dihydro-(2H)-	Benzaldehyde, 4-hydroxy- (L23)
5-Methyl-2-furaldehyde (50)	pyran-2-one (57)	Syringol, 4-vinyl- (L55)
gamma-Butyrolactone (53)	Methyl-dihydro-(2H)-pyran-2-one (60)	Guaiacyl acetone (L56)
(5H)-Furan-2-one (54)	Unknown pyran derivative (OU)	Syringol, 4-propenyl- (cis) (L69)
2-Hydroxy-1-methyl-1-	Unknown pyran derivative (OU)	Syringol, 4-propenyl- (trans) (L70)
cyclopentene-3-one (61)	3- Hydroxy-2-methyl-pyran-4-one (66)	Dihydroconiferyl alcohol (L63)
2-Furoic acid methyl ester (65)		Syringaldehyde (L64)
gamma-Lactone derivative: unknown (70)	**Sugars**	Syringyl acetone (L76)
4-Hydroxy-3-methyl-(5H)-furanone or	Anhydrosugar: unknown (72)	
3-Methyl-2,4-furandione (79)	1,4:3,6-Dianhydro-glucopyranose (83)	
5-Hydroxymethyl-2- furaldehyde (86)	1,5-Anhydro-arabinofuranose (84)	
	1,5-Anhydro-beta-D-xylofuranose (93)	
	Anhydrosugar: unknown (98)	
	Anhydrosugar: unknown (OU)	

Table 6.1. Compounds and lumps used for torrefaction modeling.

Acetic Acid	**Furfurals**	**Sugars**
Acetic acid (18)	Furan (6)	Anhydrosugar: unknown (72)
	(2H)-Furan-3-one (29)	1,4:3,6-Dianhydro-glucopyranose (83)
Alkyl Phenols	3-Furaldehyde (30)	1,5-Anhydro-arabinofuranose (84)
Phenol (L2)	2-Furaldehyde (34)	Anhydro-pento-faranose: unknown (92)
o-Cresol (L10)	2,3-Dihydro-5-methylfuran-2-one (36)	Unknown anhydrosugar (OU)
m-Cresol (L11)	4-Cyclopentene-1,3-dione (44)	
p-Cresol (L11)	Dihydro-methyl-furanone (47)	**Methoxyphenols**
Phenol, 3-ethyl- (L20)	Dihydro-methyl-furanone (48)	Guaiacol (L24)
Phenol, 4-ethyl- (L22)	(5H)-Furan-2-one (54)	Guaiacol, 4-methyl- (L34)
Anisol, 2,4-/2,5-dimethyl- (L30)	Overlapping spectra (gamma-Lactone and unknown) (56)	Guaiacol, 4-ethyl- (L40)
Phenol, 4-vinyl- (L14)		Guaiacol, 4-vinyl- (L38)
	gamma-Lactone derivative: unknown (70)	Eugenol (L45)
Light Oxygenates	4-Hydroxy-3-methyl-(5H)-furanone (79)	Isoeugenol (cis) (L46)
Light Oxygenate (OU) (Multiple peaks)	5-Hydroxymethyl-2-furaldehyde (86)	Syringol (L42)
2,3-Butanedione (11)		Syringol, 4-methyl- (L51)
Hydroxypropanone (19)	**Levoglucosan**	Vanillin (L41)
2-Propenoic acid methyl ester (23)	Levoglucosan (101)	Homovanillin (L42)
3-Hydroxypropanal (isomer of compound no. 19) (26)		Syringol, 4-ethyl- (L62)
	Pyran Derivatives	Acetoguaiacone (L50)
Isomer of compound no. 33 (31)	Isomer of compound no. 57: unknown (51)	Syringol, 4-vinyl- (L55)
1-Acetyloxypropane-2-one (38)	4-Hydroxy-5,6-dihydro-(2H)-pyran-2-one (57)	Guaiacyl Acetone (L56)
Methyl-butyraldehyde derivative (67)	Methyl-dihydro-(2H)-pyran-2-one (60)	Syringol, 4-propenyl- (trans) (L70)
	2-Hydroxymethyl-5-hydroxy-2,3-dihydro-(4H)-pyran-4-one (91)	Syringol, 4-propenyl- (cis) (L69)
Light Gases		
Light Gases (OU)		

Table 6.2. Compounds and lumps used for pyrolysis modeling.

6.3.2 Linear modeling

To more robustly model the thermochemical product yields of the 22 tested switchgrasses as functions of their predictor variables (e.g., compositional features), a multi-step approach was used employing random forests, ordinary least squares regression, and stepwise feature selection. While pure predictive ability of a model on its own can be useful in some circumstances, approaches which may result in good predictive ability (e.g., principal components regression (*177*, chapter 3)) can obscure the contribution of the predictor variables (a 'black box' model). To ensure that the models developed were fully interpretable, an ordinary least squares approach was chosen. However, the compositional data had 64 potential useful features; therefore, significant feature selection was needed in order to avoid over-fitting of the linear regression models built. The source code for the models described in this work is available in appendices B – D.

Before beginning, predictor variables which were defined as combinations of other predictor variables (e.g., non-structural carbohydrates, defined as total soluble carbohydrates plus starch) were removed as these could lead to the development of models with highly correlated predictors. Beyond removal of the known combination variables, calculated linear combinations were removed from the predictor variables using the `findLinearCombos` function from the caret package in R (*178*) (*179*) on the full compositional dataset (1096 observations).

To make initial selections for modeling features, random forests (*177*, *180*) were used to assess the most important compositional features in predicting the thermochemical lump responses. The random forests were built from the `randomForest` package in R

(*181*) with 15,000 trees and 20 random variables sampled for each split. Variable importance scoring was assessed as the mean decrease in accuracy measured by the mean standard error across the forest. The top 7 features from the variable importance scores (by percent increase of mean square error [MSE] upon deletion) were used for ordinary least squares regression (tables 6.3 and 6.4). Compositional features are described in table 3.1.

acetic acid	% Increase MSE		light oxygenates	% Increase MSE
FEST	3.84E+08		**RHA**	1.79E+05
GLC	2.98E+08		C	1.62E+05
Ca	1.70E+08		ARA	1.52E+05
total_sugar.nrel	1.29E+08		HCA	9.26E+04
KL	9.55E+07		Lignin.forage	6.42E+04
PCA	9.51E+07		UA	5.49E+04
lignin..nrel	6.57E+07		xylose.nrel	4.76E+04
alkyl phenols	% Increase MSE		methoxyphenols	% Increase MSE
pCA	3.15E+08		KL	7.22E+05
Ash.forage	1.44E+08		**UA**	4.49E+05
UA	7.84E+07		GLCS	4.22E+05
HCA	7.58E+07		FEST	2.17E+05
FEST	5.52E+07		proG..Z.Factor	1.11E+05
Ash	5.10E+07		PCA	1.00E+05
Ca	4.67E+07		Ash.forage	6.15E+04
furfurals	% Increase MSE		pyran derivatives	% Increase MSE
GLCS	5.98E+06		SUC	1.44E+08
ADL	3.86E+06		XYL	8.59E+07
ARA	3.07E+06		**PENT**	7.73E+07
SUC	2.25E+06		**GLCS**	4.60E+07
Lignin.forage	2.07E+06		GLC	2.77E+07
proG..Z.Factor	1.67E+06		CP	2.11E+07
lignin.nrel	8.48E+05		P.forage	1.63E+07
light gases	% Increase MSE		non-levoglucosan sugars	% Increase MSE
C	6.71E+06		PENT	6.49E+05
KL	5.46E+06		**Ash**	6.11E+05
GLC	3.80E+06		ASH	4.09E+05
Crude.Protein	3.76E+06		**sgratio.nrel**	4.02E+05
RHA	3.51E+06		**EE**	2.51E+05
GAL	2.08E+06		ARA	2.17E+03
XYL	1.89E+06		RHA	-1.43E+03

Table 6.3. Model variable importance scores for top seven torrefaction random forest variables.
Variables in bold were eventually used in the final models.

acetic acid	% Increase MSE	light oxygenates	% Increase MSE
PCA	8.21E+07	**FA**	1.44E+06
GLC	1.97E+07	**UA**	1.28E+06
K	1.73E+07	Ca	1.22E+06
FA	1.23E+07	KL	6.67E+05
PENT	1.04E+07	PENT	4.62E+05
Lignin.forage	1.03E+07	C	4.48E+05
Ash	8.58E+06	**glucose.nrel**	3.76E+05
alkyl phenols	% Increase MSE	methoxyphenols	% Increase MSE
pCA	4.50E+06	**ARA**	1.68E+06
glucose.nrel	2.64E+06	PENT	1.33E+06
PCA	1.25E+06	**lignin.nrel**	8.82E+05
K	8.02E+05	STA	5.26E+05
Crude.Protein	7.29E+05	PENT2	1.86E+05
Ash	3.33E+05	glucose.nrel	1.82E+05
SUC	2.96E+05	xylose.nrel	6.08E+04
furfurals	% Increase MSE	pyran derivatives	% Increase MSE
K	4.59E+05	PENT2	3.59E+06
Crude.Protein	1.19E+05	K	3.11E+06
PCA	6.74E+04	XYL	2.58E+06
GLC	2.29E+03	**GLCS**	2.24E+06
RHA	-1.68E+03	ASH	2.14E+06
PENT2	-3.01E+03	CAL	1.34E+06
Mg	-8.76E+03	PENT	1.09E+06
levoglucosan	% Increase MSE	non-levoglucosan sugars	% Increase MSE
P	2.45E+07	**K**	2.76E+05
K	1.61E+07	**FRU**	1.17E+05
xylose.nrel	1.36E+07	xylose.nrel	3.86E+04
ADL	9.72E+06	HCA	3.82E+04
KL	4.57E+06	**GAL**	3.81E+04
FEST	3.67E+06	GLC	1.84E+04
XYL	3.65E+06	**XYL**	1.41E+04
light gases	% Increase MSE		
UA	8.31E+07		
C	7.74E+07		
KL	7.72E+07		
Ca	5.21E+07		
glucose.nrel	2.96E+07		
FA	2.63E+07		
K	2.55E+07		

Table 6.4. Model variable importance scores for top seven pyrolysis random forest variables.
Variables in bold were eventually used in the final models.

Following this automatic feature selection step, ordinary least squares linear regression models were built using the selected features using the `lm` function from the `stats` package in R (*178*). To further reduce model over-fitting, stepwise feature selection minimizing the Akike information criteria statistic (AIC) (*182*) was used. Both forward & backward approaches were employed. Finally, the `vif` (variance inflation factors) function from the car package in R (*183*) was applied to the stepwise models to test for multicollinearity. Variables having an inflation factor greater than 10 were removed from the initial ordinary least squares model and the stepwise selection was re-run. The final model coefficients and their values are presented in tables 6.5 and 6.6.

acetic acid	*(Intercept)* 720300.34	**FEST** 33219.44	**Ca** -305260.1	**lignin.nrel** -10292.75	
alkyl phenols	*(Intercept)* 198885.89	**pCA** 7785.6	**Ash.forage** 17377.44	**FEST** 22921.93	**UA** -11980.6
furfurals	*(Intercept)* -4698.984	**GLCS** 1898.821	**ARA** 1850.343		
light gases	*(Intercept)* 69998.6703	**Crude.Protein** 3133.2839	**XYL** -227.4472		
light oxygenates	*(Intercept)* 42916.211	**RHA** 4948.003			
methoxyphenols	*(Intercept)* 66015.173	**UA** -1241.789			
pyran derivatives	*(Intercept)* -31770.8978	**PENT** 950.7257	**GLCS** 9845.1746		
non-levoglucosan sugars	*(Intercept)* -6225	**sgratio.nrel** 45954	**EE** 1801		

Table 6.5. Final model coefficients developed for the torrefaction thermochemical lump indicated.

acetic acid	*(Intercept)* 428149.75	**K** 112052.83	**Ash** -7003.11	**PENT2** -881.59		
alkyl phenols	*(Intercept)* 19390.61	**pCA** 1314.24	**K** 24239.00	**SUC** -585.41		
furfurals	*(Intercept)* -16555.59	**K** 7914.94	**GLC** 91.53	**SSE** 413.09		
levoglucosan	*(Intercept)* 51319.39	**P** -621733.33	**K** -45507.10	**xylose.nrel** 510989.78	**ADL** 1173.02	
light gases	*(Intercept)* 2435647.92	**C** -4906.00	**Ca** 183902.51	**glucose.nrel** -599861.90	**FA** -5708.51	**K** 112343.10
light oxygenates	*(Intercept)* 31635.30	**FA** -939.54	**UA** 2185.86	**glucose.nrel** -88782.91		
methoxyphenols	*(Intercept)* 33233.79	**ARA** -563.70	**lignin.nrel** 703.30			
pyran derivatives	*(Intercept)* 87857.55	**GLCS** 4066.25				
non-levoglucosan sugars	*(Intercept)* -4652.18	**K** 3347.06	**FRU** 390.16	**GAL** 610.90	**XYL** 77.21	

Table 6.6. Final model coefficients developed for the pyrolysis thermochemical lump indicated.

6.4 Results & Discussion

6.4.1 Linear modeling

The predictive thermochemical model R^2 values were calculated from the fit of the predictions of the final stepwise models from the compositional data to the observed thermochemical data and are shown in table 6.7. Models with an R^2 value greater than or approximately equal to 0.5 were accepted and those below this value were rejected as too poorly fitted. A total of four linear models with good predictive ability and interpretable coefficients (that is, predicted relationships between composition & product distribution which are in agreement with literature or are otherwise reasonably hypothesized to have a relationship) were developed for the torrefaction products, and

three were developed for pyrolysis. The signs of their coefficients and the schemas these coefficients fit are shown in tables 6.8 – 6.11.

350°C Torrefaction models	R^2 value	500°C Pyrolysis models	R^2 value
acetic acid	0.488	acetic acid	0.466
alkyl phenols	0.811	alkyl phenols	0.739
furfurals	0.501	furfurals*	0.379
light gases†	0.733	levoglucosan*	0.418
light oxygenates*	0.110	light gases†	0.528
methoxyphenols*	0.093	light oxygenates*	0.388
pyran derivatives	0.738	methoxyphenols	0.645
non-levoglucosan sugars†	0.436	pyran derivatives	0.314
		non-levoglucosan sugars*	0.375

*Unacceptable predictive ability. †Uninterpretable coefficients.

Table 6.7. Torrefaction and pyrolysis stepwise model R^2 values for re-prediction of observed values.

model	**Ca**	**FEST**	**lignin.nrel**	**Ash.forage**	**pCA**	**UA**	**ARA**	**GLCS**	**PENT**
acetic acid	-	+	-						
alkyl phenols		+		+	+	-			
furfurals							+	+	
pyran derivatives								+	+

Table 6.8. Torrefaction model coefficients and their signs.

model	**Ash**	**K**	**PENT2**	**pCA**	**SUC**	**ARA**	**lignin.nrel**
acetic acid	-	+	-				
alkyl phenols		+		+	-		
methoxyphenols						-	+

Table 6.9. Pyrolysis model coefficients and their signs.

model	Ca	FEST	lignin.nrel	Ash.forage	pCA	UA	ARA	GLCS	PENT
acetic acid	2	1	1						
alkyl phenols		1		2	1	3			
furfurals							1	1	
pyran derivatives								1	1

Table 6.10. Torrefaction model coefficients and their relationship schemas.

model	Ash	K	PENT2	pCA	SUC	ARA	lignin.nrel
acetic acid	2	2	1				
alkyl phenols		2		1	1		
methoxyphenols						1	1

Table 6.11. Pyrolysis model coefficients and their relationship schemas.

6.4.2 Torrefaction Schema 1 Relationships

Schema 1 relationships, or the direct conversion of biomass components to thermochemical products, are predicted by four of the torrefaction models (table 6.10). The pyran model predicts increasing yields of pyrans with increasing amounts of pentose & soluble glucose measurements. Both pentose sugars and glucose have been reported by others in the literature to form pyran products on fast pyrolysis (*79*, *184*), and the identified pyrans in this lump are both C_5 and C_6. The observed pyrans here are therefore most likely primary products of the decomposition of these sugar types, with moderate to low contribution from the other sugars (e.g., non-glucose hexoses). The

furfural lump model is similar in that the two coefficients are sugar measurements (arabinose & soluble glucose) from which furfurals may be directly produced; however, the predictive ability is lower in comparison with the pyrans. These direct conversion of components to products are explicit schema 1 relationships.

It is worth noting that the compositional predictors for the torrefaction furfurals model (arabinose and soluble glucose) are much lower mass fractions of the original biomass than those of the torrefaction pyrans model (all pentoses + soluble glucose). Figure 6.1 shows the boxplot of the distributions of these sugars in the association panel. While the FID area of the furfurals lump for the 22 tested samples is lower than that of the pyrans lump, the difference is not large enough to reflect the disparities underlying distribution of the predictor sugar compositions. It is possible that there are secondary reactions between the pyrans or the sugars and other components within the biomass (e.g., inorganics) not captured in this predictive model which catalyze ring closure reactions in a schema 2 effect, shifting products out of the pyrans group and into the furfurals group. Patwardhan et al. (*88*) observed a mild catalytic effect of magnesium & calcium salts which increased the production of 5-hydroxymethyl furfural away from levoglucosan (a pyranose anhydrosugar) upon pyrolysis of cellulose, and a similar effect may be occurring here.

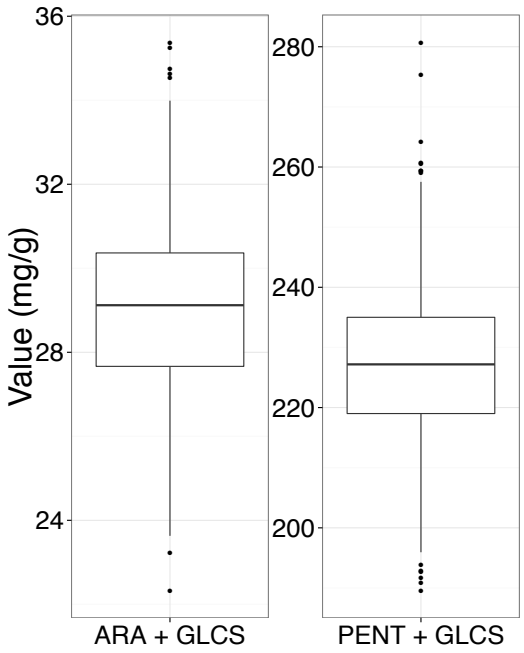


Figure 6.1. Boxplot distributions of the two compositional sugars for the furfurals (left) and pyran (right) models.

The negative sign of the "lignin.nrel" coefficient in the torrefaction acetic acid model implies a schema 1 relationship between composition and thermochemical products (table 6.10). Acetic acid in bio-oil has been reported as a product of primarily hemicellulose pyrolysis (and as relatively minor products of lignin and cellulose pyrolysis) (*12*, *30*, *79*). While production of acetic acid from the polysaccharides results from fragmentation of the sugar monomers or from acetyl groups present on hemicellulose side chains, no hemicellulose measurements are incorporated in the model. However, the negative sign of the "lignin.nrel" coefficient indicates that biomass which contains more polysaccharides (as an overall weight percent) would produce more acetic acid, which is consistent with what others have observed.

The only positive predictive relationship in the torrefaction acetic acid model is the measurement of esterified ferulates. Esterified ferulates are the ferulic acids which have

been incorporated into the cell wall, forming cross-linkages between the lignin and hemicelluloses (*56*, *57*, *62*, *99*) (figure 6.3), in contrast with the measured predictors pCA (p-coumaric acid) and FA (ferulic acid), which are not yet incorporated into the cell wall. The acid group of the ferulic acid forms the ester bond with hemicellulose sugar chains, and the phenolic & methoxy group at the other end of the molecule are incorporated into the lignin polymer via a variety of possible lignin coupling reactions (forming bonds such as β-O-4, β-5, 5-5, 5-O-4, etc) (*55*). Direct production of some of the acetic acid from these molecules may be due to decomposition of the polysaccharide-lignin crosslink, with bond scission between carbons 1 & 2. A speculative mechanism for such a reaction is shown in figure 6.2.

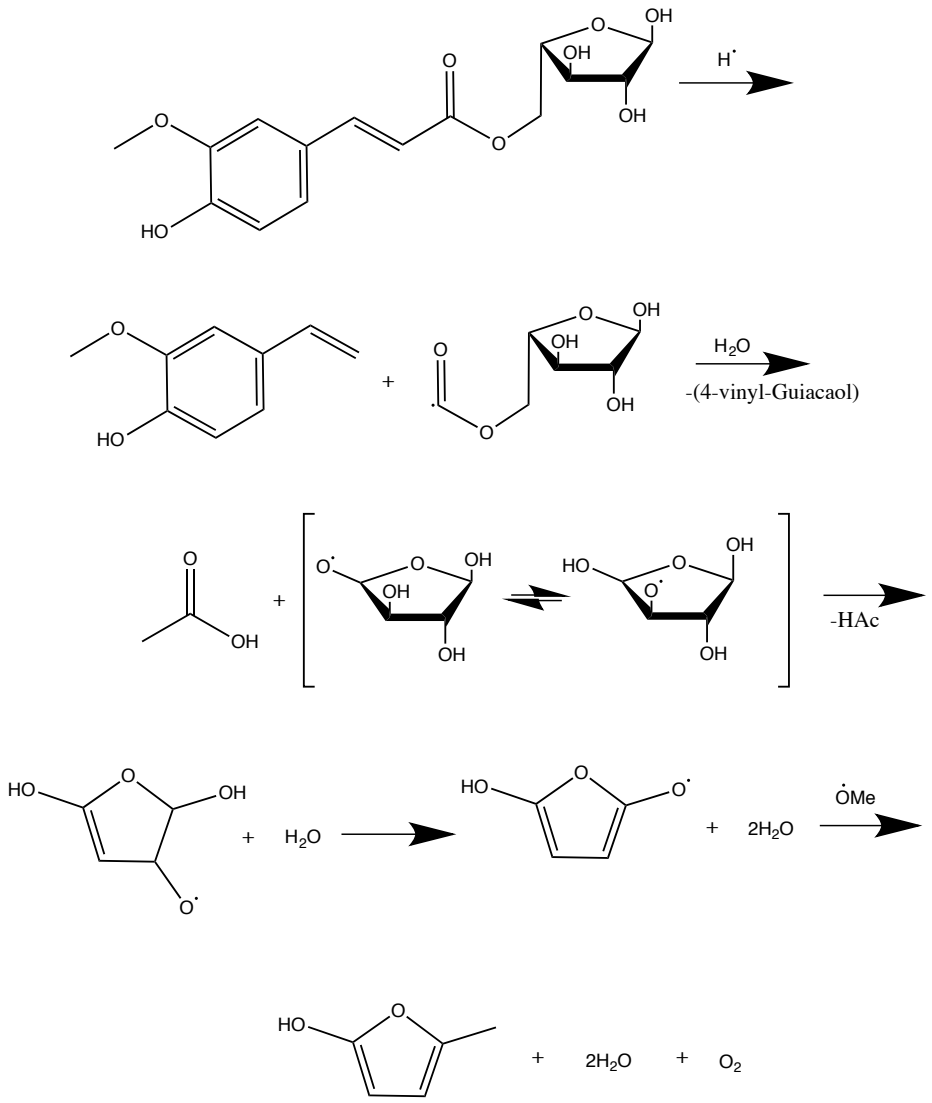


Figure 6.2. Speculative reaction mechanism of acetic acid, phenolic, and furan derivative formation from ferulic acid esterified to an arabinose monosaccharide via a hydrogen radical initiated pathway.

The alkyl phenol model is the fourth torrefaction model which predicts schema 1 relationships. Both the unincorporated p-coumaric acid and the esterified ferulates (figure 6.3) have positive coefficients in the model. Pyrolysis of the p-coumaric acid (with removal of the carboxylic acid) can lead directly to the major alkyl phenolic compound, 4-vinylphenol (*115*). The esterified ferulic acid could also be a source of the alkyl phenols, but a removal of the methoxy substituent would be required.

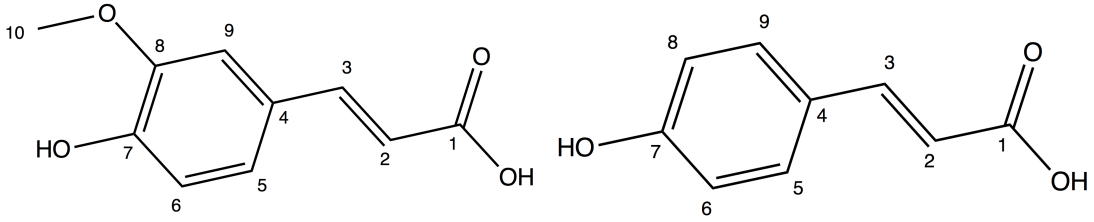


Figure 6.3. Ferulic acid (left) and p-coumaric acid (right).
The acid end of ferulic acid can bond with hemicellulose sugars, and the phenolic end with lignin, to form ester crosslinks between these two biopolymers.

6.4.3 Pyrolysis Schema 1 Relationships

Two of the three pyrolysis models also predict schema 1 relationships between composition and thermochemical products (table 6.11). As discussed above, acetic acid has been reported as a product of primarily hemicellulose pyrolysis and the positive sign of the pentose coefficient in the pyrolysis model corroborates this relationship. The pyrolysis alkyl phenol model, like the torrefaction alkyl phenol model, predicts a schema 1 relationship with the pCA, for the same reasons. Also for the pyrolysis alkyl phenols model, the inclusion and negative coefficient for sucrose suggests an implicit schema 1 relationship much like the lignin in the torrefaction acetic acid – increasing amounts of hemicellulose compete with lignin content in the biomass. Finally, the pyrolysis methoxyphenol model has a positive coefficient for with the "lignin.nrel" lignin measurement. These relationships are consistent with developed literature on pyrolysis products of biomass.

6.4.4 Torrefaction Schema 2 Relationships

Schema 2 relationships are also predicted by the developed torrefaction models. The torrefaction model for acetic acid predicts that increasing amounts of calcium lowers acetic acid yield (table 6.8). This effect may be due to secondary cracking reactions of the acetic acid on the calcium to light gases (CO & CO_2), decreasing the acetic acid

yield. The mass balances reported by Patwardhan et al. (*88*) for inorganic-doped cellulose powder showed the highest yield of 'unaccounted' for the $CaCl_2$ case, and the loss is attributed to formation of gaseous species and water. It is also possible that the calcium acts as a dehydration catalyst for the acetic acid, converting it into acetaldehyde and into the light oxygenates compound lump. While Patwardhan et al. (*88*) did not observe increasing acetaldehyde formation with added calcium in cellulose pyrolysis, acetic acid is not a major primary product of cellulose pyrolysis, unlike hemicellulose pyrolysis.

Another torrefaction schema 2 relationship is predicted between the total ash content of the switchgrass and the alkyl phenol lump. Increasing yields with increases in the total ash content of the switchgrass is somewhat unexpected, as inorganics have been shown to primarily increase the kinds of secondary reactions which lead to char & light gas formation. However, Raveendran and Fahmi (*87*, *185*) independently compared the TGA profiles of various untreated and demineralized biomasses and found that the initial decomposition temperature was increased after demineralization. The inorganics in the biomass act as catalysts for the thermal decomposition of the cell wall polysaccharide/lignin composite matrix. Higher inorganic content in the biomass would promote higher rates of decomposition of the cell wall, allowing for enhanced transport of the lignin decomposition products out of the reacting particle, limiting the char- and light gas-forming secondary reactions.

6.4.5 Pyrolysis Schema 2 Relationships

Two of the three pyrolysis models predict schema 2 relationships (table 6.11), all involving inorganic components. The acetic acid model predicts increasing yields with

potassium and decreasing yields with total ash content; while the promotion of acetic acid formation may suggest enhanced decomposition of hemicellulose sugars via secondary decomposition reactions with potassium (*88*), the acetic acid may be over-cracking to CO/CO_2, much like the effect of calcium on torrefaction acetic acid yields. Finally, the increasing yield of alkyl phenols with increased potassium content may be similar for that proposed for torrefaction above – enhanced decomposition of the cell walls leading to fewer secondary reactions.

6.4.6 Schema 3 Relationships

One possible schema 3 relationship is predicted by the developed torrefaction models (none were apparent in the pyrolysis models) (table 6.10). Increased uronic acid (UA) content is positively correlated with lower yields of the alkyl phenols. Uronic acid is known to act as a crosslinker between lignin and hemicellulose (*71*), in a similar functional role as the esterified ferulates. From a plot of the esterified ferulates and the uronic acid (figure 6.4a), it appears there is a strong inverse relationship between these components in the full switchgrass population. However, this relationship is not present between the unincorporated p-coumaric acid or ferulic acid (figure 6.4b & figure 6.4c). The uronic acid may thus be acting as a competitive crosslinker, limiting the esterification of the ferulates and thus their contribution to the overall alkyl phenol yield.

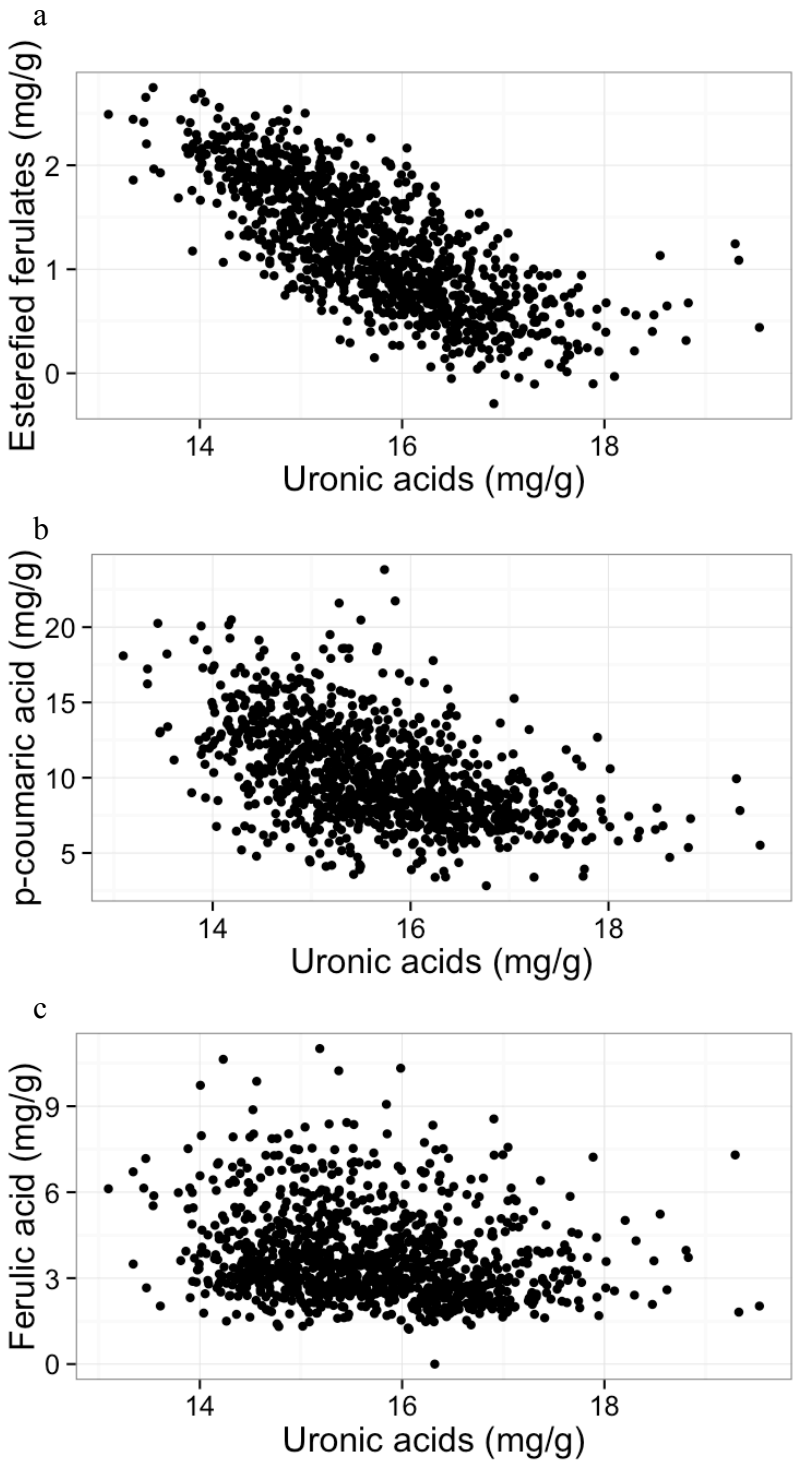


Figure 6.4. Uronic acids vs. esterified ferulates in the association panel.
In (a), a strong inverse relationship and similar functional role implies a competitive relationship between these components of the biomass in crosslinking lignin to hemicelluloses. (b) and (c), the unincorporated ferulates, do not show a strong correlation with the uronic acids.

6.4.7 Bad models – things we couldn't predict

The light oxygenate compound lump consists of molecules which result from secondary reactions of primary thermochemical products and radical species in the vapors or cracking via reactions with inorganics present in char – e.g., a 'model 2' relationship between the composition and the thermochemical products. As the modeling approach used did not consider interaction terms (which would have increased the variables to fit from 64 to over 400), the inability to model the response of this lump for pyrolysis and for torrefaction is somewhat unsurprising. Other models which had poor predictive ability were the torrefaction methoxyphenols and the pyrolysis furfurals, levoglucosan, pyran, and sugars. Further modeling work on these compound lumps may be possible by collecting data on more switchgrass samples.

While the predictive ability of the model with regards to torrefaction & pyrolysis light gases and torrefaction sugars was good, the interpretability of the coefficients for these models was poor. The coefficients of these three models are shown in 6.5 and 6.6; of these, the pyrolysis light gas model is the only one with coefficients which have been demonstrated or theorized by other work in the literature to contribute to the product lump's formation; however, it also includes components which do not reflect observations extant in the literature. This plus the high number of included predictors suggests that this model may be over-fitted to the data and is rejected on those grounds. The other two models (torrefaction light gas & torrefaction sugars) do not include any relationships suggested by other literature. These are therefore likely non-causal correlations, or artifacts of the implemented modeling methodology. Like the light

oxygenates above, further data collection to improve the modeling process could shed more light on relationships between compound lumps and the biomass composition.

6.4.8 Prediction of larger association panel population

Torrefaction predictions

Following the development and evaluation of the predictive models, the four torrefaction models with good predictive ability – acetic acid, alkyl phenols, furfurals, and pyrans – were used to predict the yield of 1,096 complete cases (i.e., no missing predictor data) from the approximately 1300 member association panel. Histograms showing the distributions of the predicted values are shown in figure 6.5. These figures indicate that the test samples used captured the predicted diversity of the association panel's thermochemical product distribution well, as the test samples' observed yields span the predicted yield distributions.

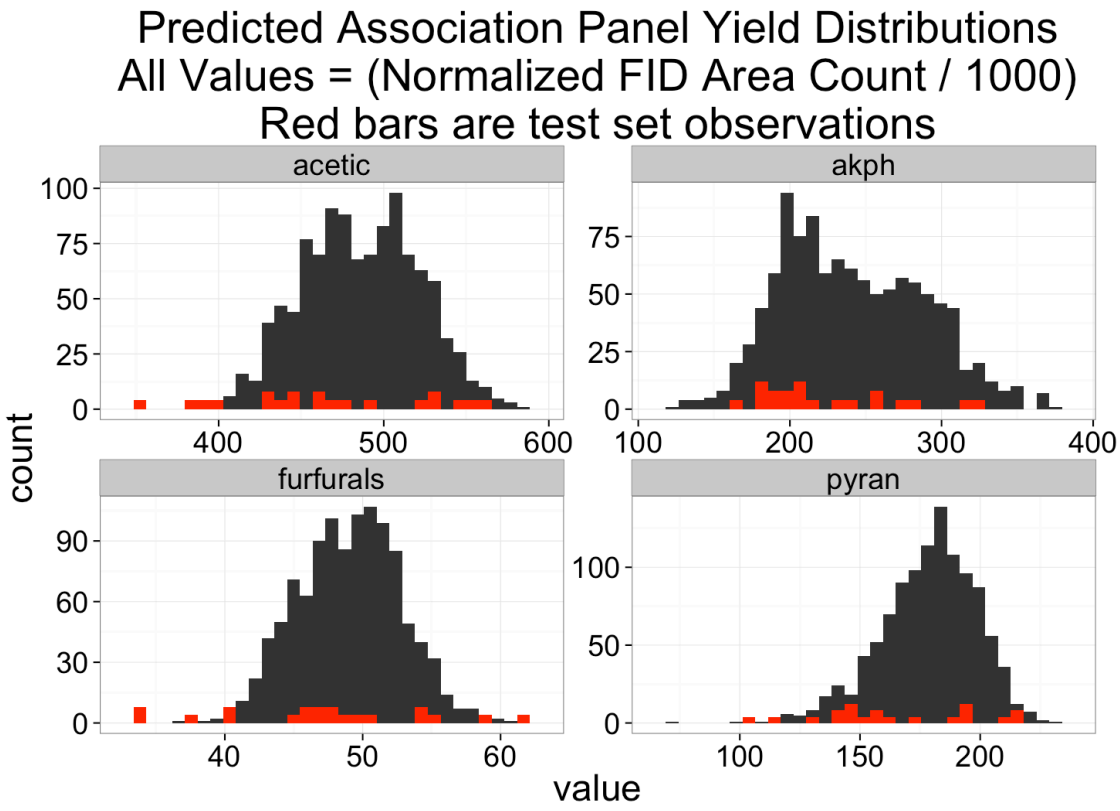


Figure 6.5. Histograms of the predicted yield values of the indicated compound lumps for the 1,096 complete cases of the approximately 1,300 member association panel. Red bars indicate the observed values of the test set, with their height increased for visibility purposes. The observed values span almost the full range of the predicted values for all four models.

To simplify the predicted lumps for further exploration of the predictions (specifically, that three-variable graphs are more easily interpretable than four-variable graphs), the furfurals & pyrans lumps were combined into a single lump, leaving us with three lumps – acetic acid, alkyl phenols, and furfurals/pyrans. These three predicted lumps are plotted together in figure 6.6. The predicted yields of these three lumps show some mild correlation, with high yields of the alkyl phenols generally correlating with high yields of acetic acid and generally higher levels of the combined furfurals & pyrans.

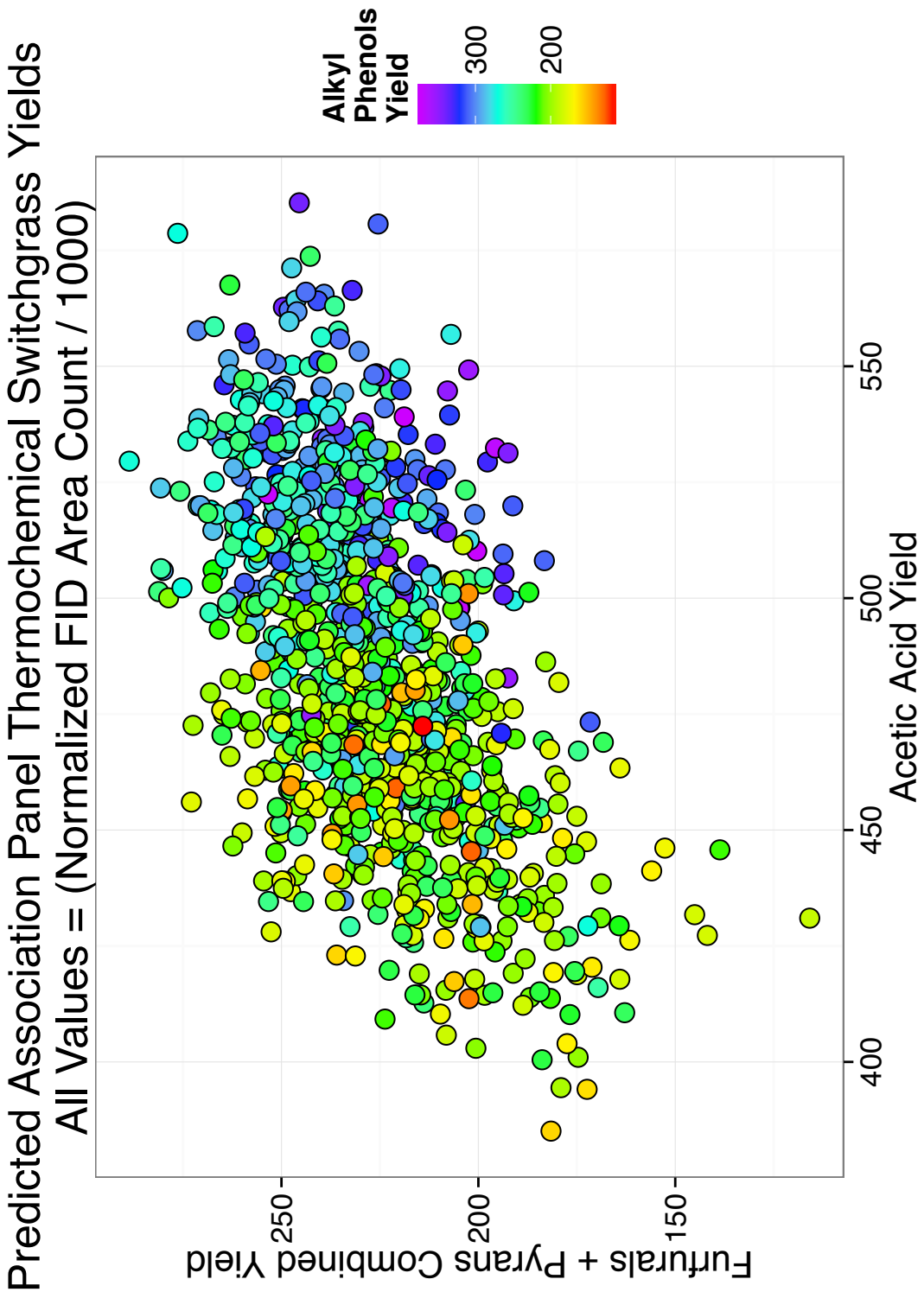


Figure 6.6. Predicted torrefaction compound lump yields for the 1,096 complete cases of the approximately 1,300 member association panel.
Pyrans and furfurals have been combined for simplicity.

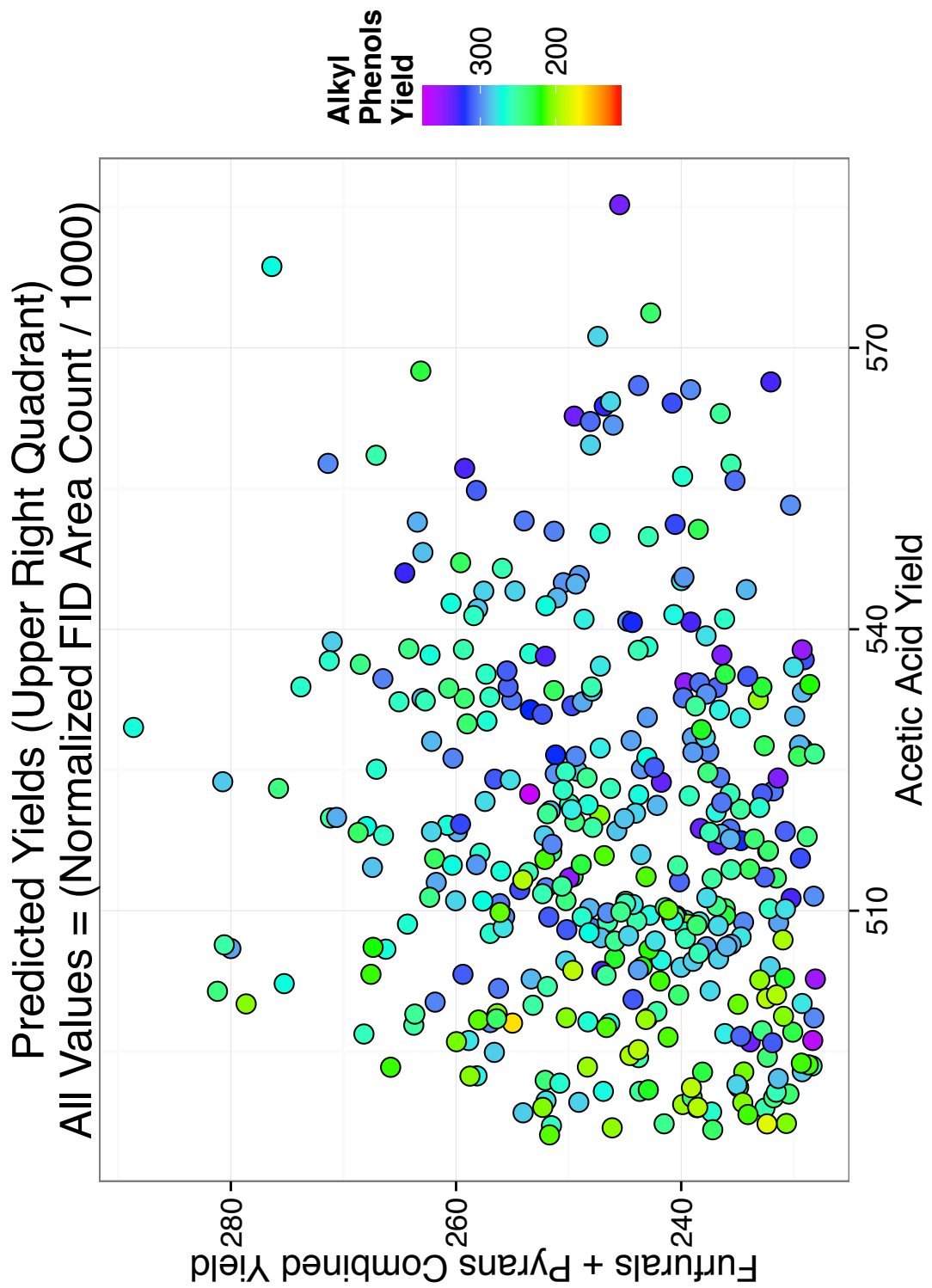


Figure 6.7. Predicted torrefaction compound lump yields for the 1,096 complete cases of the approximately 1,300 member association panel, replotted to show only the upper quadrant of figure 6.6.
Pyrans and furfurals have been combined for simplicity.

The composition of the switchgrasses in this upper quadrant (figure 6.6) are of particular interest for improving the staged thermal fractionation process as discussed in chapter 5. Higher yields of acetic acid, furfurals, and pyrans at these torrefaction conditions is desirable, as this implies better thermal segregation of these compounds into the two lower temperature stages. Figure 6.8 shows the underlying compositional features for these four predictive models, with overlaid highlights of the high and low phenolics yields for switchgrasses which appear in the upper quadrant plot (figure 6.7). From these distributions, it appears that the pCA and esterified ferulates are predicted to have the most influence over the yield of the alkyl phenols for the switchgrasses with high predicted acetic acid, furfurals, and pyrans yields. In consideration of the optimal segregation strategy recommended in chapter 5, switchgrasses with higher amounts of esterified ferulates and pCA may be desirable feedstocks for torrefaction.

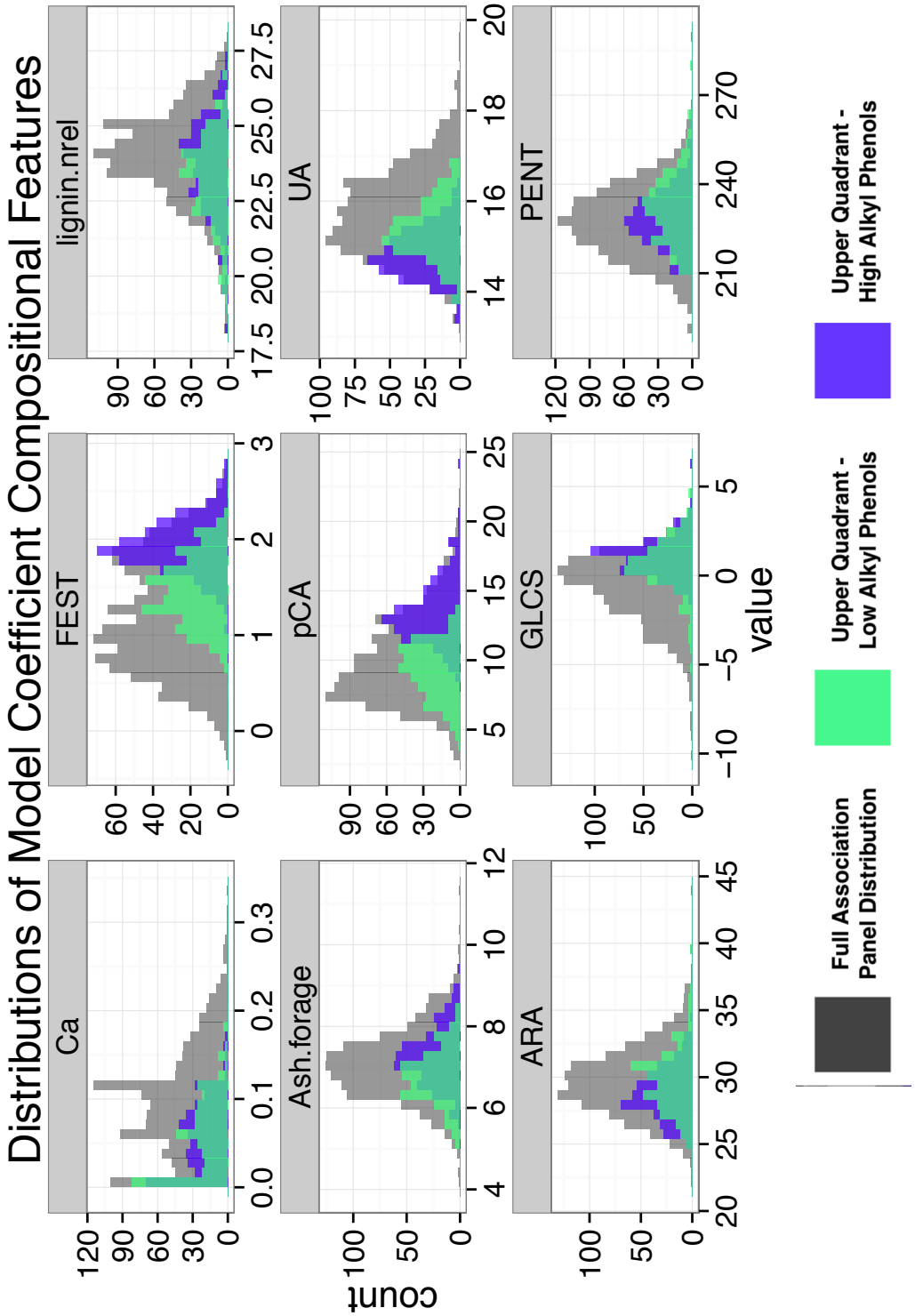


Figure 6.8. Compositional feature distributions from the 1,096 complete cases of the approximately 1,300 member association panel which are used in the 4 predictive models.
The colored shaded areas are the panel members predicted to be in the upper quadrant of figure 6.6 (e.g., in figure 6.7).

Pyrolysis predictions

In the same fashion, the yields of the 1,096 member association panel are predicted for the three pyrolysis models with good predictive ability. Histograms showing the distributions of the predicted values are shown in figure 6.9. Similar to the torrefaction models, these figures indicate that the test samples used captured the predicted diversity of the association panel's thermochemical product distribution well.

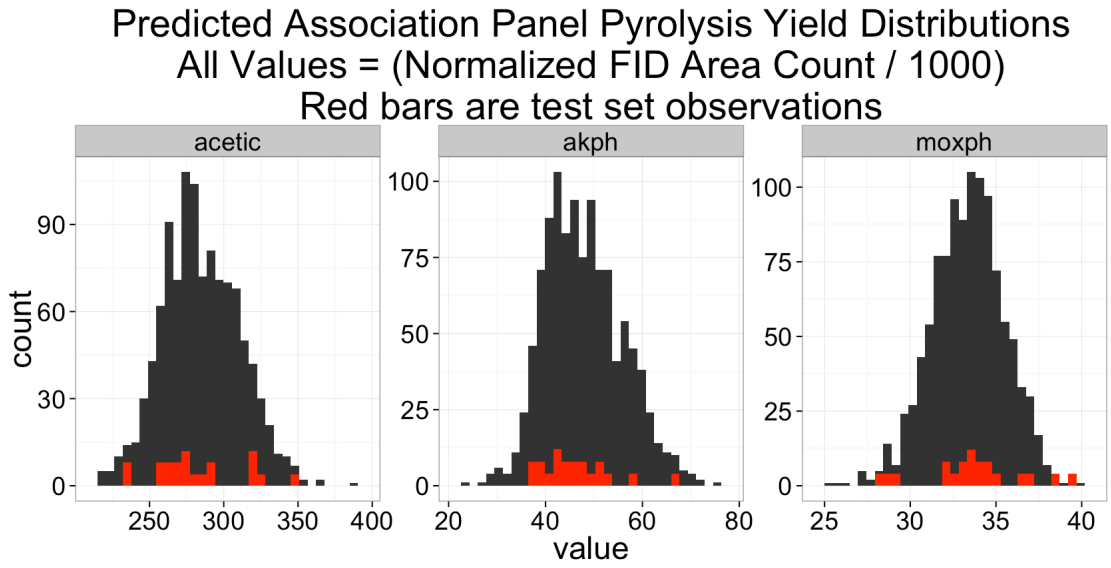


Figure 6.9. Histograms of the predicted yield values of the indicated compound lumps for the 1,096 complete cases of the approximately 1,300 member association panel. Red bars indicate the observed values of the test set, with their height increased for visibility purposes. The observed values span almost the full range of the predicted values for all four models.

Similar to the torrefaction models, the two phenolic lumps' yields are combined and plotted against the acetic acid yields in figure 6.10, with the potassium content (the only compositional features which promote both phenolics and acetic acid yields) as the color axis. These yields show a high degree of correlation and suggest that there are no trade-offs between phenolic and acetic acid yields for pyrolysis of switchgrass.

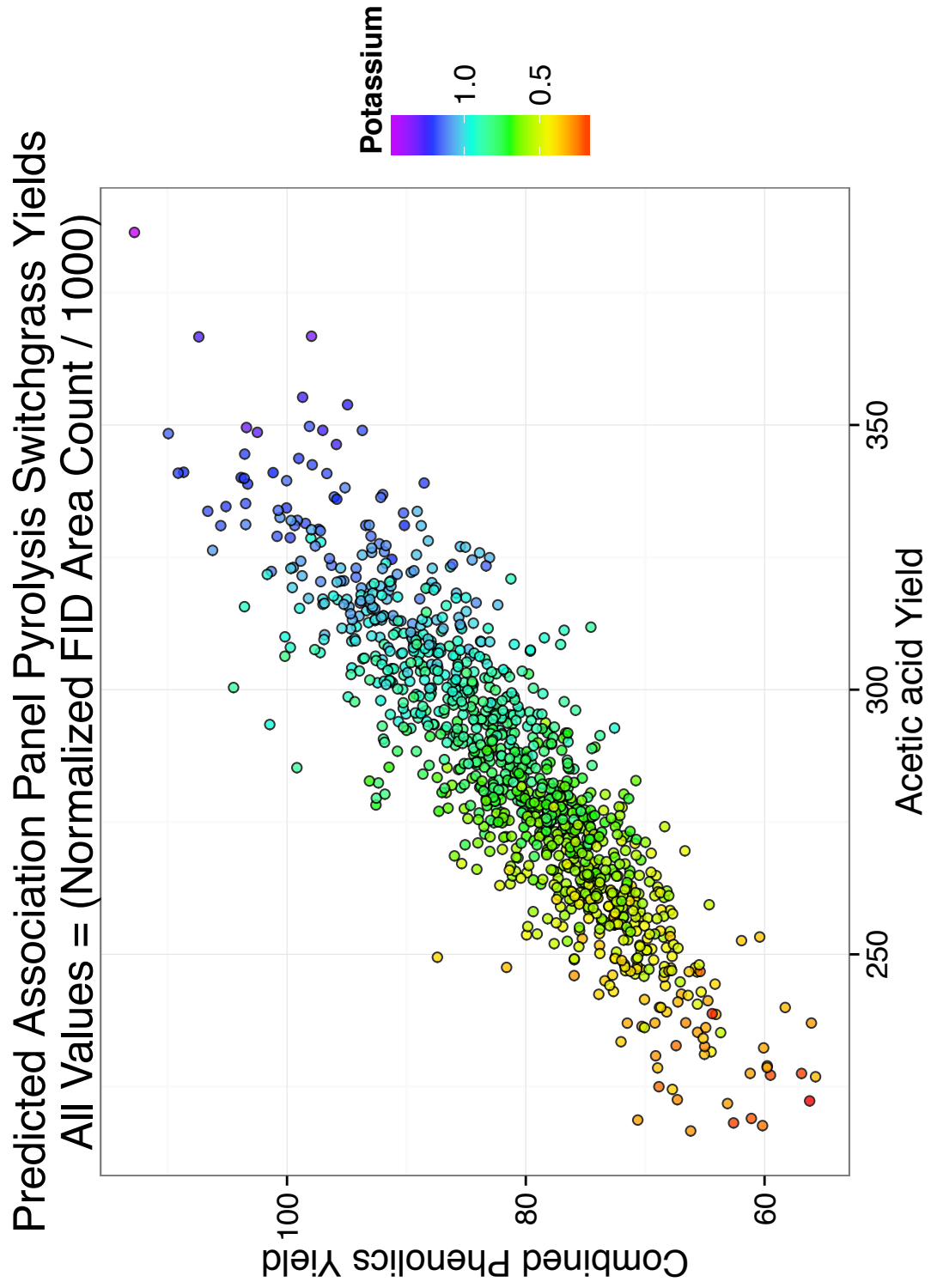


Figure 6.10. Predicted pyrolysis compound lump yields for the 1,096 complete cases of the approximately 1,300 member association panel.

6.5 Conclusions

As discussed in the experimental section, these models linking compositional features to thermochemical product distribution were developed in such a way as to preserve the interpretability of the models, and not necessarily to have the best predictive ability. Of the ten models with good predictive ability developed (for torrefaction: acetic acid, alkyl phenols, furfurals, light gases, pyran derivatives, and sugars; and for pyrolysis: acetic acid, alkyl phenols, light gases, and methoxyphenols), seven were found to incorporate compositional features whose behavior is in agreement with the literature (dis-including torrefaction light gases & sugars, and pyrolysis light gases). This combination of good predictive ability and agreement with the literature have led us to conclude that these seven models are likely to be modeling actual phenomena, especially given the feature elimination steps performed to avoid over-fitting. Understanding how biomass compositional features impact the thermochemical product distribution is challenging due to a variety of factors. Some thermochemical products can be formed by multiple source biopolymers (ex: furfurals from hemicellulose & cellulose), obscuring the relative contribution from each (competing schema 1 relationships). Even the known primary decomposition products (e.g., levoglucosan or phenolics) can interact with others, creating secondary reaction products which can be similar to other primary products (e.g., schema 2 relationships, such as cellulose decomposition intermediates forming furfurals by reacting with inorganics, or calcium cracking low molecular-weight species further to light gases). These overlapping products and interactions prohibit some of the most straightforward analysis techniques, such as direct correlation analysis. Compounding the issue, many studies in the

literature on this topic have used segregated biopolymers (i.e., extracted cellulose/hemicellulose/lignin), which do not reveal the possible interactions between these components (e.g., schema 3 relationships) that have been shown to occur in true biomass feedstocks (*92*).

This screening of a well-characterized single species is the first of its kind in the literature to our knowledge. While the aggregate analysis of the products of biomasses of known composition can mitigate some of the issues mentioned above, this type of approach presents its own set of challenges, primarily related to the amount of compositional data involved and the limitations of the thermochemical experimental methods. With 64 compositional features which could have been considered, the number of observations would ideally be in the hundreds to permit a 'traditional' linear modeling approach, but time and equipment constraints on the thermochemical data collection approach restricted us to 22 samples. This prompted the need for the somewhat sophisticated automatic feature selection techniques used, and prohibited the development of models which take into account interaction terms between thermochemical products and compositional features, limiting the degree to which secondary reaction effects can be explored. This is likely the reason that we were unable to model the light oxygenates and light gases well, as discussed above.

The compositional diversity of the switchgrasses was initially hypothesized to result in a broad range of outcomes for the thermochemical product distribution; in both the test set observations and the predicted association panel yields, we observed this to be correct (Figure 6.11). These large differences in the product distributions, attributable to differences in the underlying compositions (or structures), suggest that compositionally

tuning feedstocks for specific catalytic upgrading strategies could result in greatly enhanced process performance, and underscores the need for further research into compositional influences on product yields.

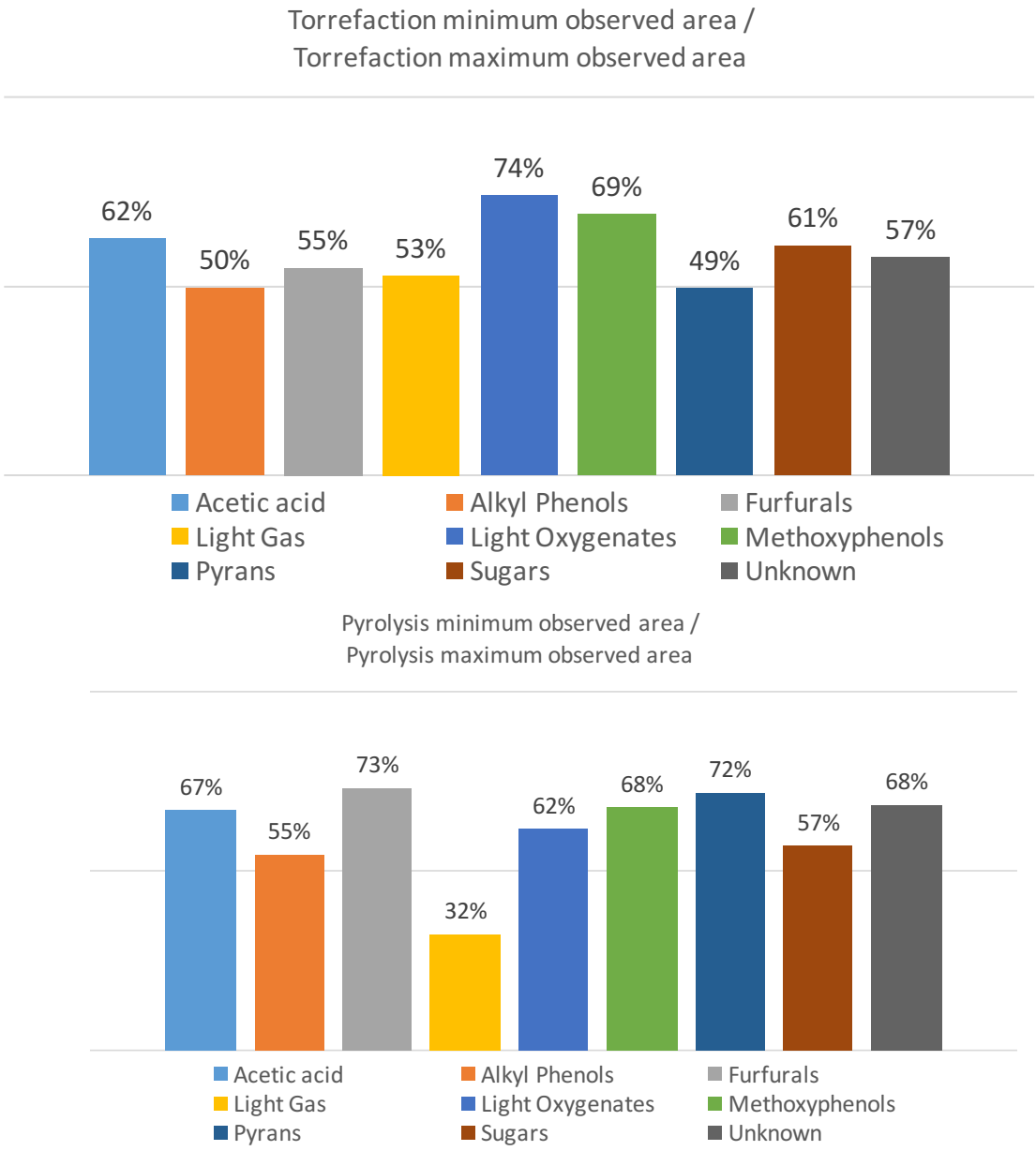


Figure 6.11. Minimum values of the observation set as a percentage of the maximum values of the observation set for torrefaction and pyrolysis thermochemical data. Large differences in the yield distributions were observed for most of the compound lumps.

6.6 Acknowledgements

The work has been financially supported by the NIFA-USDA Sun Grant Initiative (AB-5-68110.OU1), the National Science Foundation (EPSCoR 0814361), and the US Department of Energy (DE-FG36GO88064). The help of Adam Stevens & Tim Pegg at OU and Malay Saha & Charles Brummer of the Samuel Roberts Nobel Foundation was also instrumental in this work.

Chapter 7: The effect of genetic modification of the lignin biosynthesis pathway on low temperature pyrolysis product yields

Extracted from a manuscript in preparation

Christopher Waters, Fan Lin, Laura Bartley, Richard Mallinson, Lance Lobban

7.1 Introduction

In chapters 5 and 6, we have seen that there are means to exert control over the thermochemical degradation products via process and feedstock compositional factors. The study in chapter 5 showed that it is possible to achieve segregation of thermochemical products into separate product streams, which can then have catalytic valorization strategies appropriately tailored in order to improve desirable outcomes (e.g., maximizing carbon yield, improving process economics, etc.). Complementing this, the study in chapter 6 showed that not only do different organisms within a single species have different thermochemical product distributions but also revealed the compositional factors which may be responsible for the observable differences. However, the role of heritable traits in the biomass was not studied in the previous study with switchgrass, and may be another area in which thermochemical biomass feedstock improvements may be realized.

One of the desirable segregations to achieve via staged thermal fractionation was the 'clean' removal of hemicellulose via a low-temperature first stage treatment. At the conditions used in that study (270°C, 20 minutes), small amounts of both cellulose and lignin were decomposed and volatilized in addition to the hemicellulose. This initial low temperature treatment can likely be optimized to minimize cellulose degradation using solely time and temperature parameters, but TGA studies of lignin decomposition indicate that some lignin decomposition will be unavoidable at conditions which

decompose hemicellulose (*27-29*). While the use of a low-lignin feedstock could help to minimize the selectivity to lignin-derived phenolics in a first stage product stream, high amounts of lignin-derived phenolics are desirable from a whole-process viewpoint as the high C:O ratio and profligate C-C bonds could lead to relatively low-input fuel range molecules. However, if the thermal stability of the lignin could be altered to suppress decomposition at lower temperatures, especially without reducing the total lignin amount, both of these desirable outcomes could be achieved.

As discussed in section 2.1, lignin is a complex, cross-linked polymer comprised primarily of three monomer (***H***, ***G***, and ***S***) units cross-coupled via oxidative reactions during incorporation in the cell wall (*55*). Figure 7.1 shows an overview of these three lignin monomers and their resulting structures in the polymer. The most frequently occurring polymer unit is the β–O–4 structure (**A**), accounting for more than half of the inter-unit linkages; other commonly occurring units are β–5 (**B**), β–β (**C**), 5–5 (**D**), 5–O–4 (**E**), and β–1 (**F**) (other units are various special cases, and are included by the author for completeness). Increased production of ***G*** monomers over ***S*** monomers (i.e., a lower ***S***/***G*** ratio) leads to an increased prevalence of β–5, 5-5, and 5-O-4 linkages in the lignin polymer over β–O–4, β–β, and β–1 linkages due to the increased availability of the 5 position as a reaction site (*55*). Table 7.1 summarizes the results from DFT calculations of the enthalpy of dissociation for these six bonds. These suggest that, in general, we should expect the initiation of the radical depolymerization reactions of the lignin biopolymer to be less favorable at lower temperatures for lower ***S***/***G*** ratios. This would result in lower selectivity to lignin decomposition products at lower temperatures. As lignin synthesis is a regulated biological process (*55*), the production

of all three monomers has been demonstrated to be impacted by heritable traits. Figure 7.2 shows an overview of the lignin biosynthesis pathways established and the enzymes (when known) that catalyze various steps in the process.

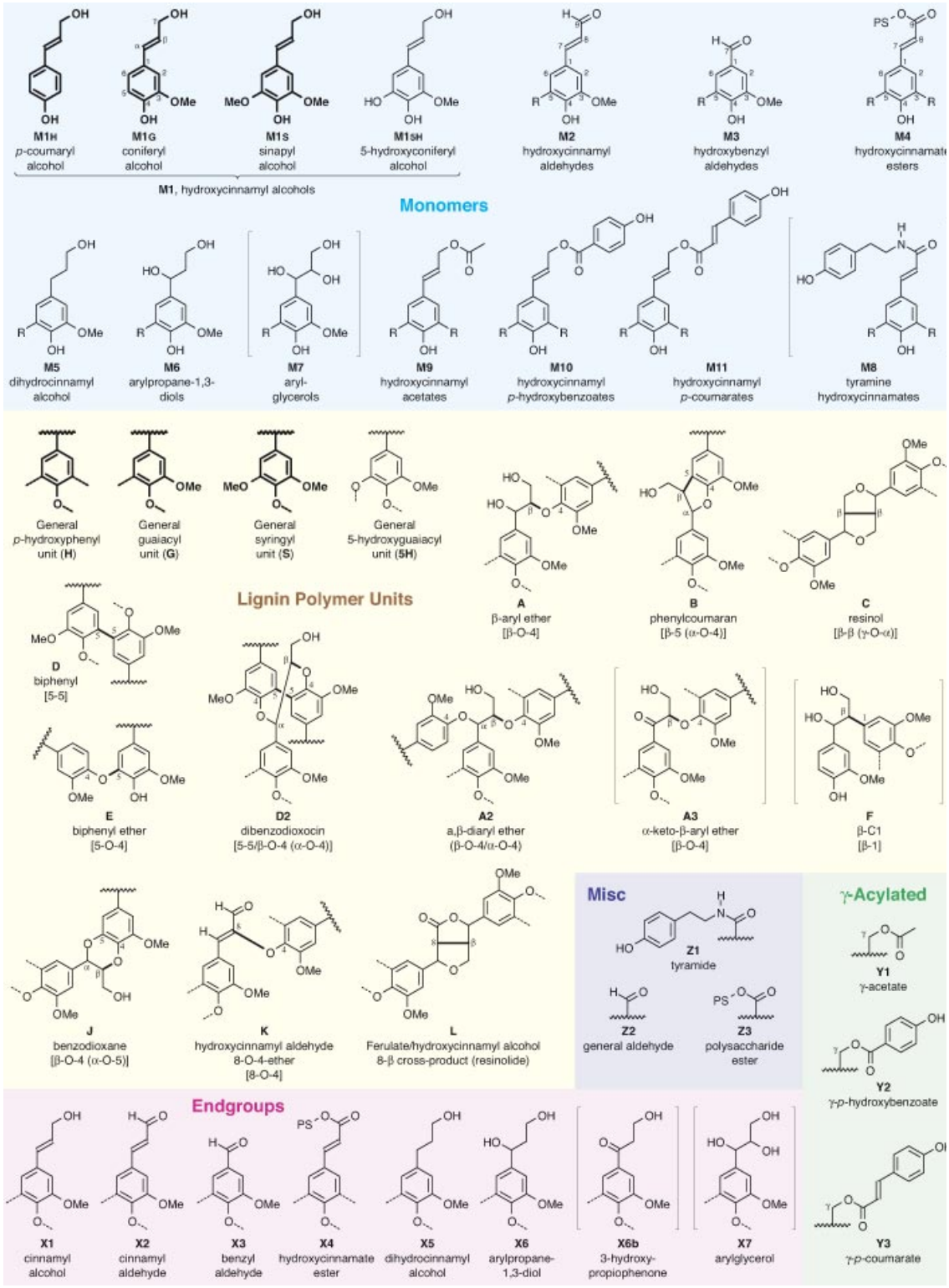


Figure 7.1. Lignin monomers and structures in the polymer. From (*55*).

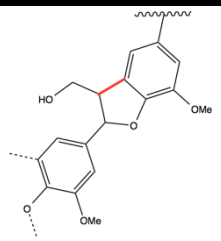
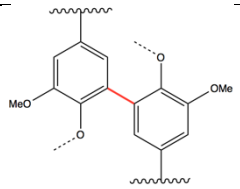
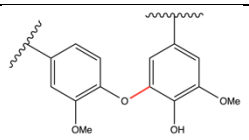
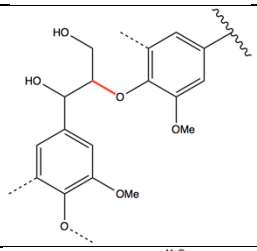
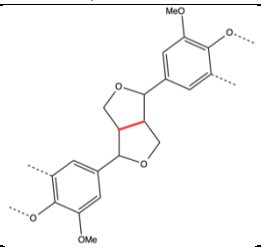
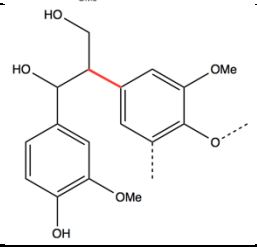
Bond	Structure	ΔH_{diss} , kcal/mol, 298K	Reference
β-5		101.6 – 107.1	(*186*)
5-5		111.8 – 118.1	(*186*)
5-O-4		77.7 – 82.5	(*187*)
β–O–4		67.7 – 71.3	(*186*)
β-β		81.1 – 82.6	(*188*)
β-1		64.7 – 69.1	(*187*)

Table 7.1. Summarized results from DFT calculations of the enthalpy of dissociation for six common lignin crosslinks.

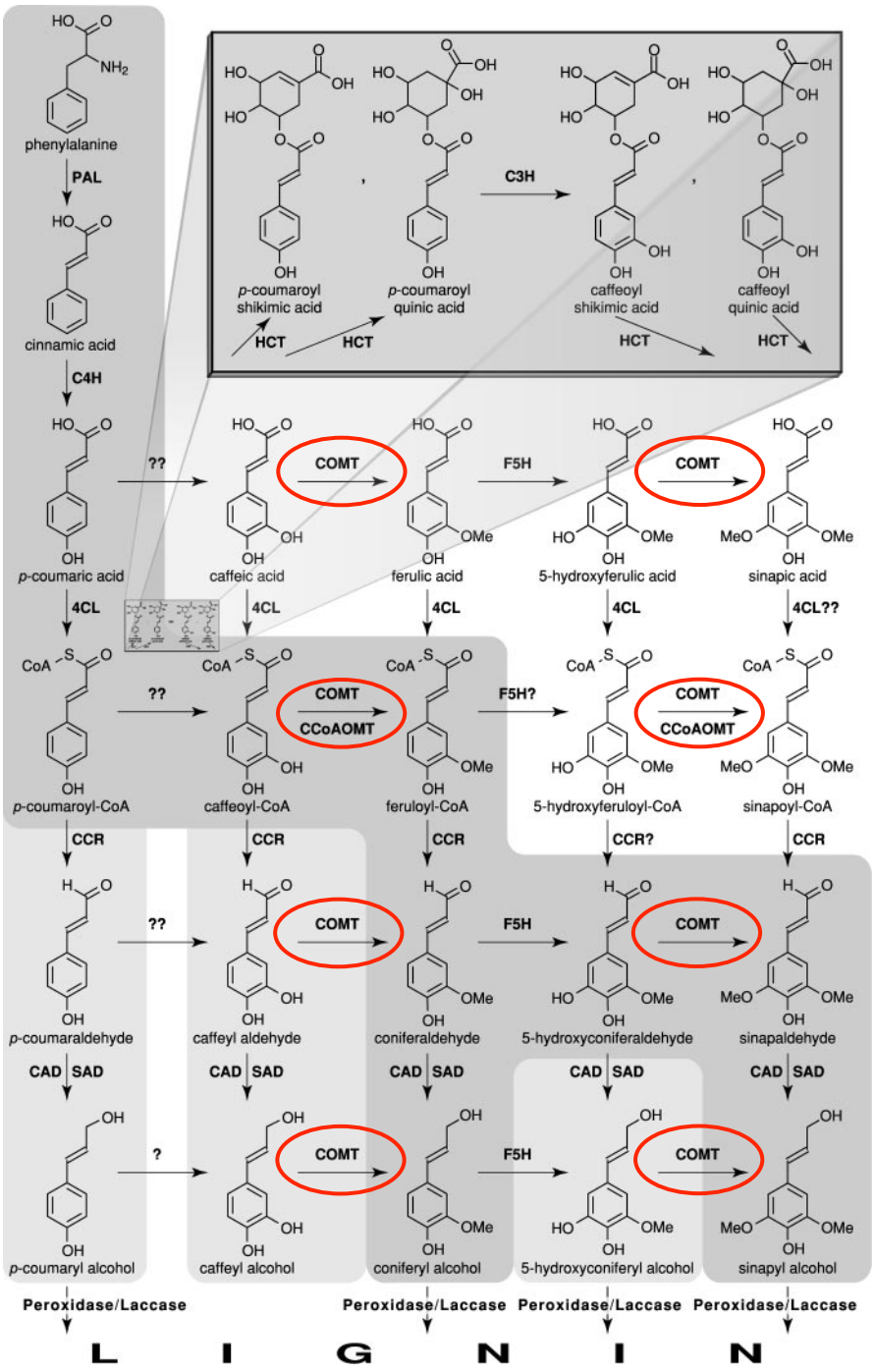


Figure 7.2. Phenylpropanoid and monolignol biosynthetic pathways. From (*55*)

In this study, we compare the low-temperature thermochemical products of a genetically modified switchgrass variant to those of the unmodified wild-type. The

genetic modification down-regulates the production of Caffeic acid O-methyltransferase (COMT), an enzyme which catalyzes the production of ***S*** monomers (figure 7.2) (*55*). This lowers the ***S***/***G*** ratio of the resulting lignin which is incorporated into the cell wall. We hypothesize that this modification will result in lower phenolic product yield for the down-regulated mutants as compared with their unmodified wild-type. To our knowledge, this is the first reported study of the effect of genetic modification to modify a major cell wall biopolymer on thermochemical product distributions.

7.2 Experimental

C. N. Stewart and associates generously provided eight samples (four wild-types and four mutants) from the COMT2-RNAi knock down line of transgenic switchgrass as described and characterized in Baxter et. al. (*126*), along with their corresponding biochemical characterization data. Table 7.2 lists the compositional information for these samples. The wild-types and mutants exhibit significant differences in their ***S***/***G*** ratios, while the total lignin content is only slightly lower in the mutants. The four wild-types and four mutants are all from cloned plants of their respective category.

Each sample was prepared for thermal decomposition in the method described in section 3.2.1. Each biological replicate was run with three technical replicates in the pyroprobe apparatus. The average sample size was 1038 mg with a 5.04% standard deviation. The pyrolysis temperature was 290°C with a hold time of 120 seconds. Vapors were collected & analyzed as described in section 3.2. The identified compounds and their lumps as used in this work are listed in table 7.3.

	Sample ID	Struct. Ash	Lignin	Acetyl	Total Carbohyd.	Cellulose	Hemi-cellulose	Xylan	Galactan	Arabinan	Mannan	S/G ratio
Wild-types	3	2.0%	22.0%	4.6%	68.3%	38.2%	30.1%	26.7%	0.5%	2.6%	0.3%	0.69
	5	1.8%	22.1%	4.7%	67.7%	38.5%	29.2%	26.0%	0.6%	2.5%	0.2%	0.76
	11	3.5%	21.9%	4.4%	68.6%	38.8%	29.8%	25.3%	1.3%	3.1%	0.1%	0.68
	14	2.3%	22.4%	4.5%	70.0%	40.3%	29.7%	26.0%	1.1%	2.6%	0.0%	0.77
Mutants	4	1.9%	20.1%	4.7%	73.4%	40.8%	32.6%	29.2%	0.6%	2.6%	0.2%	0.42
	7	1.6%	20.4%	4.8%	73.1%	40.7%	32.4%	28.4%	1.2%	2.8%	0.0%	0.45
	9	2.5%	20.6%	5.0%	72.7%	40.1%	32.7%	28.2%	1.2%	3.2%	0.1%	0.44
	10	2.3%	20.8%	4.7%	67.6%	37.6%	30.0%	26.3%	0.7%	2.7%	0.4%	0.48

Table 7.2. Switchgrass compositional information. From (*126*)

Acetic Acid	**Levoglucosan**	**Methoxyphenols**
Acetic acid (18)	Levoglucosan (101)	Guaiacol (L24)
		Guaiacol, 4-ethyl- (L40)
Alkyl Phenols	**Light Oxygenates**	Guaiacol, 4-vinyl- (L38)
Phenol (L2)	Light Oxygenate (OU) (Multiple peaks)	Syringol (L42)
Phenol, 4-ethyl- (L22)		Eugenol (L45)
Phenol, 4-vinyl- (L14)	2-Propenoic acid methyl ester (23)	Vanillin (L41)
		C10H10O2 (L43)
Furfurals	3-Hydroxypropanal (26)	Syringol, 4-vinyl- (L55)
2-Furaldehyde (34)	Isomer of compound no. 33 (31)	Syringol, 4-propenyl- (cis) (L69)
2-Furfuryl alcohol (37)	1-Acetyloxypropane-2-one (38)	
Dihydro-methyl-furanone (47)		Syringol, 4-propenyl- (trans) (L70)
Dihydro-methyl-furanone (48)	**Pyran Deriv.**	Sinapyl alcohol (cis) (L79)
5-Methyl-2-furaldehyde (50)	4-Hydroxy-5,6-dihydro- (2H)-pyran-2-one (57)	Syringyl acetone (L76)
gamma-Butyrolactone (53)		
(5H)-Furan-2-one (54)	Methyl-dihydro-(2H)-pyran-2-one (60)	**Sugars**
		Anhydrosugar: unknown (72)

Table 7.3. Compounds identified and used in this study.

7.3 Results

The thermochemical products from the 8 samples run are shown in figure 7.3. Comparable amounts of methoxyphenols were produced by both the wild-types & mutants; however, mutants produced much less alkyl phenols per milligram of switchgrass. A paired t-test for means was performed to assess if the differences in the means of the normalized yields were significant ($p > 0.1$). The results are shown in table 7.4 and indicate that there is not a significant difference in the yields per gram of biomass of any of the compound groups except for the alkyl phenols.

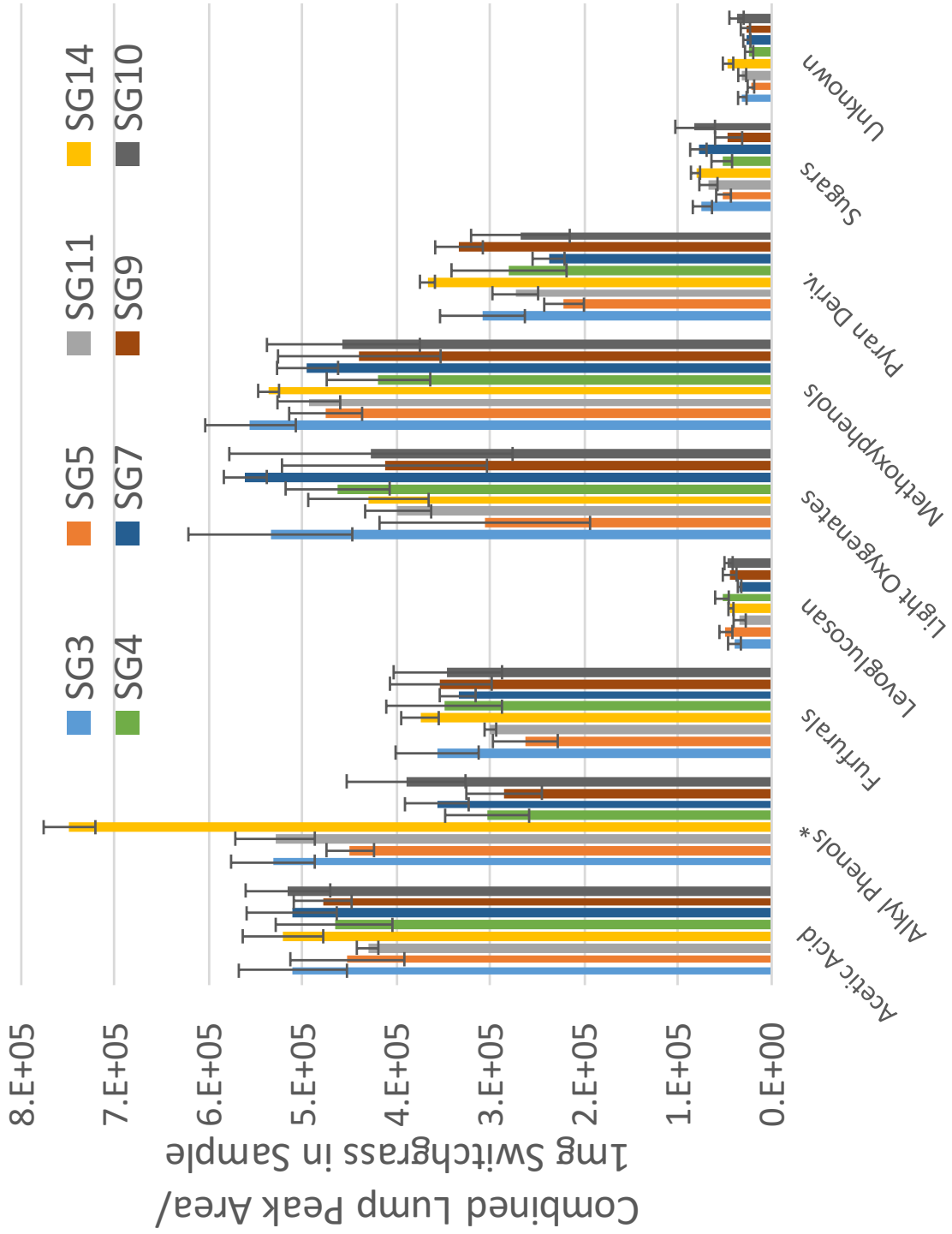


Figure 7.3. Combined lump yields for all compounds and samples in this study. Yield values are (Summed compound peak areas) / (mg of raw switchgrass). All samples run at 290°C, 120s. Each value reported is the mean of three technical replicates; error bars indicate 95% confidence interval of the mean for these three replicates.

The two phenolic compound groups (alkyl phenols and methoxyphenols) were summed together and the result was normalized per milligram of lignin in the sample; this result is shown in figure 7.4. All of the wild-types tested yielded more phenolics per milligram of lignin than did the mutants, although the yield of sample SG5 was virtually identical to the highest yielding sample of the wild-type (SG7).

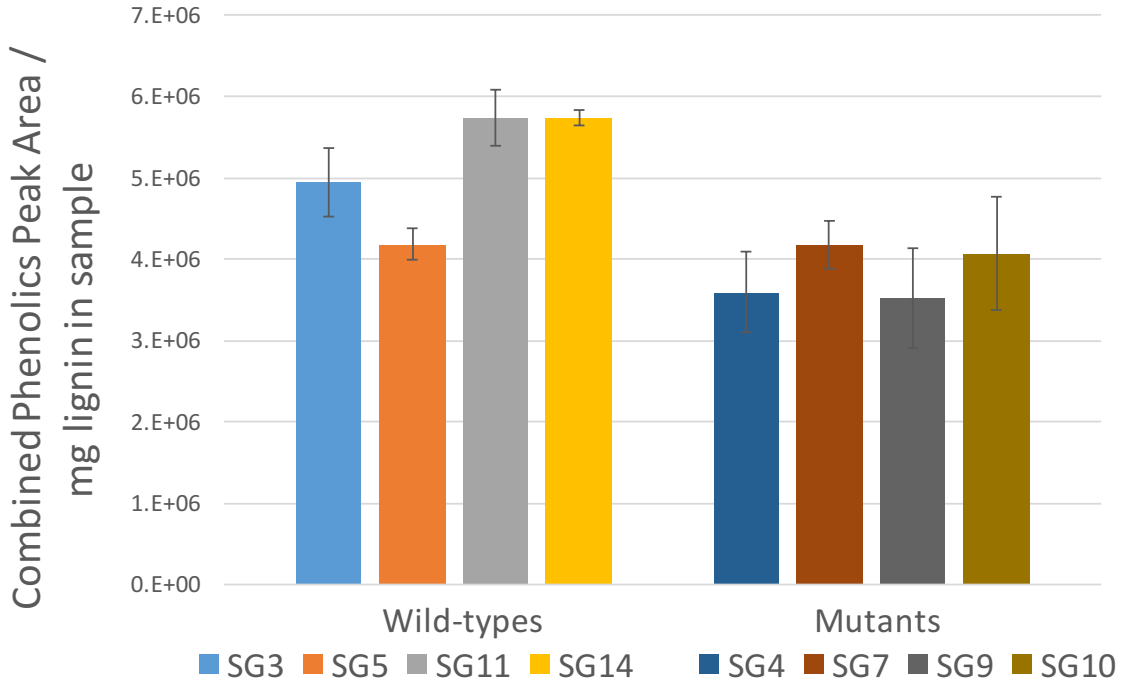


Figure 7.4. Wild-type vs. Mutant phenolic compound yields.
Yield values are (Summed phenolic compound peak areas) / [(mg of raw switchgrass) * (wt % lignin)]. All samples run at 290°C, 120s

A paired t-test for means was performed to assess if the differences in the means of these lignin-normalized combined phenolics yields were significant ($p > 0.1$). The results are shown in table 7.4 and indicate that there is a significant difference in the yields per gram of lignin in the biomass of the phenolics at the conditions tested.

t-Test: Paired Two Sample for Means		
	Wild-type	*Mutant*
Mean	4883371.75	3839352.509
Variance	4.2142E+11	1.09436E+11
Observations	4	4
Pearson Correlation	0.0227	
Hypothesized Mean Difference	0	
df	3	
t Stat	2.8925	
P(T<=t) one-tail	0.0314	
t Critical one-tail	2.3534	
P(T<=t) two-tail	0.0629	
t Critical two-tail	3.1824	

Table 7.4. t-test of sample means of phenolic yields normalized to the amount of lignin in switchgrass samples.

Figure 7.5 shows the contributions of each of the identified phenolic compounds to the total lignin-normalized yield results, color-coded by the originating monomer unit. 4-vinyl phenol (***H*** monomer derived) and 4-vinyl guaiacol (***G*** monomers derived) dominate the results, and few dimethoxy-substituted phenolic products (***S*** monomer derived) are observed. Most of the differences observed in the normalized phenolic yields of the mutants and wild-types are due to the difference in the 4-vinyl phenol (figure 7.6).

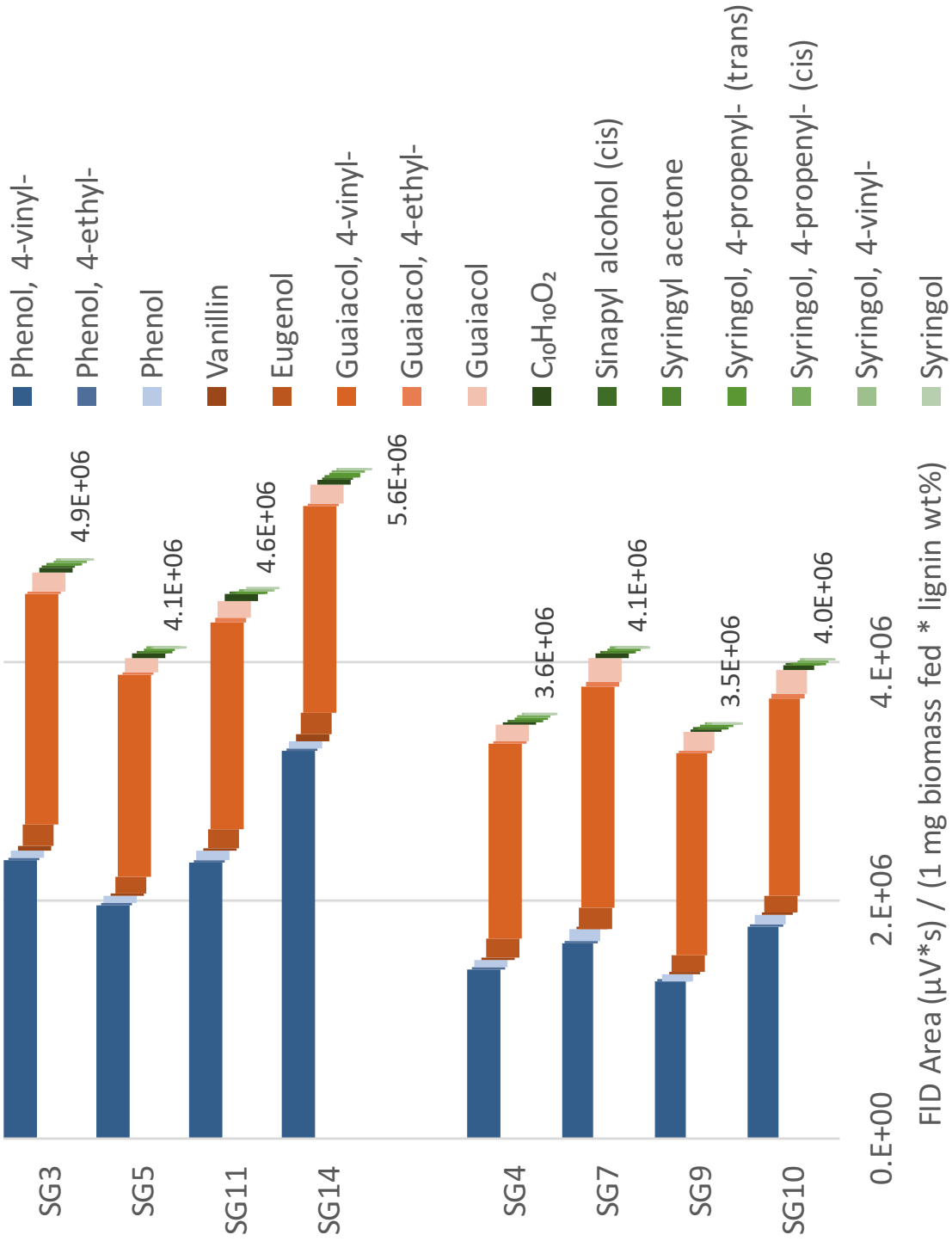


Figure 7.5. Yields of all tracked phenolic compounds colored by originating monomer Yield values are (Summed phenolic compound peak areas) / [(mg of raw switchgrass) * (wt % lignin)]. All samples run at 290°C, 120s

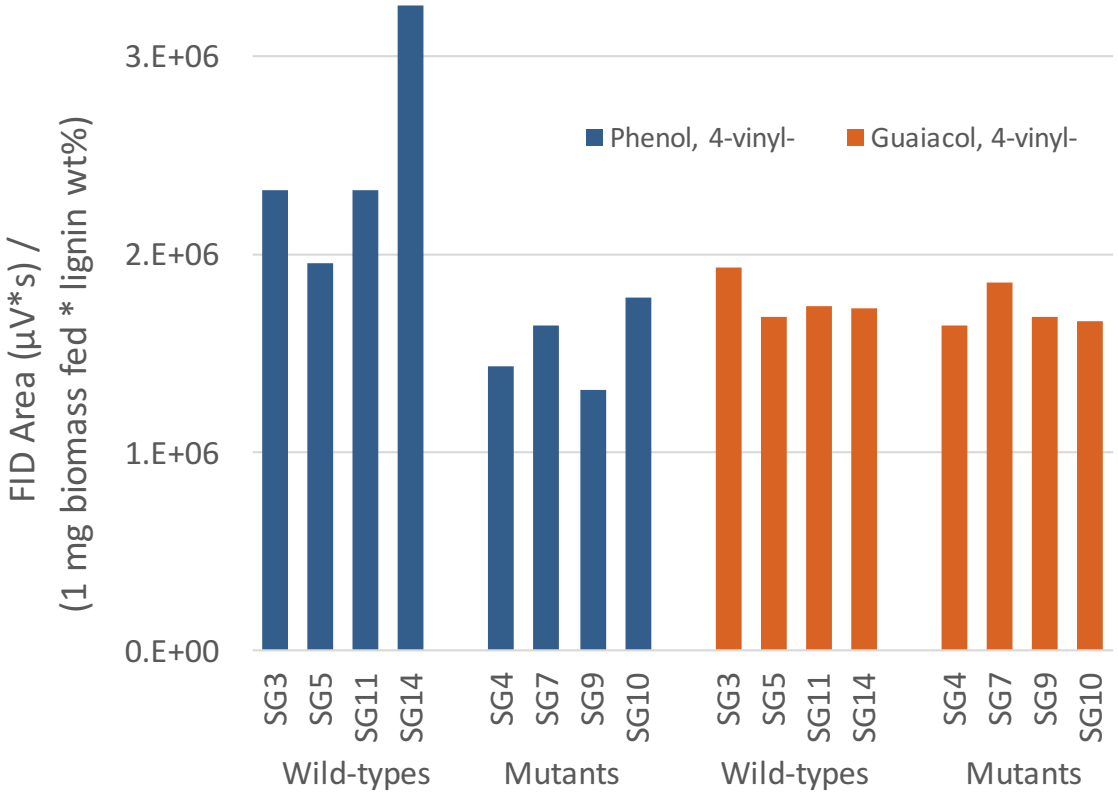


Figure 7.6. Yields of two most abundant phenolic compounds in this study. Yield values are (compound peak area) / [(mg of raw switchgrass) * (wt % lignin)]. All samples run at 290°C, 120s

The ***H*** monomer production pathway is not impacted by the COMT enzyme (figure 7.2) (*55*). The decreased abundance in the mutants (figure 7.6) as compared to the wild-type suggest that a larger fraction of these ***H*** monomers may have been incorporated into the lignin polymer via stronger crosslinks as a result of the mutation. The relatively unchanged 4-vinyl guaiacol yields may be due to the effects of the mutation cancelling each other out – increased abundance of ***G*** monomers (which should increase the yield of 4-vinyl guaiacol) is offset by the decrease in decomposition due to their being incorporated into the polymer via more thermally resistant linkages. The combined ***G*** monomers derived products (figure 7.7) are not statistically significantly different

between the wild-types and mutants (table 7.5). For both wild-types and mutants, yield of ***S*** derived products is minimal (figure 7.8), but the difference is statistically significant ($p < 0.1$) (table 7.6).

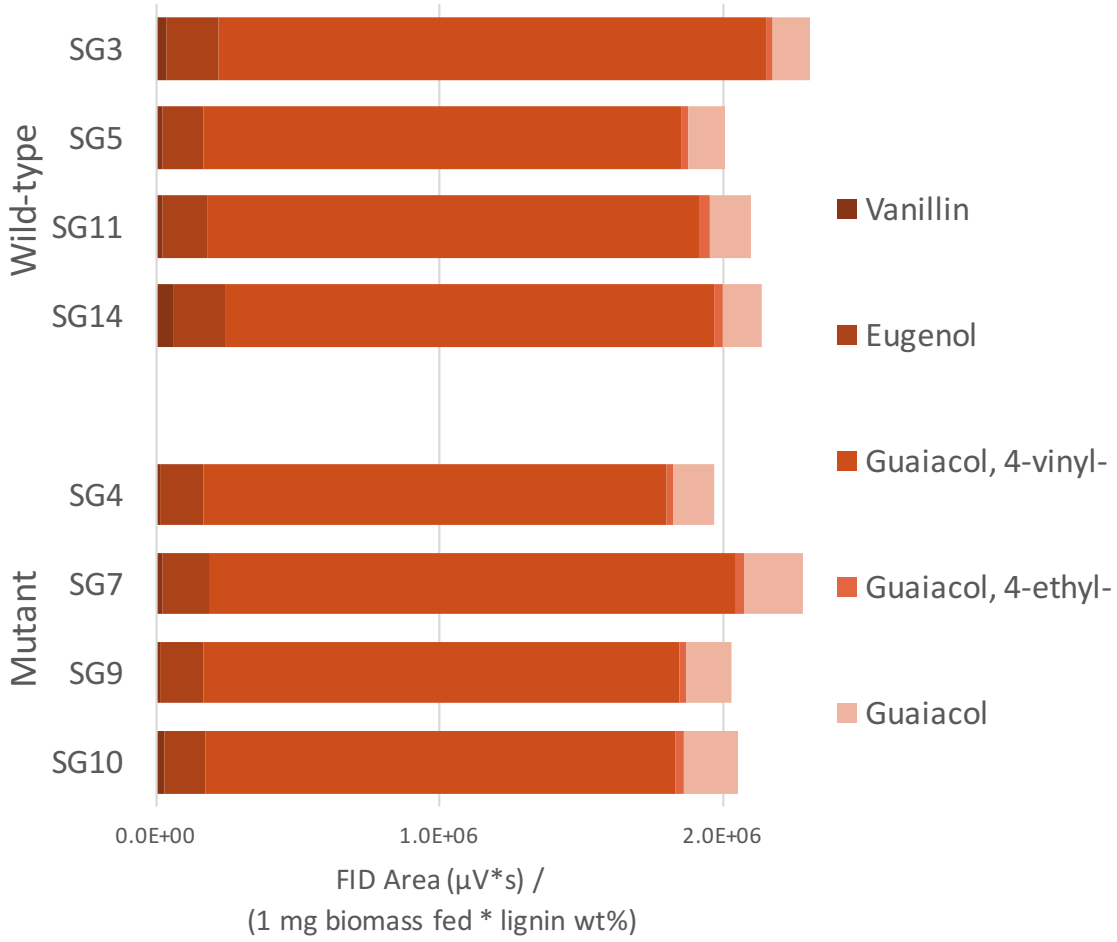


Figure 7.7. Yields of phenolic compounds derived from ***G*** monomers.
Yield values are (compound peak area) / [(mg of raw switchgrass) * (wt % lignin)]. All samples run at 290°C, 120s

	Wild-type	*Mutant*
Mean	2144057.52	2082788.177
Variance	19488378847	18828723648
Observations	4	4
Pearson Correlation	-0.8156	
Hypothesized Mean Difference	0	
df	3	
t Stat	0.4646	
P(T<=t) one-tail	0.3369	
t Critical one-tail	2.3534	
P(T<=t) two-tail	0.6739	
t Critical two-tail	3.1824	

Table 7.5. t-Test: Paired Two Sample for Means – G-unit products only.

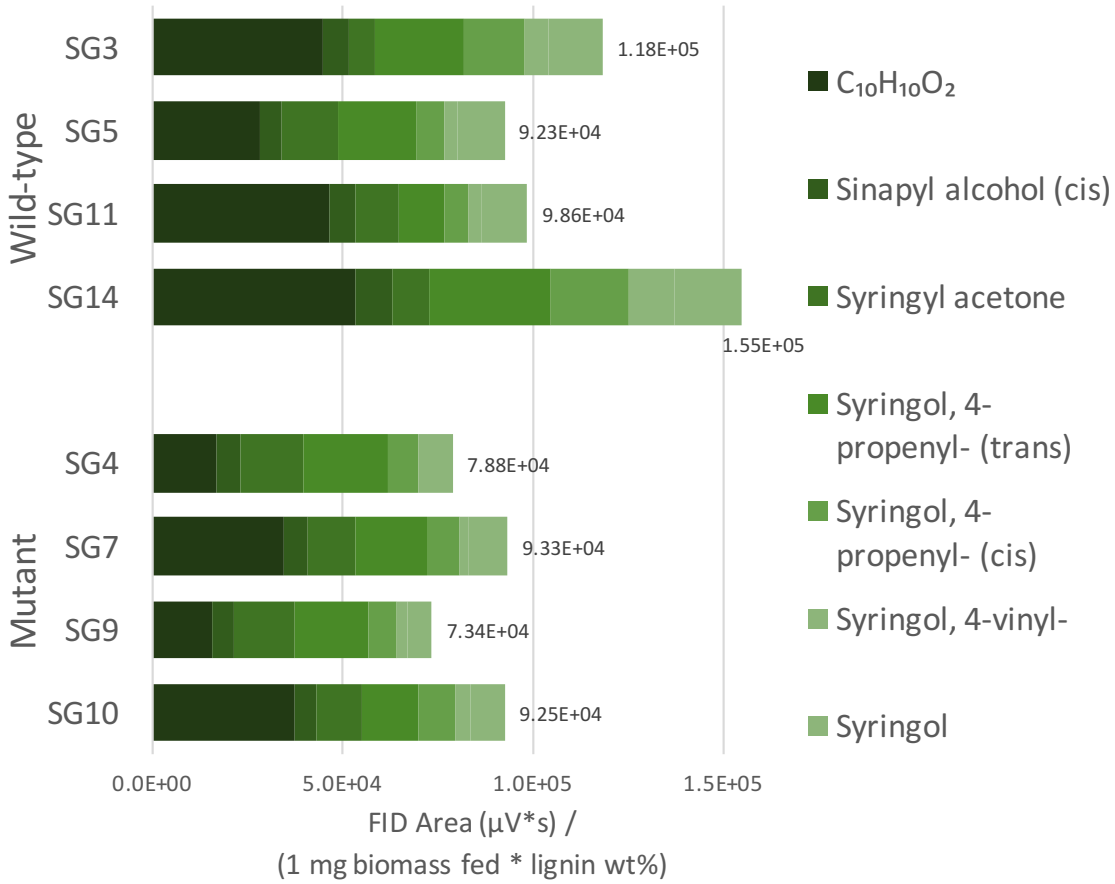


Figure 7.8. Yields of phenolic compounds derived from ***S*** monomers.
Yield values are (compound peak area) / [(mg of raw switchgrass) * (wt % lignin)].
All samples run at 290°C, 120s

	Wild-type	*Mutant*
Mean	115919.7377	84509.69338
Variance	784382489.7	99189479.33
Observations	4	4
Pearson Correlation	0.3368	
Hypothesized Mean Difference	0	
df	3	
t Stat	2.3817	
P(T<=t) one-tail	0.0487	
t Critical one-tail	2.3534	
P(T<=t) two-tail	0.0975	
t Critical two-tail	3.1824	

Table 7.6. t-Test: Paired Two Sample for Means – S-unit products only.

7.4 Discussion

As hypothesized, statistically significant differences are observed in the lignin-normalized phenolics product yields between the COMT mutant samples and the wild-type samples. This could be a result of the decreased abundance of β–O–4 inter-unit linkages in the mutants due to the suppression of ***S*** monomer formation. The phenolic species most responsible for the observed differences is 4-vinyl phenol, a product of ***H*** monomers and the most abundant single phenolic thermochemical product in all of the wild-type samples. The mutants, although enriched in ***G*** monomers compared with the wild-type, did not produce significantly less ***G*** derived phenolic products.

7.4.1 Alternative hypothesis

Another possible explanation for the lower yield of the phenolic products could be lowered levels of p-coumaric acid (pCA) in the mutant switchgrasses resulting from the mutation. Palmer et al. (*189*) observed lower wall-bound pCA levels across two different COMT-deficient sorghum mutants as compared to their wild-types. These lower pCA levels were accompanied by higher levels of ferulic acid (FA). It is possible

that lower pCA levels in the mutants could be responsible for the observed decrease in production of 4-vinyl phenol via torrefaction. However, no increase in 4-vinyl guaiacol (a product of the FA) is observed. This could be indicative that incorporation of the esterified FA is primarily by the more thermally recalcitrant bonds, or simply that these samples do not exhibit higher levels of esterified FA. As no measurement of the esterified hydroxycinnamic acids (HCA) is available for the samples used in this study, this alternative hypothesis remains untested.

While this study shows that the COMT deficient switchgrass may be a favorable mutant for achieving lower phenolic yields in low-temperature torrefaction, some important unanswered questions remain open for investigation, and several are enumerated below:

1. The alternative hypothesis – that lower pCA levels in the mutant are responsible for the observed differences – is untested, as no esterified HCA measurements are available for these samples. If pCA levels are no different between the wild-types and mutants, then the observed differences are more likely due to enhanced thermal stability of the lignin polymer backbone, or incorporation of the esterified ferulates into the lignin via the more thermally resistant cross-linkages (β–5, 5-5, etc.). Additionally, pre-treatment of these samples to remove HCA esters before torrefaction would help dis-aggregate the impacts of the HCA on the observed phenolics products.
2. The effect of the mutation on the subsequent fast pyrolysis of these samples is untested and therefore unknown. If the total phenolic yield across both torrefaction and subsequent pyrolysis is much lower for the mutant switchgrasses, regardless of the underlying cause, the question of suitability of

this mutation for a staged thermal fractionation process becomes more complex, for similar reasons as discussed in section 5.5.

3. Lignin biosynthesis literature indicates that the decreased ***S***/***G*** ratio should result in fewer β–O–4 bonds. However, directly quantifying the inter-unit linkages in the samples tested here would serve as a more direct confirmation of this assumption. The methodology for this quantification is extant in the literature and involves NMR characterization of the lignin (*190-192*). Adding this supporting data will strengthen the case made here.
4. As the differences in phenolic yields are ascribed to differences in bond dissociation enthalpies, it is reasonable to conclude that these differences may become less apparent or vanish altogether with increasing temperatures at which the thermal degradation is performed (the inverse may also be true – the differences in yields may increase with even lower temperatures). It may be of interest to establish an upper temperature bound at which these differences cease to become statistically significant to aid in optimization decision-making.

7.5 Acknowledgements

The work has been financially supported by the NIFA-USDA Sun Grant Initiative (AB-5-68110.OU1), the National Science Foundation (EPSCoR 0814361), and the US Department of Energy (DE-FG36GO88064). Much thanks to the Stewart group for the provision of the switchgrasses and their compositional data as used in this study.

Chapter 8: Looking ahead

8.1 Further optimization of staged thermal fractionation

8.1.1 Stage 1 optimization

Chapter 5 has dealt with the optimization of the thermal segregation of the cellulose and lignin components of the biomass. While there is evidence that the first stage thermal treatment used in that work was effective at removing the hemicellulose portion of the biomass, establishing the optimal conditions for doing this important step was not explored. In the context of the staged thermal fractionation strategy as laid out chapters 1 and 5, an optimal stage 1 separation should achieve complete degradation of the hemicellulose with minimal lignin degradation and at conditions which are not detrimental to the total process yield of cellulose-derived products (primarily stage 3 levoglucosan yields), and at conditions which will minimize the required energy input of the stage 1 torrefaction process (i.e., the lowest temperature and shortest time at which these objectives can be achieved).

Starting at the optimal stage 2 conditions as established in chapter 5, untreated red oak biomass (the same as used in chapters 4 & 5) should be sequentially torrefied and pyrolyzed at successively decreasing torrefaction temperatures until a maximal level of levoglucosan production in the pyrolysis step is achieved. This establishes the highest temperature possible for the process which will prevent the condensation and repolymerization reactions responsible for loss of yield as discussed in section 5.4.2. The final hold time for the torrefaction process should be long enough to ensure no thermal gradient is present in the biomass particles; 5 minutes is suggested but some simple heat transfer calculations may provide a better estimate.

Following the establishment of the maximum stage 1 temperature, the final stage 1 hold time should be successively reduced until the production of acetic acid in the successive pyrolysis step begins to increase. While some acetic acid is produced at higher temperatures (likely from lignin, as discussed in section 5.4), increased acetic acid production at the pyrolysis stage signals the incomplete conversion of hemicellulose. The highest possible temperature at the shortest possible time to achieve breakdown of the hemicellulose with no accompanying cellulose degradation should also minimize the amount of lignin-derived phenolics products into the first stage thermal products stream.

By determining an optimal stage 1 temperature in this manner and combining with the optimal stage 2 conditions as established in chapter 5, a best case for purely thermal segregation is established which can then be used as a basis of comparison for other segregation strategies (e.g., pre- and post-treatments).

8.1.2 Optimization with other feedstocks

All of the optimization of stage 1 and stage 2 as performed and described up to this point have used red oak as the feedstock. As discussed in section 5.5, oak is unlikely to ever be a major biofuel feedstock and as such it is suggested that the research group at the University of Oklahoma establish a better standard feedstock for the thermochemical biofuels research efforts. Switchgrass would be an obvious choice as the standard feedstock for the research group given its favorable properties as described in section 6.3 and the degree to which we understand the thermochemical product spectrum as a result of the efforts in chapters 6 and 7. Other suggested feedstocks are poplar, one of the only hardwoods in serious consideration as a biofuels feedstock due

to its rapid growth rates (*2*, *193*); eastern red cedar (*Juniperus virginiana*), a pioneer invader species which is destructive to grasslands and could be a candidate for blending with prairie grass feedstocks (*194*); and any of the southern yellow pines, as their extensive current use in the forest products industries provides readily available residues for biofuels production (*193*).

Any change in feedstock will necessitate re-optimization of the thermal stage conditions due to differences in the hemicellulose and lignin compositions. The stage 1 optimization as outlined above will need to be repeated to establish the final hold time, as the temperature is optimized not for the hemicellulose decomposition but rather to avoid any cellulose decomposition. The stage 2 optimization conditions should be re-examined as well, as the differences in lignin composition may change the distribution of the phenolics between stages 2 and 3. For example, oak (a woody dicot) contains ***H***, ***G***, and ***S*** lignins (section 2.1, (*55*)); cedar or pine (woody gymnosperms) do not produce significant amounts of ***S*** lignin and may have more thermally stable lignin as compared with oak for the reasons discussed in chapter 7.

8.2 Further investigation into biomass compositional effects

8.2.1 Direct analysis of NIR spectra

The modeling approach used in chapter 6 relies on compositional data which itself is derived from several models fitted to predict compositional features from near-infrared spectroscopy. At the time of this writing the underlying spectroscopic data remains unavailable; however, when it is obtained, it is suggested that the modeling methodology be re-factored to improve the predictive ability of the models developed. The spectral data for the 22 thermochemically characterized samples can be used to

develop a ‘black box’ predictive model using principal components analysis pre-processing; the application of a model with tunable parameters (tuned via cross validation) is suggested (e.g., LASSO (*195*), or elastic net (*196*)) to improve their predictive ability. In this manner, it should be possible to develop models which have good predictive ability for all of the thermochemical lumps, expanding to including models which had poor predictive ability and those which were rejected due to uninterpretable coefficients using the stepwise OLS approach as applied in chapter 6. Once these ‘black box’ models are generated from the 22 characterized samples and their spectral data, the complete association panel’s spectral data can then be used to predict the entire thermochemical product distribution, beyond just the three or four lumps that were predicted in chapter 7. This predicted data can then be combined with the available NIRS compositional data to develop interpretable models in the same fashion as the work in chapter 6, except that due to the much larger number of observations, the feature selection algorithms used can be discarded. Additionally, with a much larger set of observations, interactions between thermochemical lumps and compositional features can be modeled. Finally, having the predicted thermochemical data for the full association panel will allow for the first attempt at mapping trends in the product distribution to genotypes. In short, this revised approach which will be possible once the spectral data is acquired will allow for more insights and hypothesis formation than what has been presented in this work.

8.3 Improving data collection

8.3.1 Optimization of the GC program

The program used for the GC temperature ramp is detailed in section 3.2.3. The total time of this GC program is roughly 99 minutes. As shown below in figure 8.1, it should be possible to reduce this program time by as much as 25 minutes without losing any data relevant to the research. Peaks after 75 minutes are either too small to matter or consist of contamination which has been introduced into the system (e.g., skin oils from careless handling). Savings of 25 minutes per analysis over the 132 analyses done in chapter 6 results in a net savings of 55 hours of instrument time.

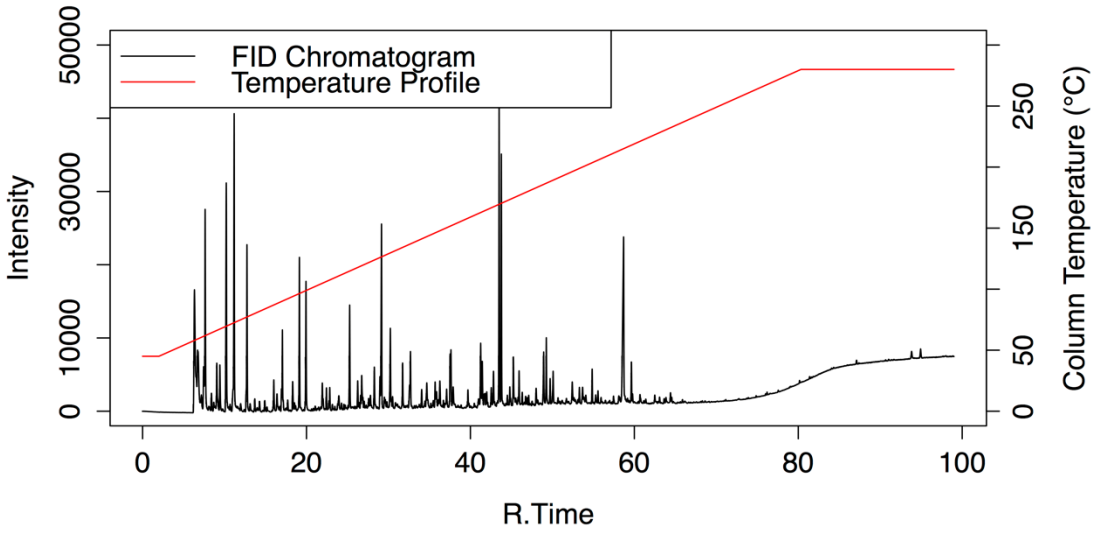


Figure 8.1. Typical biomass pyrolysis chromatogram.

Additionally, it may be possible to optimize the temperature profile to achieve better separation of peaks and improve data quality. Figure 8.2 shows a subsection of the chromatogram in figure 8.1, indicating numerous overlapping peaks. In areas of the chromatogram with many overlapping peaks, slowing down the temperature ramp or

introducing some isothermal sections into the program may assist in reducing the overlap and allow for better identification.

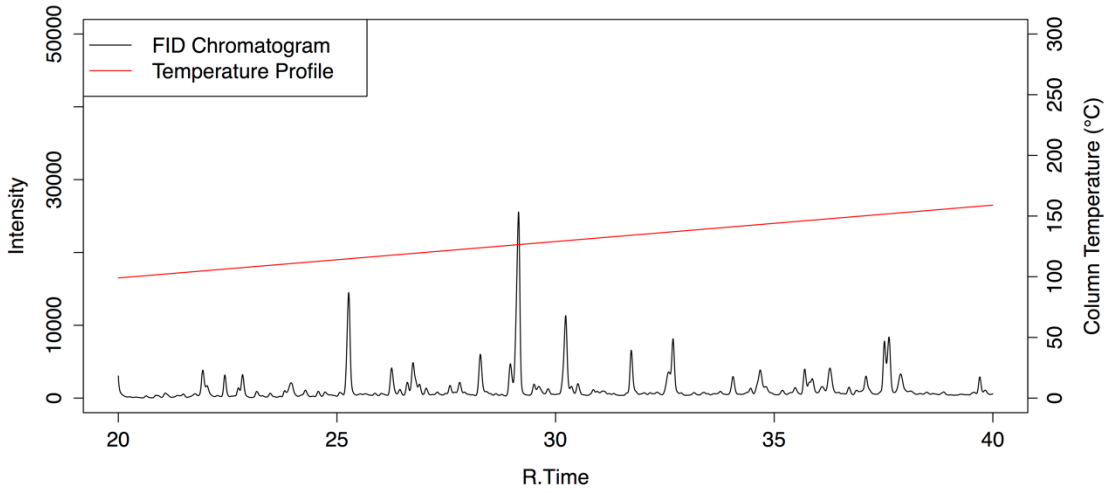


Figure 8.2. Subset of biomass pyrolysis chromatogram showing overlapping peaks.

8.3.2 Reducing or eliminating the need for technical replicates in the pyroprobe

For the studies in chapters 5, 6, and 7, all data points as reported are the mean of three observations, with error estimates as appropriate. As discussed in section 3.2.5, yield data consists of the integrated peak areas of the chromatography, summed by lump and normalized to the amount of the biomass in each sample tube. However, by dividing each compound's peak area by the sum of the total peak areas of the chromatogram (or, after applying a carbon conversion as discussed in section 3.2.6) the thermochemical product selectivity of each individual experiment can be measured without introducing another source of error (the measurement of the mass of the tube). For the torrefied switchgrass samples from chapter 7, a comparison of the error estimates (95% confidence interval of the mean) as a proportion of the mean value between yield and selectivity are shown for each compound lump in figure 8.3. For many of the compound

groups tracked, the selectivity error is smaller than the yield error for the same data, and for many measurements, the size of the error interval is close to 0% of the mean.

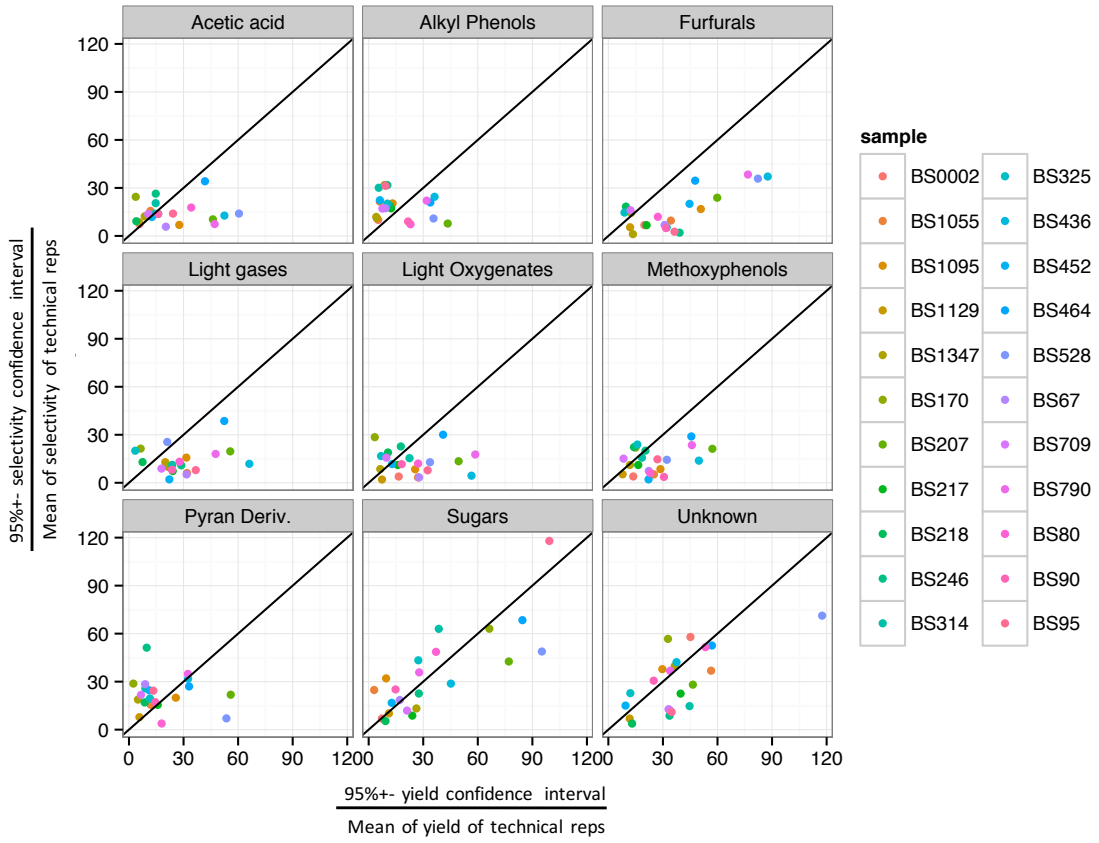


Figure 8.3. Comparison of switchgrass torrefaction error estimates for yield and selectivity.
The line through each plot has intercept = 0 and slope = 1.

While the error estimates of the selectivity are demonstrably improved over those of the yields, this is of little use as the knowledge of the yields is most desirable. If the total mass loss of a sample due to thermal decomposition was determinable using either available instruments (e.g., TGA) or a new instrument acquisition, this data could be combined with selectivity measurements from the pyroprobe, which when used with compound calibration could be combined to get total compound yields. By combining historical data from the pyroprobe apparatus, standard selectivity errors as a percentage of each compound lump could be applied to future experiments. By obtaining yield data

from mass loss experiments and limiting the pyroprobe to selectivity measurements, the need for technical replicates in the pyroprobe would be eliminated, increasing throughput by up to 300%. Alternatively, careful design of experiments which would rely solely on selectivity metrics could leverage these same savings.

8.4 Further investigation into catalytic upgrading approaches and requirements

8.4.1 Switchgrass variants for catalyst deactivation experiments

In chapter 4, the effects of catalyst & process parameters on the rates of zeolite catalyst deactivation during biomass upgrading were explored. For all experiments in that work, the starting biomass and thermal conversion conditions were unchanged. In light of the results from chapters 5-7, it is suggested that investigations into the catalyst performance (especially deactivation rates) be performed in order to demonstrate the degree to which different thermochemical product distributions can impact these important process parameters. One specific example would be the rates of acidic zeolite deactivation upon reaction with a torrefaction product stream from a COMT mutant switchgrass vs. a wild-type. Mullen et al. (*197*) suggests that higher levels of alkyl phenols can lead to increased coke formation on acidic zeolites; if that is the case, a product stream with lower alkyl phenol yields (as in the mutant) should improve catalyst performance, and tying this directly to a specific feedstock genetic modification would be a first.

8.4.2 Torrefaction liquid stability exploration

The discussion of the study in chapter 5 largely focused on the implications for catalytic upgrading of the process streams from staged thermal fractionation. However, an often-cited problem with bio-oil liquids is the high degree of chemical instability which leads

to self-polymerization, making efficient processing a challenging engineering problem. Due to the segregation of much of the acetic acid into the first stage product stream, the optimized staged thermal fractionation product streams as described in chapter 5 may exhibit significantly improved physical properties related to storage, transportation, and handling as compared with fast pyrolysis bio-oil. While the pyroprobe apparatus as described in chapter 3.2 is unsuitable for generation of liquid samples, the micropyrolysis unit as described in chapter 3.1 has previously been used to generate liquids which were evaluated for some of these properties including viscosity and molecular weight increase upon aging (*135*). Similar studies evaluating these properties for the optimized staged product streams are suggested.

References

1. A. V. Bridgwater, D. Meier, D. Radlein, An overview of fast pyrolysis of biomass. *Organic Geochemistry* (1999).

2. R. D. Perlack *et al.*, "Biomass as feedstock for a bioenergy and bioproducts industry: The technical feasibility of a billion-ton annual supply" (2005),, doi:10.2172/1216415.

3. NSF, "Breaking the Chemical and Engineering Barriers to Lignocellulosic Biofuels: Next Generation Hydroccarbon Biorefineries" (Washington, DC, 2008),, doi:10.2172/1218335.

4. C. Somerville, Biofuels. *Current Biology*. **17**, R115–R119 (2007).

5. C. Somerville *et al.*, Toward a Systems Approach to Understanding Plant Cell Walls. *Science*. **306**, 2206–2211 (2004).

6. G. W. Huber, S. Iborra, A. Corma, Synthesis of transportation fuels from biomass: chemistry, catalysts, and engineering. *Chem. Rev.* (2006).

7. M. S. Mettler, D. G. Vlachos, P. J. Dauenhauer, Top ten fundamental challenges of biomass pyrolysis for biofuels. *Energy Environ. Sci.* **5**, 7797–13 (2012).

8. O. Faix, I. Fortmann, J. Bremer, D. Meier, *Thermal degradation products of wood* (1991), vol. 49.

9. O. Faix, I. Fortmann, J. Bremer, D. Meier, Thermal degradation products of wood. *Holz als Roh-und Werkstoff*. **49**, 213–219 (1991).

10. A. Oasmaa, S. Czernik, Fuel oil quality of biomass pyrolysis oils state of the art for the end users. *Energy & Fuels* (1999), doi:10.1021/ef980272b.

11. A. Demirbas, Competitive liquid biofuels from biomass. *APPLIED ENERGY*. **88**, 17–28 (2011).

12. P. R. Patwardhan, R. C. Brown, B. H. Shanks, Understanding the Fast Pyrolysis of Lignin. *ChemSusChem*. **4**, 1629–1636 (2011).

13. J. P. Diebold, "A review of the chemical and physical mechanisms of the storage stability of fast pyrolysis bio-oils" (1999),, doi:10.2172/753818.

14. E. Furimsky, *Catalytic hydrodeoxygenation* (2000).

15. J. Meng, J. Park, D. Tilotta, S. Park, The effect of torrefaction on the chemistry of fast-pyrolysis bio-oil. *Bioresource technology*. **111**, 439–446 (2012).

16. T. N. Pham, D. Shi, D. E. Resasco, Evaluating strategies for catalytic upgrading of pyrolysis oil in liquid phase. *Applied Catalysis B: Environmental* (2014), doi:10.1016/j.apcatb.2013.01.002.

17. D. E. Resasco, What Should We Demand from the Catalysts Responsible for Upgrading Biomass Pyrolysis Oil? *J. Phys. Chem. Lett.* **2**, 2294–2295 (2011).

18. A. V. Bridgwater, Catalysis in thermal biomass conversion. *Applied Catalysis A: General*. **116**, 5–47 (1994).

19. T. R. Carlson, T. P. Vispute, G. W. Huber, Green Gasoline by Catalytic Fast Pyrolysis of Solid Biomass Derived Compounds. *ChemSusChem*. **1**, 397–400 (2008).

20. T. R. Carlson, G. A. Tompsett, W. C. Conner, G. W. Huber, Aromatic Production from Catalytic Fast Pyrolysis of Biomass-Derived Feedstocks. *Top Catal*. **52**, 241–252 (2009).

21. S. Wan *et al.*, Decoupling HZSM-5 Catalyst Activity from Deactivation during Upgrading of Pyrolysis Oil Vapors. *ChemSusChem*. **8**, 552–559 (2014).

22. D. E. Resasco, S. P. Crossley, Implementation of concepts derived from model compound studies in the separation and conversion of bio-oil to fuel. *Catalysis Today*. **257**, 185–199 (2015).

23. V. Srinivasan, S. Adhikari, S. A. Chattanathan, S. Park, Catalytic Pyrolysis of Torrefied Biomass for Hydrocarbons Production. *Energy & Fuels*. **26**, 7347–7353 (2012).

24. P. de Wild, H. Reith, E. Heeres, Biomass pyrolysis for chemicals. *Biofuels*. **2**, 185–208 (2011).

25. S. Chang *et al.*, Effect of torrefaction temperature on product distribution from two-staged pyrolysis of biomass. *Energy & Fuels* (2012), doi:10.1021/ef201872y.

26. A. A. Boateng, C. A. Mullen, Fast pyrolysis of biomass thermally pretreated by torrefaction. *Journal of Analytical and Applied Pyrolysis* (2013), doi:10.1016/j.jaap.2012.12.002.

27. H. Yang, R. Yan, H. Chen, D. H. Lee, C. Zheng, Characteristics of hemicellulose, cellulose and lignin pyrolysis. *Fuel*. **86**, 1781–1788 (2007).

28. M. J. Antal Jr., G. Varhegyi, Cellulose Pyrolysis Kinetics: The Current State of Knowledge. *Ind. Eng. Chem. Res.* **34**, 703–717 (1995).

29. G. Varhegyi, M. J. Antal Jr., E. Jakab, P. Szabó, Kinetic modeling of biomass pyrolysis. *Journal of Analytical and Applied Pyrolysis*. **42**, 73–87 (1997).

30. P. R. Patwardhan, R. C. Brown, B. H. Shanks, Product distribution from the fast pyrolysis of hemicellulose. *ChemSusChem* (2011), doi:10.1002/cssc.201000425/full.

31. P. R. Patwardhan, D. L. Dalluge, B. H. Shanks, Distinguishing primary and secondary reactions of cellulose pyrolysis. *Bioresource technology* (2011), doi:10.1016/j.biortech.2011.02.018.

32. D. Mohan, C. U. Pittman, P. H. Steele, Pyrolysis of Wood/Biomass for Bio-oil: A Critical Review. *Energy & Fuels*. **20**, 848–889 (2006).

33. P. Tanger, J. L. Field, C. E. Jahn, M. W. DeFoort, J. E. Leach, Biomass for thermochemical conversion: targets and challenges. *Front. Plant. Sci.* **4** (2013), doi:10.3389/fpls.2013.00218.

34. M. A. ONeill, W. S. York, in *Annual Plant Reviews*, J. K. C. Rose, Ed. (John Wiley & Sons, 2009), p. 400.

35. K. P. Vogel *et al.*, Quantifying Actual and Theoretical Ethanol Yields for Switchgrass Strains Using NIRS Analyses. *Bioenerg. Res.* **4**, 96–110 (2011).

36. P. R. Adler, M. A. Sanderson, A. A. Boateng, P. I. Weimer, H. J. G. Jung, Biomass yield and biofuel quality of switchgrass harvested in fall or spring. *Agron J*. **98**, 1518–1525 (2006).

37. H. M. El-Nashaar, G. M. Banowetz, S. M. Griffith, M. D. Casler, K. P. Vogel, Genotypic variability in mineral composition of switchgrass. *Bioresource technology*. **100**, 1809–1814 (2009).

38. M. P. Singh *et al.*, Mineral composition and biomass partitioning of sweet sorghum grown for bioenergy in the southeastern USA. *Biomass and Bioenergy*. **47**, 1–8 (2012).

39. Y. L. Zhao, Y. Steinberger, M. Shi, Han, L. P., G. H. Xie, Changes in stem composition and harvested produce of sweet sorghum during the period from maturity to a sequence of delayed harvest dates. *Biomass and Bioenergy*. **39**, 261–273 (2012).

40. M. Pauly, K. Keegstra, Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J*. **54**, 559–568 (2008).

41. H. V. Scheller, P. Ulvskov, Hemicelluloses. *Annu Rev Plant Biol*. **61**, 263–289 (2010).

42. M. E. Vega-Sánchez *et al.*, Loss of Cellulose Synthase-Like F6 Function Affects Mixed-Linkage Glucan Deposition, Cell Wall Mechanical Properties, and Defense Responses in Vegetative Tissues of Rice. *Plant Physiol.* **159**, 56–69 (2012).

43. J. Vogel, Unique aspects of the grass cell wall. *Current Opinion in Plant Biology*. **11**, 301–307 (2008).

44. P. McKendry, Energy production from biomass (part 1): overview of biomass. *Bioresource technology*. **83**, 37–46 (2002).

45. C. Somerville, Cellulose Synthesis in Higher Plants. *Annu. Rev. Cell Dev. Biol.* **22**, 53–78 (2006).

46. R. H. Newman, S. J. Hill, P. J. Harris, Wide-Angle X-Ray Scattering and Solid-State Nuclear Magnetic Resonance Data Combined to Test Models for Cellulose Microfibrils in Mung Bean Cell Walls. *Plant Physiol.* **163**, 1558–1567 (2013).

47. F. Shafizadeh, G. D. McGinnis, C. W. Philpot, Thermal degradation of xylan and related model compounds. *Carbohyd Res*. **25**, 23–33 (1972).

48. O. D. Mante, S. P. Babu, T. E. Amidon, A comprehensive study on relating cell-wall components of lignocellulosic biomass to oxygenated species formed during pyrolysis. *Journal of Analytical and Applied Pyrolysis*. **108**, 56–67 (2014).

49. M. E. Vega-Sánchez, Y. Verhertbruggen, H. V. Scheller, P. C. Ronald, Abundance of mixed linkage glucan in mature tissues and secondary cell walls of grasses. *Plant signaling & behavior*. **8**, e23143 (2013).

50. M. D. C. Rodriguez-Gacio, R. Iglesias-Fernandez, P. Carbonero, A. J. Matilla, Softening-up mannan-rich cell walls. *J Exp Bot*. **63**, 3975–3988 (2012).

51. M. Pauly *et al.*, Hemicellulose biosynthesis. *Planta*. **238**, 627–642 (2013).

52. F. Chen, Y. Tobimatsu, D. Havkin-Frenkel, R. A. Dixon, J. Ralph, A polymer of caffeyl alcohol in plant seeds. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1772–1777 (2012).

53. F. Chen *et al.*, Novel seed coat lignins in the Cactaceae: structure, distribution and implications for the evolution of lignin diversity. *The Plant Journal*. **73**, 201–211 (2013).

54. H. Kawamoto, S. Horigoshi, S. Saka, Pyrolysis reactions of various lignin model dimers. *Journal of Wood Science*. **53**, 168–174 (2007).

55. W. Boerjan, J. Ralph, M. Baucher, Lignin biosynthesis. *Annu Rev Plant Biol*. **54**, 519–546 (2003).

56. D. L. Petrik *et al.*, p-Coumaroyl-CoA:monolignol transferase (PMT) acts specifically in the lignin biosynthetic pathway in Brachypodium distachyon. *Plant J*. **77**, 713–726 (2014).

57. F. Lu *et al.*, Naturally p-Hydroxybenzoylated Lignins in Palms. *Bioenerg. Res.*, 1–19 (2015).

58. B. Vanholme *et al.*, Breeding with rare defective alleles (BRDA): a natural Populus nigra HCT mutant with modified lignin as a case study. *New Phytologist*. **198**, 765–776 (2013).

59. K. V. Sarkanen, H. Chang, G. G. Allan, Species Variation in Lignins .3. Hardwood Lignins. *Tappi*. **50**, 587–& (1967).

60. F. Lu, J. Ralph, Preliminary evidence for sinapyl acetate as a lignin monomer in kenaf. *Chemical Communications*, 90–91 (2002).

61. M. V. Bule, A. H. Gao, B. Hiscox, S. Chen, Structural Modification of Lignin and Characterization of Pretreated Wheat Straw by Ozonation. *J. Agric. Food Chem.* **61**, 3916–3925 (2013).

62. R. D. Hatfield, J. M. Marita, K. Frost, Characterization of p-coumarate accumulation, p-coumaroyl transferase, and cell wall changes during the development of corn stems. *J. Sci. Food Agric.* **88**, 2529–2537 (2008).

63. J. Ralph, J. H. Grabber, R. D. Hatfield, Lignin-Ferulate Cross-Links in Grasses - Active Incorporation of Ferulate Polysaccharide Esters into Ryegrass Lignins. *Carbohyd Res*. **275**, 167–178 (1995).

64. W. Lan *et al.*, Tricin, a Flavonoid Monomer in Monocot Lignification. *Plant Physiol.* **167**, 1284–1295 (2015).

65. A. A. Tortosa Masiá, B. J. P. Buhre, R. P. Gupta, T. F. Wall, Characterising ash of biomass and waste. *Fuel Processing Technology*. **88**, 1071–1081 (2007).

66. S. V. Vassilev, D. Baxter, L. K. Andersen, C. G. Vassileva, An overview of the chemical composition of biomass. *Fuel*. **89**, 913–933 (2010).

67. M. Dick-Pérez *et al.*, Structure and Interactions of Plant Cell-Wall Polysaccharides by Two- and Three-Dimensional Magic-Angle-Spinning Solid-State NMR. *Biochemistry*. **50**, 989–1000 (2011).

68. L. Tan *et al.*, An Arabidopsis Cell Wall Proteoglycan Consists of Pectin and Arabinoxylan Covalently Linked to an Arabinogalactan Protein. *The Plant Cell*. **25**, 270–287 (2013).

69. D. Mikkelsen, B. M. Flanagan, S. M. Wilson, A. Bacic, M. J. Gidley, Interactions of Arabinoxylan and (1,3)(1,4)-β-Glucan with Cellulose Networks. *Biomacromolecules*. **16**, 1232–1239 (2015).

70. Mirko Bunzel, John Ralph, Fachuang Lu, A. Ronald D Hatfield, H. Steinhart, Lignins and Ferulate−Coniferyl Alcohol Cross-Coupling Products in Cereal

Grains. *J. Agric. Food Chem.* **52**, 6496–6502 (2004).

71. T.-Q. Yuan, S.-N. Sun, F. Xu, R.-C. Sun, Characterization of lignin structures and lignin-carbohydrate complex (LCC) linkages by quantitative 13C and 2D HSQC NMR spectroscopy. *J. Agric. Food Chem.* **59**, 10604–10614 (2011).

72. X. Du *et al.*, Analysis of lignin-carbohydrate and lignin-lignin linkages after hydrolase treatment of xylan-lignin, glucomannan-lignin and glucan-lignin complexes from spruce wood. *Planta*. **239**, 1079–1090 (2014).

73. N. C. Carpita *et al.*, Cell Wall Architecture of the Elongating Maize Coleoptile. *Plant Physiol.* **127**, 551–565 (2001).

74. L. V. Kozlova, M. V. Ageeva, N. N. Ibragimova, T. A. Gorshkova, Arrangement of mixed-linkage glucan and glucuronoarabinoxylan in the cell walls of growing maize roots. *Annals of botany*. **114**, 1135–1145 (2014).

75. N. Kido *et al.*, The matrix polysaccharide (1;3,1;4)-beta-D-glucan is involved in silicon-dependent strengthening of rice cell wall. *Plant & cell physiology*. **56**, 268–276 (2015).

76. F.-X. Collard, J. Blin, A review on pyrolysis of biomass constituents: Mechanisms and composition of the products obtained from the conversion of cellulose, hemicelluloses and lignin. *Renewable and Sustainable Energy Reviews*. **38**, 594–608 (2014).

77. X. L. Zhang, W. H. Yang, C. Q. Dong, Levoglucosan formation mechanisms during cellulose pyrolysis. *Journal of Analytical and Applied Pyrolysis*. **104**, 19–27 (2013).

78. R. Vanholme, R. Van Acker, W. Boerjan, Potential of Arabidopsis systems biology to advance the biofuel field. *Trends in Biotechnology*. **28**, 543–547 (2010).

79. U. Räisänen, I. Pitkänen, H. Halttunen, Formation of the main degradation compounds from arabinose, xylose, mannose and arabinitol during pyrolysis. *Journal of Thermal Analysis and Calorimetry*. **72**, 481–488 (2003).

80. C. A. Mullen, A. A. Boateng, K. B. Hicks, N. M. Goldberg, R. A. Moreau, Analysis and Comparison of Bio-Oil Produced by Fast Pyrolysis from Three Barley Biomass/Byproduct Streams. *Energy & Fuels*. **24**, 699–706 (2010).

81. C. Sáiz-Jiménez, J. W. De Leeuw, Lignin pyrolysis products: their structures and their significance as biomarkers. *Organic Geochemistry* (1986).

82. M. Li *et al.*, Structural characterization of alkaline hydrogen peroxide pretreated grasses exhibiting diverse lignin phenotypes. *Biotechnol Biofuels*. **5**, 38 (2012).

83. V. Mendu *et al.*, Identification and thermochemical analysis of high-lignin feedstocks for biofuel and biochemical production. *Biotechnol Biofuels*. **4**, 1 (2011).

84. I.-Y. Eom *et al.*, Characterization of primary thermal degradation features of lignocellulosic biomass after removal of inorganic metals by diverse solvents. *Bioresource technology*. **102**, 3437–3444 (2011).

85. R. Lou, S. Wu, G. Lv, A. Zhang, Factors Related to Minerals and Ingredients Influencing the Distribution of Pyrolysates Derived from Herbaceous Biomass. *Bioresources* (2013).

86. F. Ronsse, X. Bai, W. Prins, R. C. Brown, Secondary reactions of levoglucosan and char in the fast pyrolysis of cellulose. *Environ. Prog. Sustainable Energy*. **31**, 256–260 (2012).

87. R. Fahmi *et al.*, The effect of alkali metals on combustion and pyrolysis of Lolium and Festuca grasses, switchgrass and willow. *Fuel*. **86**, 1560–1569 (2007).

88. P. R. Patwardhan, J. A. Satrio, R. C. Brown, B. H. Shanks, Influence of inorganic salts on the primary pyrolysis products of cellulose. *Bioresource technology*. **101**, 4646–4655 (2010).

89. M. Müller-Hagedorn, H. Bockhorn, L. Krebs, U. Müller, A comparative kinetic study on the pyrolysis of three different wood species. *Journal of Analytical and Applied Pyrolysis*. **68-69**, 231–249 (2003).

90. G. R. Ponder, G. N. Richards, T. T. Stevenson, Influence of linkage position and orientation in pyrolysis of polysaccharides: A study of several glucans. *Journal of Analytical and Applied Pyrolysis*. **22**, 217–229 (1992).

91. M. Kleen, G. Gellerstedt, Influence of Inorganic Species on the Formation of Polysaccharide and Lignin Degradation Products in the Analytical Pyrolysis of Pulps. *Journal of Analytical and Applied Pyrolysis*. **35**, 15–41 (1995).

92. J. Zhang *et al.*, Cellulose–Hemicellulose and Cellulose–Lignin Interactions during Fast Pyrolysis. *ACS Sustainable Chem. Eng.* **3**, 293–301 (2015).

93. D. Mourant *et al.*, Mallee wood fast pyrolysis: Effects of alkali and alkaline earth metallic species on the yield and composition of bio-oil. *Fuel*. **90**, 2915–2922 (2011).

94. R. J. M. Westerhof, N. J. M. Kuipers, S. R. A. Kersten, W. P. M. van Swaaij, Controlling the Water Content of Biomass Fast Pyrolysis Oil. *Ind. Eng. Chem. Res.* **46**, 9238–9247 (2007).

95. L. Burhenne, M. Damiani, T. Aicher, Effect of feedstock water content and

pyrolysis temperature on the structure and reactivity of spruce wood char produced in fixed bed pyrolysis. *Fuel*. **107**, 836–847 (2013).

96. R. He, X. P. Ye, B. C. English, J. A. Satrio, Influence of pyrolysis condition on switchgrass bio-oil yield and physicochemical properties. *Bioresource technology*. **100**, 5305–5311 (2009).

97. A. Pictet, J. Sarasin, Sur la distillation de la cellulose et de l'amidon sous pression reduite. *Helvetica Chimica Acta* (1918), doi:10.1002/hlca.19180010109/abstract.

98. A. G. W. Bradbury, Y. Sakai, F. Shafizadeh, A kinetic model for pyrolysis of cellulose. *Journal of Applied Polymer Science*. **23**, 3271–3280 (1979).

99. B. W. Penning *et al.*, Validation of PyMBMS as a High-throughput Screen for Lignin Abundance in Lignocellulosic Biomass of Grasses. *Bioenerg. Res.* **7**, 899–908 (2014).

100. P. R. Patwardhan, J. A. Satrio, R. C. Brown, Influence of inorganic salts on the primary pyrolysis products of cellulose. *Bioresource technology* (2010), doi:10.1016/j.biortech.2010.01.112.

101. I.-Y. Eom *et al.*, Effect of essential inorganic metals on primary thermal degradation of lignocellulosic biomass. *Bioresource technology*. **104**, 687–694 (2012).

102. Z. Jin, K. S. Katsumata, T. B. T. Lam, K. Iiyama, Covalent linkages between cellulose and lignin in cell walls of coniferous and nonconiferous woods. *Biopolymers*. **83**, 103–110 (2006).

103. Y. Zhou, H. Stuart-Williams, G. D. Farquhar, C. H. Hocart, The use of natural abundance stable isotopic ratios to indicate the presence of oxygen-containing chemical linkages between cellulose and lignin in plant cell walls. *Phytochemistry*. **71**, 982–993 (2010).

104. O. D. Mante, S. P. Babu, T. E. Amidon, A comprehensive study on relating cell-wall components of lignocellulosic biomass to oxygenated species formed during pyrolysis. *Journal of Analytical and Applied Pyrolysis*. **108**, 56–67 (2014).

105. J. G. Pohlmann, E. Osório, A. C. F. Vilela, M. A. Diez, A. G. Borrego, *Integrating physicochemical information to follow the transformations of biomass upon torrefaction and low-temperature carbonization* (2014), vol. 131.

106. A. V. Bridgwater, D. Meier, D. Radlein, *An overview of fast pyrolysis of biomass* (1999).

107. K. M. Bryden, K. W. Ragland, C. J. Rutland, Modeling thermally thick

pyrolysis of wood. *Biomass and Bioenergy*. **22**, 41–53 (2002).

108. K. O. Davidsson, J. Pettersson, *Birch wood particle shrinkage during rapid pyrolysis* (2002).

109. D. S. Scott, J. Piskorz, The continuous flash pyrolysis of biomass. *Can. J. Chem. Eng.* **62**, 404–412 (1984).

110. H. S. Fogler, *Elements of chemical reaction engineering* (Prentice-Hall, Upper Saddle River, NJ, ed. 4, 2006).

111. K. B. Hicks, K. P. Vogel, A. A. Boateng, Pyrolysis of switchgrass (Panicum virgatum) harvested at several stages of maturity. *Journal of Analytical and Applied Pyrolysis*. **75**, 55–64 (2006).

112. L. Liu, X. P. Ye, A. R. Womac, S. Sokhansanj, Variability of biomass chemical composition and rapid analysis using FT-NIR techniques. *Carbohydrate Polymers*. **81**, 820–829 (2010).

113. R. Sykes *et al.*, in *Biofuels* (Humana Press, Totowa, NJ, 2009), vol. 581 of *Methods in Molecular Biology*, pp. 169–183.

114. J. L. Wegrzyn *et al.*, Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (Populus trichocarpa, Salicaceae) secondary xylem. *New Phytologist*. **188**, 515–532 (2010).

115. B. W. Penning *et al.*, Genetic Determinants for Enzymatic Digestion of Lignocellulosic Biomass Are Independent of Those for Lignin Abundance in a Maize Recombinant Inbred Population. *Plant Physiol.* **165**, 1475–1487 (2014).

116. L. E. Bartley, P. C. Ronald, Plant and microbial research seeks biofuel production from lignocellulose. *California Agriculture*. **63**, 178–184 (2009).

117. F. Yang *et al.*, Engineering secondary cell wall deposition in plants. *Plant Biotechnol J*. **11**, 325–335 (2013).

118. N. D. Bonawitz, C. Chapple, The Genetics of Lignin Biosynthesis: Connecting Genotype to Phenotype. *Annu. Rev. Genet.* **44**, 337–363 (2010).

119. D. Chiniquy *et al.*, XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17117–17122 (2012).

120. L. E. Bartley *et al.*, Overexpression of a BAHD Acyltransferase, OsAt10, Alters Rice Cell Wall Hydroxycinnamic Acid Content and Saccharification. *Plant Physiol.* **161**, 1615–1633 (2013).

121. A. Schultink, D. Naylor, M. Dama, M. Pauly, The role of the plant-specific

ALTERED XYLOGLUCAN9 protein in Arabidopsis cell wall polysaccharide O-acetylation. *Plant Physiol.* **167**, 1271–1283 (2015).

122. Q. Zhao, R. A. Dixon, Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends in Plant Science*. **16**, 227–233 (2011).

123. J. F. Ma *et al.*, A silicon transporter in rice. *Nature*. **440**, 688–691 (2006).

124. N. Yamaji, J. F. Ma, A Transporter at the Node Responsible for Intervascular Transfer of Silicon in Rice. *Plant Cell*. **21**, 2878–2883 (2009).

125. R. Zhong, Z. H. Ye, Secondary Cell Walls: Biosynthesis, Patterned Deposition and Transcriptional Regulation. *Plant & cell physiology*. **56**, 195–214 (2015).

126. H. L. Baxter *et al.*, Two-year field analysis of reduced recalcitrance transgenic switchgrass. *Plant Biotechnol J*. **12**, 914–924 (2014).

127. H. Baxter *et al.*, Field Evaluation of Transgenic Switchgrass Plants Overexpressing PvMYB4 for Reduced Biomass Recalcitrance. *Bioenerg. Res.*, 1–12 (2015).

128. L. E. Bartley, X. Tao, C. Zhang, H. Nguyen, in *Compendium of Bioenergy Plants Switchgrass* (… of Bioenergy Plants: …, 2014), p. 109.

129. C. Lee, Q. Teng, R. Zhong, Z.-H. Ye, The four Arabidopsis reduced wall acetylation genes are expressed in secondary wall-containing cells and required for the acetylation of xylan. *Plant & cell physiology*. **52**, 1289–1301 (2011).

130. G. Xiong, K. Cheng, M. Pauly, Xylan O-Acetylation Impacts Xylem Development and Enzymatic Recalcitrance as Indicated by the Arabidopsis Mutant tbl29. *Molecular Plant*. **6**, 1373–1375 (2013).

131. A. Zheng *et al.*, Effect of torrefaction on structure and fast pyrolysis behavior of corncobs. *Bioresource technology*. **128**, 370–377 (2013).

132. S. W. Banks, D. J. Nowakowski, A. V. Bridgwater, Fast pyrolysis processing of surfactant washed Miscanthus. *Fuel Processing Technology*. **128**, 94–103 (2014).

133. M. M. I. Sheikh *et al.*, Effect of torrefaction for the pretreatment of rice straw for ethanol production. *J. Sci. Food Agric.* **93**, 3198–3204 (2013).

134. J. A. Cunha *et al.*, Waste biomass to liquids: Low temperature conversion of sugarcane bagasse to bio-oil. The effect of combined hydrolysis treatments. *Biomass and Bioenergy*. **35**, 2106–2116 (2011).

135. S. Wan *et al.*, Direct catalytic upgrading of biomass pyrolysis vapors by a dual function Ru/TiO2 catalyst. *AIChE J.* (2013), doi:10.1002/aic.14038.

136. P. J. Dauenhauer, A. D. Paulsen, M. S. Mettler, The Role of Sample Dimension and Temperature in Cellulose Pyrolysis. *Energy & Fuels*. **27**, 2126–2134 (2013).

137. N. Duong, thesis, Norman, OK (2014).

138. A. Romualdo S Fukushima, R. D. Hatfield, Extraction and Isolation of Lignin for Utilization as a Standard to Determine Lignin Concentration Using the Acetyl Bromide Spectrophotometric Method. *J. Agric. Food Chem.* **49**, 3133–3139 (2001).

139. N. Santoro *et al.*, A High-Throughput Platform for Screening Milligram Quantities of Plant Biomass for Lignocellulose Digestibility. *Bioenerg. Res.* **3**, 93–102 (2010).

140. K. P. Vogel *et al.*, Quantifying Actual and Theoretical Ethanol Yields for Switchgrass Strains Using NIRS Analyses. *Bioenerg. Res.* **4**, 96–110 (2010).

141. J. A. Guretzky, J. T. Biermacher, B. J. Cook, M. K. Kering, J. Mosali, Switchgrass for forage and bioenergy: harvest and nitrogen rate effects on biomass yields and nutrient composition. *Plant Soil*. **339**, 69–81 (2010).

142. G. W. Huber *et al.*, Catalytic pyrolysis of solid biomass and related biofuels, Armoatic, and Olefin Compounds.

143. F. A. Agblevor, S. Beis, O. Mante, N. Abdoulmoumine, Fractional Catalytic Pyrolysis of Hybrid Poplar Wood. *Ind. Eng. Chem. Res.* **49**, 3533–3538 (2010).

144. A. Corma, G. HUBER, L. SAUVANAUD, P. OCONNOR, Processing biomass-derived oxygenates in the oil refinery: Catalytic cracking (FCC) reaction pathways and role of catalyst. *Journal of Catalysis*. **247**, 307–327 (2007).

145. H. Zhang, Y.-T. Cheng, T. P. Vispute, R. Xiao, G. W. Huber, Catalytic conversion of biomass-derived feedstocks into olefins and aromatics with ZSM-5: the hydrogen to carbon effective ratio. *Energy Environ. Sci.* **4**, 2297 (2011).

146. P. T. Williams, N. Nugranad, Comparison of products from the pyrolysis and catalytic pyrolysis of rice husks. *Energy*. **25**, 493–513 (2000).

147. P. A. Horne, P. T. Williams, The effect of zeolite ZSM-5 catalyst deactivation during the upgrading of biomass-derived pyrolysis vapours. *Journal of Analytical and Applied Pyrolysis*. **34**, 65–85 (1995).

148. E. F. Iliopoulou, S. D. Stefanidis, Catalytic upgrading of biomass pyrolysis vapors using transition metal-modified ZSM-5 zeolite. *Applied Catalysis B: Environmental*. **127**, 281–290 (2012).

149. J. Jae *et al.*, Investigation into the shape selectivity of zeolite catalysts for biomass conversion. *Journal of Catalysis*. **279**, 257–268 (2011).

150. A. N. Mlinar, P. M. Zimmerman, F. E. Celik, M. Head-Gordon, A. T. Bell, Effects of Brønsted-acid site proximity on the oligomerization of propene in H-MFI. *Journal of Catalysis*. **288**, 65–73 (2012).

151. T. R. Carlson, Y.-T. T. Cheng, J. Jae, G. W. Huber, Production of green aromatics and olefins by catalytic fast pyrolysis of wood sawdust - Energy & Environmental Science (RSC Publishing) DOI:10.1039/C0EE00341G. *Energy Environ. Sci.* (2011).

152. D. J. Mihalcik, C. A. Mullen, A. A. Boateng, Screening acidic zeolites for catalytic fast pyrolysis of biomass and its components. *Journal of Analytical and Applied Pyrolysis*. **92**, 224–232 (2011).

153. C. Mukarakate *et al.*, Real-time monitoring of the deactivation of HZSM-5 during upgrading of pine pyrolysis vapors. *Green Chem.* **16**, 1444–1461 (2014).

154. C. Pereira, R. J. Gorte, Method for distinguishing Brønsted-acid sites in mixtures of H-ZSM-5, H-Y and silica-alumina. *Applied Catalysis A: General*. **90**, 145–157 (1992).

155. J. Bedard, H. Chiang, A. Bhan, Kinetics and mechanism of acetic acid esterification with ethanol on zeolites. *Journal of Catalysis*. **290**, 210–219 (2012).

156. T. Q. Hoang, X. Zhu, T. Sooknoi, D. E. Resasco, R. G. Mallinson, A comparison of the reactivities of propanal and propylene on HZSM-5. *Journal of Catalysis*. **271**, 201–208 (2010).

157. D. J. Parrillo, C. Lee, R. J. Gorte, Heats of adsorption for ammonia and pyridine in H-ZSM-5: evidence for identical Brønsted-acid sites. *Applied Catalysis A: General*. **110**, 67–74 (1994).

158. D. Freude, M. Hunger, H. Pfeifer, W. Schwieger, 1H MAS NMR studies on the acidity of zeolites. *Chemical Physics Letters*. **128**, 62–66 (1986).

159. P. Sazama *et al.*, Effect of aluminium distribution in the framework of ZSM-5 on hydrocarbon transformation. Cracking of 1-butene. *Journal of Catalysis*. **254**, 180–189 (2008).

160. U. Olsbye *et al.*, Conversion of methanol to hydrocarbons: how zeolite cavity and pore size controls product selectivity. *Angew. Chem. Int. Ed.* **51**, 5810–5831 (2012).

161. R. Gounder, E. Iglesia, Catalytic Consequences of Spatial Constraints and Acid Site Location for Monomolecular Alkane Activation on Zeolites. *J. Am. Chem.*

Soc. **131**, 1958–1971 (2009).

162. T. R. Carlson, J. Jae, Y.-C. Lin, G. A. Tompsett, G. W. Huber, Catalytic fast pyrolysis of glucose with HZSM-5: The combined homogeneous and heterogeneous reactions. *Journal of Catalysis*. **270**, 110–124 (2010).

163. W. H. Chen, P. C. Kuo, Torrefaction and co-torrefaction characterization of hemicellulose, cellulose and lignin as well as torrefaction of some basic constituents in biomass. *Energy* (2011), doi:10.1016/j.energy.2010.12.036.

164. W. H. Chen, P. C. Kuo, Isothermal torrefaction kinetics of hemicellulose, cellulose, lignin and xylan using thermogravimetric analysis. *Energy* (2011), doi:10.1016/j.energy.2011.09.022.

165. P. A. Horne, P. T. Williams, Influence of temperature on the products from the flash pyrolysis of biomass. *Fuel* (1996).

166. W. J. Grigsby, G. Varhegyi, C. Di Blasi, Thermogravimetric analysis and devolatilization kinetics of wood. *Ind. Eng. Chem. Res.* (2002), doi:10.1016/j.carbpol.2014.01.086.

167. A. Broido, M. A. Nelson, Char yield on pyrolysis of cellulose. *Combustion and Flame* (1975).

168. T. Melkior *et al.*, NMR analysis of the transformation of wood constituents by torrefaction. *Fuel*. **92**, 271–280 (2012).

169. N. Baccile *et al.*, Structural Characterization of Hydrothermal Carbon Spheres by Advanced Solid-State MAS 13C NMR Investigations. *J. Phys. Chem. C*. **113**, 9644–9654 (2009).

170. Y.-C. Lin, J. Cho, G. A. Tompsett, P. R. Westmoreland, G. W. Huber, Kinetics and Mechanism of Cellulose Pyrolysis (2009).

171. J.-L. Wen, S.-L. Sun, T.-Q. Yuan, F. Xu, R.-C. Sun, Understanding the chemical and structural transformations of lignin macromolecule during torrefaction. *APPLIED ENERGY*. **121**, 1–9 (2014).

172. S. B. McLaughlin, L. Adams Kszos, Development of switchgrass (panicum virgatum) as a bioenergy feedstock in the United States. *Biomass and Bioenergy*. **28**, 515–535 (2005).

173. M. G. Ward, J. K. Ward, B. E. Anderson, Grazing selectivity and in vivo digestibility of switchgrass strains selected for differing digestibility. *J. Anim Sci.* (1989).

174. C. R. Krueger, D. C. Curtis, Evaluation of big bluestem, indiangrass, sideoats grama, and switchgrass pastures with yearling steers. *Agron J* (1979).

175. J. H. Fike *et al.*, Long-term yield potential of switchgrass-for-biofuel systems. *Biomass and Bioenergy*. **30**, 198–206 (2006).

176. C. L. Porter Jr, An analysis of variation between upland and lowland switchgrass, Panicum virgatum L., in central Oklahoma. *Ecology* (1966).

177. M. Kuhn, K. Johnson, *Applied predictive modeling* (New York : Springer, 2013).

178. R Core Team, R: A language and environment for statistical computing, (available at https://www.R-project.org/).

179. M. Kuhn, Caret package. *Journal of Statistical Software* (2008).

180. L. Breiman, Random Forests. *Machine Learning*. **45**, 5–32 (2001).

181. A. Liaw, M. Wiener, Classification and regression by randomForest. *R news* (2002).

182. T. Hastie, D. Pregibon, "Shrinking trees" (AT&T Bell Laboratories, 1990).

183. J. Fox, S. Weisberg, *An R Companion to Applied Regression* (Sage, Thousand Oaks, CA, 2011; http://socserv.socsci.mcmaster.ca/jfox/Books/Companion).

184. P. R. Patwardhan, J. A. Satrio, R. C. Brown, B. H. Shanks, Product distribution from fast pyrolysis of glucose-based carbohydrates. *Journal of Analytical and Applied Pyrolysis*. **86**, 323–330 (2009).

185. K. Raveendran, A. Ganesh, K. C. Khilar, Influence of mineral matter on biomass pyrolysis characteristics. *Fuel*. **74**, 1812–1822 (1995).

186. S. Kim *et al.*, Computational Study of Bond Dissociation Enthalpies for a Large Range of Native and Modified Lignins. *J. Phys. Chem. Lett.* **2**, 2846–2852 (2011).

187. R. Parthasarathi, R. A. Romero, A. Redondo, S. Gnanakaran, Theoretical Study of the Remarkably Diverse Linkages in Lignin. *J. Phys. Chem. Lett.* **2**, 2660–2666 (2011).

188. T. Elder, Bond Dissociation Enthalpies of a Pinoresinol Lignin Model Compound. *Energy & Fuels* (2014), doi:10.1021/ef402310h.

189. N. A. Palmer *et al.*, Genetic background impacts soluble and cell wall-bound aromatics in brown midrib mutants of sorghum. *Planta*. **229**, 115–127 (2008).

190. K. M. Holtman, H. M. Chang, H. Jameel, J. F. Kadla, Quantitative 13C NMR Characterization of Milled Wood Lignins Isolated by Different Milling Techniques. *Journal of Wood Chemistry and Technology*. **26**, 21–34 (2006).

191. Ewellyn A Capanema, A. Mikhail Y Balakshin, J. F. Kadla, A Comprehensive Approach for Quantitative Lignin Characterization by NMR Spectroscopy. *J. Agric. Food Chem.* **52**, 1850–1860 (2004).

192. Ewellyn A Capanema, A. Mikhail Yu Balakshin, J. F. Kadla, Quantitative Characterization of a Hardwood Milled Wood Lignin by Nuclear Magnetic Resonance Spectroscopy. *J. Agric. Food Chem.* **53**, 9639–9649 (2005).

193. M. Hinchee *et al.*, Short-rotation woody crops for bioenergy and biofuels applications. *In Vitro Cell.Dev.Biol.-Plant*. **45**, 619–629 (2009).

194. Z. Barth, Invasion of the Eastern Red Cedar. *Rangelands* (2002), doi:10.2307/4001794.

195. R. Tibshirani, Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (...* (1996), doi:10.2307/2346178.

196. H. Zou, T. Hastie, Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. **67**, 301–320 (2005).

197. C. A. Mullen, A. A. Boateng, Catalytic pyrolysis-GC/MS of lignin from several sources. *Fuel Processing Technology*. **91**, 1446–1458 (2010).

Appendix A: Source code for implementation of torrefaction weight loss model

```
## Implementation of oak weight loss model as described in

## W. J. Grigsby, G. Varhegyi, C. Di Blasi,

## Thermogravimetric analysis and devolatilization kinetics of wood.

## Industrial & Engineering Chemistry Research (2002),

## doi:10.1021/ie0201157.

##

## Last Modified March 17, 2016

## Christopher Waters

## cwaters@ou.edu

## University of Oklahoma

## College of Engineering

## School of Chemical, Biological, and Materials Engineering

## Center for Biofuels Refining

## Pyrolysis Lab Group

## Script for the R programming language



## Output 10 digits when printing values

options(digits=10)
```

```
## Global constants

R <- (8.314 * 10^-3) # kJ/mol * K

t_0 <- 0 # Initial time, seconds; do not change this


# Wt % of original biomass

c_ext <- 0.00 #extractives

c_hc <- 0.3024 #hemicellulose

c_c <- 0.277 #cellulose

c_L <- 0.2267 #lignin


## Kinetic constants, found in Table 3

## Extractives

A_ext <- 10^10.05 #s^-1

E_ext <- 127 #kJ/mol


## Hemicellulose

A_hc <- 10^6.84 #s^-1

E_hc <- 100 #kJ/mol


## Cellulose

A_c <- 10^17.97 #s^-1
```

```
E_c <- 236 #kJ/mol


## Lignin
A_L <- 10^0.6 #s^-1
E_L <- 46 #kJ/mol


## Kinetic parameters defined as functions of temperature, according to
## Equation 1 in the Kinetic Modeling section
k_ext <- function(T) {
  return(A_ext * exp(-E_ext / (R * T)))
}


k_hc <- function(T) {
  return(A_hc * exp(-E_hc / (R * T)))
}


k_c <- function(T) {
  return(A_c * exp(-E_c / (R * T)))
}


k_L <- function(T) {
```

```
  return(A_L * exp(-E_L / (R * T)))
}

## This function will numerically calculate the mass loss of the fractions ----
runModel <- function(T, t_f, dt) {
  ## T:   Isothermal temperature at which the model will run.
  ##      Temperatures must be passed in as Celsius.
  ##      This could theoretically be passed in as a temperature profile in a
  ##      more advanced implementation.
  ##
  ## t_f: Final time.
  ##      All time unis in this model are in seconds.
  ##      This is for how long the mass losses will be calculated.
  ##
  ## dt:  Time step for the model
  ##      This is also in seconds.

  ## Conversion of input temperature to Kelvin
  T <- T + 273

  ## First, calculate kinetic parameters at the selected temperature
```

```
k_ext_i <- k_ext(T)

k_hc_i <- k_hc(T)

k_c_i <- k_c(T)

k_L_i <- k_L(T)


## This defines the structure of the data frame where the results go

results.frame <- data.frame(t = 0,  a_ext = 0,  a_hc = 0,  a_c = 0,  a_L = 0)


## Sets up the iteration step limits for the numeric solution.

step_0 <- t_0 / dt

step_f <- t_f / dt


## Initial boundary conditions: No mass loss.

a_ext <- 0

a_hc <- 0

a_c <- 0

a_L <- 0


## This loop calculates each iteration of the model.

for(iteration in step_0:step_f) {

    if(iteration == 0) {
```

```
    ## This is the boundary condition and we know the a_'s here (0's)
    next ## just skip it
  }


## A brief algebraic exposition follows, for clarity
## From equation 1 in the kinetic modeling section:
##                          dαj/dt = kj*(1-αj)
## Applying the fundamental theorem of calculus:
##              dα/dt = lim(dt -> 0) of (α(t+dt) - a(t)) / dt
## Substituting
##       k*(1-α[t]) = lim(dt -> 0) of (α[t + dt] - α[t]) / dt
## Rewriting, as we want to look backwards in time when we iterate
##       k*(1-α[t]) = lim(dt -> 0) of (α(t) - α[t - dt]) / dt
## Dropping the limit
##               k*(1-α[t]) = (α[t] - α[t - dt]) / dt
## Rearranging and solving for current iteration's α
##               k*(1-α[t]) = (α[t] - α[t - dt]) / dt
##             dt * k * (1 - α[t]) = (α[t] - α[t - dt])
##           (dt * k) - (dt * k * α[t]) = α[t] - α[t - dt]
##           (dt * k) - (dt * k * α[t]) - α[t] = -α[t - dt]
## Sign flip
```

```
##          -(dt * k) + (dt * k * α[t]) + α[t] = α[t - dt]
##           (dt * k * α[t]) + α[t] = α[t - dt] + (dt * k)
##           (dt * k + 1) * α[t] = α[t - dt] + (dt * k)
## Arriving at the final expression
##          α[t] = (α[t - dt] + (dt * k)) / ((dt * k) + 1)

## The a_j variables here will overwrite every time the model iterates.
## R always starts counting at 1, but our model begins at t = 0. Since
## we skip the first iteration (at t = 0), the a_j values stored in
## the results.frame on the row equal to the current value of [iteration]
## are at the time step before the current one being calculated.
## The results are appended to the bottom of results.frame.

a_ext <- (results.frame$a_ext[iteration] + (dt * k_ext_i)) /
                                          ((dt * k_ext_i) + 1)
a_hc <- (results.frame$a_hc[iteration] + (dt * k_hc_i)) /
                                        ((dt * k_hc_i) + 1)
a_c <- (results.frame$a_c[iteration] + (dt * k_c_i)) /
                                      ((dt * k_c_i) + 1)
a_L <- (results.frame$a_L[iteration] + (dt * k_L_i)) /
                                      ((dt * k_L_i) + 1)
```

```
    # Binding to the frame
    results.frame <- rbind.data.frame(results.frame, c((iteration * dt),
                                                       a_ext, a_hc, a_c, a_L))
  }

  ## The overall mass loss is not part of the converison frame, so we can just
  ## print it out at the end.
  final.values <- tail(results.frame, n=1)[2:5]

  mass.loss <- final.values$a_ext * c_ext +
               final.values$a_hc * c_hc +
               final.values$a_c * c_c +
               final.values$a_L * c_L

  print(paste("Mass loss: ", mass.loss * 100, "%", sep=""))

  return(results.frame)
}

## This function makes a quick & dirty plot of the model ----
```

```
plotModel <- function(T, t_f, dt) {
  ## Notice that we call the modeling function here.
  result <- runModel(T, t_f, dt)
  ## Plot the first component, then overlay the rest on the same plot object.
  plot(result$t, result$a_ext, type = "l", col="red",
       xlim = c(0, max(result$t)), ylim = c(0, 1))
  points(result$t, result$a_hc, type = "l", col="blue")
  points(result$t, result$a_c, type = "l", col="green")
  points(result$t, result$a_L, type = "l", col="brown")
}

## This function makes a plot of the model using ggplot2 ----
ggPlotModel <- function(T, t_f, dt) {
  require(ggplot2)
  require(reshape2)
  result <- runModel(T, t_f, dt)
  m.result <- melt(result, id.vars = "t")

  p <- ggplot(data = m.result) + theme_bw(base_size = 30)

  p + geom_path(aes(x = t, y = 1 - value, col = variable))
```

```
}


ggPlotModel(270, 1250, 1)

ggPlotModel(350, 300, 1)

ggPlotModel(500, 60, 1)
```

Appendix B: Source code for linear models data preparation

```
# Copyright Statement:

# This work is licensed under a Creative Commons

# Attribution-NonCommercial-ShareAlike 3.0 Unported License.



# torrefaction data preparation.r

# Version 1.0

# November 14, 2015

# Christopher Waters

# cwaters@ou.edu

# University of Oklahoma

# College of Engineering

# School of Chemical, Biological, and Materials Engineering

# Center for Biofuels Refining

# Pyrolysis Lab Group



# A collection of utility functions for data preparation written in the R

# programming language.



rm(list = ls()) #Removes all items from the workspace
```

```
library(reshape2)


# Association panel load ——

sg.comp <- read.csv("original composition data.csv")

sg.comp <- sg.comp[complete.cases(sg.comp),]


# Compositional subset load - 22 obs. that go with thermochemical samples——

sg.comp.sub <- read.csv("original subset composition data.csv")


# Thermochemical data load ——

tc <- read.csv("torrefaction switchgrass data.csv")

tc[which(tc$lumps=="Unid."),]$lumps <- "Unknown"


# Do means of each compound per sample replicate ——

tc.c.m <- melt(tc)


# rename samples to just the switchgrass identifier

 tc.c.m$variable <- as.character(tc.c.m$variable)

for(i in 1:nrow(tc.c.m)){

  tc.c.m$variable[i] <- unlist(strsplit(tc.c.m$variable[i], split="\\_"))[3]

}
```

```
tc.c.m$variable <- as.factor(tc.c.m$variable)


# Aggregate the mean & recast, and get rid of stuff we don't need
tc.c.mean <- aggregate(tc.c.m, by = list(tc.c.m$variable,
                                         tc.c.m$Name), FUN=mean)[,c(1,2,6)]
colnames(tc.c.mean) <- c("sample", "compound", "value")
tc.mean <- dcast(tc.c.mean, compound ~ sample)
tc.mean <- as.data.frame(t(tc.mean))
colnames(tc.mean) <- as.character(unlist(tc.mean[1,]))
tc.mean <- tc.mean[-1,]
for(i in 1:ncol(tc.mean)){
  tc.mean[,i] <- as.double(as.character(tc.mean[,i]))
}
rm(tc.c.mean)


# Dimensionality reduction: Thermochemical data means, combined on lumps ——
tc.c.lump <- aggregate(tc.c.m, by = list(tc.c.m$variable,
                                         tc.c.m$lumps), FUN=mean)[,c(1,2,6)]
colnames(tc.c.lump) <- c("sample", "lump", "value")
tc.lump <- dcast(tc.c.lump, lump ~ sample)
tc.lump <- as.data.frame(t(tc.lump))
```

```
colnames(tc.lump) <- as.character(unlist(tc.lump[1,]))
tc.lump <- tc.lump[-1,]
# Fix the data to be numeric
for(i in 1:ncol(tc.lump)){
  tc.lump[,i] <- as.double(as.character(tc.lump[,i]))
}
rm(tc.c.m)
rm(tc.c.lump)
rm(i)
# Manually assigned "friendlier" names for the columns
# (should make modeling easier)
colnames(tc.lump) <- c("acetic", "akph", "furfurals", "lightgas",
                       "lightoxy", "moxph", "pyran", "sugars", "unknown")



# This function combines the TC & Comp data ----
combine.datasets <- function(tc = tc.lump, comp = sg.comp.sub) {
  # Function extracts the order of the rows from tc, then sorts comp by that
  # order, then appends the data frame.
  plots <- as.factor(as.numeric(substring(rownames(tc), 3)))
  comp <- comp[match(plots, comp$Plot),]
```

```
  return.frame <- cbind.data.frame(tc,comp)

  return(return.frame)

}

sg.combined <- combine.datasets()
```

Appendix C: Source code for torrefaction models

```
# Copyright Statement:

# This work is licensed under a Creative Commons

# Attribution-NonCommercial-ShareAlike 3.0 Unported License.



# torrefaction modeling for thesis.r

# Version 1.0

# April 1, 2016

# Christopher Waters

# cwaters@ou.edu

# University of Oklahoma

# College of Engineering

# School of Chemical, Biological, and Materials Engineering

# Center for Biofuels Refining

# Pyrolysis Lab Group



## Modeling of torrefaction data for thesis written in the R programming language.



rm(list = ls()) ## This clears out the global environment

library(stats)

library(randomForest)
```

```
library(caret)

library(MASS)

library(car)

library(ggplot2)

# Be sure you know if you're working with Torr or Pyro!

setwd("/Users/cwaters/Dropbox/Biomass Pyrolysis/Manuscripts/Switchgrass

modeling/Data/torrefaction/")

source("torrefaction data preparation.r")


## Prepare data for modeling ----

## These column numbers are hard-coded to sg.combined so be careful.

tc.columns <- 1:9

compdata.columns <- c(16:30, 32, 34, 41:43, 45:46, 48, 61:64, 68, 70, 72:76,

                      78:80, 82:85, 90, 93:96) ## For the 'original' data


## Thermochemical collinearity check

tclc <- findLinearCombos(sg.combined[,tc.columns])

tclc ## no dependent variable suggested for removal


modeldata.acetic <- sg.combined[,c(1, compdata.columns)]

modeldata.akph <- sg.combined[,c(2, compdata.columns)]
```

```
modeldata.furfurals <- sg.combined[,c(3, compdata.columns)]
modeldata.lightgas <- sg.combined[,c(4, compdata.columns)]
modeldata.lightoxy <- sg.combined[,c(5, compdata.columns)]
modeldata.moxph <- sg.combined[,c(6, compdata.columns)]
modeldata.pyran <- sg.combined[,c(7, compdata.columns)]
modeldata.sugars <- sg.combined[,c(8, compdata.columns)]
modeldata.unknown <- sg.combined[,c(9, compdata.columns)]
phenolics <- (modeldata.akph[,1] + modeldata.moxph[,1])
modeldata.phenolics <- cbind.data.frame(phenolics,
                                        sg.combined[,compdata.columns])

## Clean up stuff from this section
rm(list = c("compdata.columns", "full.sg.removal", "tc.columns",
            "tclc", "variables.to.remove"))

## Modeling function ----

buildModel <- function(modeldata, plots = FALSE, rf.variables = 7,
                       return.model = "step", custom.call = NULL) {
  ## The final approach used was random forest -> Top 7 important variables ->
  ## OLS model -> forward/backward selection -> Test for collinearity -> Final
```

```
## Model 'call' built from the column names of the data passed in
## This makes it easier for our interpretation later
dependent <- paste(colnames(modeldata)[1], "~")
call.full <- paste(dependent, ".")

# Center and scale data
# This can help us assess the relative importance of each final model
# coefficient. It also prevents us from calculating predicted values by hand
# later. It's optional and can be commented out
# modeldata <- data.frame(scale(modeldata[,-1]), modeldata[1]) ## scaling
modeldata <- data.frame(modeldata[,-1], modeldata[1]) ## For no scaling

## Generate the random forest object
set.seed(20)
model.RF <- randomForest(as.formula(call.full), data=modeldata,
                         importance=T, mtry=20, ntree=15000)

if(return.model == "RF"){
  return(model.RF)
}
```

```
## Pull the top 7 and make the final call for the OLS model
var.importance <- as.data.frame(model.RF$importance[,1])
important.variables <- rownames(head(var.importance[order(-var.importance),,
                                                    drop=FALSE], rf.variables))
independent <- paste(important.variables, collapse=" + ")
call.ols <- paste(dependent, independent)

rm(dependent)
rm(independent)

## Ordinary least squares using the 7 automatically selected features
model.OLS <- lm(as.formula(call.ols), data=modeldata)

## This is here in case we want to try some custom calls
##  - e.g., after collinearity checks
if(is.null(custom.call) == FALSE) {
  model.OLS <- lm(as.formula(custom.call), data=modeldata)
}

if(return.model == "OLS"){
```

```
  return(model.OLS)
}

## Original model probably had insignificant features
## Now doing a stepwise selection to try to improve model
model.step <- step(model.OLS, direction="both", trace = 0)

## Variance inflation score check - prints out
print(paste(colnames(modeldata)[ncol(modeldata)],
            "step model variance inflation scores"))

if(class(try(vif(model.step)))=="try-error") {
  print("Only two variables: Nothing to do")
}
else{print(vif(model.step))}

if(plots == TRUE) {
  print("doing plots")

  predicted.var <- colnames(modeldata)[ncol(modeldata)]
```

```
print(varImpPlot(model.RF,
                 main=paste("Random Forest", predicted.var,
                            "Importance Charts")))

## Value fitting
## Only use this if you're making PDF, otherwise plots won't run
par(mfrow=c(1,2))

print(plot(modeldata[,1] ~ model.OLS$fitted.values,
           main=paste("OLS prediction for", predicted.var),
           xlab = "Predicted", ylab = "Observed"))

print(plot(modeldata[,1] ~ model.step$fitted.values,
           main=paste("Stepwise Prediction for", predicted.var),
           xlab = "Predicted", ylab = "Observed"))

par(mfrow=c(1,1))

## Bar plots for discussion
##OLS Coefficients
barplot(rev(sort(model.OLS$coefficients[-1])),
```

```
                main=paste("OLS Coefficients for", predicted.var,
                           "\n (Scaled & Centered)"))

    ## Stepwise coefficients
    print(barplot(rev(sort(model.step$coefficients[-1])),
                  main=paste("Stepwise OLS Coefficients for", predicted.var,
                             "\n (Scaled & Centered)")))
  }

  return(model.step)
}

## Final, "Locked in" calls to avoid high VIF's ----
acetic.model <- buildModel(modeldata.acetic,
                           custom.call = "acetic ~ FEST + Ca + lignin.nrel")
akph.model <- buildModel(modeldata.akph,
                         custom.call = "akph ~ pCA + Ash.forage + FEST + UA")
furfurals.model <- buildModel(modeldata.furfurals,
                              custom.call = "furfurals ~ GLCS + ARA")
lightgas.model <- buildModel(modeldata.lightgas)
lightoxy.model <- buildModel(modeldata.lightoxy,
```

```
                                    custom.call = "lightoxy ~ RHA + C + ARA + HCA +
                                                  Lignin.forage + xylose.nrel")
moxph.model <- buildModel(modeldata.moxph,
                                    custom.call = "moxph ~  KL + UA + GLCS + FEST +
                                                  PCA + Ash.forage")
pyran.model <- buildModel(modeldata.pyran,
                                    custom.call = "pyran ~ PENT + GLCS")
sugars.model <- buildModel(modeldata.sugars,
                                    custom.call = "sugars ~ sgratio.nrel + EE + ASH")


## Print coefficients ----
acetic.model$coefficients
akph.model$coefficients
furfurals.model$coefficients
lightgas.model$coefficients
lightoxy.model$coefficients
moxph.model$coefficients
pyran.model$coefficients
sugars.model$coefficients


## Dump out performance info ----
```

```
r2.values <- as.data.frame(rbind(cbind('acetic', summary(acetic.model)$r.squared),
                                 cbind('akph', summary(akph.model)$r.squared),
                                 cbind('furfurals',
summary(furfurals.model)$r.squared),
                                 cbind('lightgas',
summary(lightgas.model)$r.squared),
                                 cbind('lightoxy',
summary(lightoxy.model)$r.squared),
                                 cbind('moxph', summary(moxph.model)$r.squared),
                                 cbind('pyran', summary(pyran.model)$r.squared),
                                 cbind('sugars', summary(sugars.model)$r.squared)))

r2.values

## Transformed coefficient table ----

## Let's pull a dataset of all of the 'good' model coefficients together
coeff.frm <- data.frame()
coeff.frm <- rbind.data.frame(coeff.frm,
                              cbind("acetic",
                                    names(acetic.model$coefficients)[-1],
```

```
                                        as.numeric(sign(acetic.model$coefficients[-
1])))))
coeff.frm <- rbind.data.frame(coeff.frm,
                                    cbind("akph",
                                        names(akph.model$coefficients)[-1],
                                        as.numeric(sign(akph.model$coefficients[-1]))))
coeff.frm <- rbind.data.frame(coeff.frm,
                                    cbind("furfurals",
                                        names(furfurals.model$coefficients)[-1],
                                        as.numeric(sign(furfurals.model$coefficients[-
1]))))
coeff.frm <- rbind.data.frame(coeff.frm,
                                    cbind("pyran",
                                        names(pyran.model$coefficients)[-1],
                                        as.numeric(sign(pyran.model$coefficients[-1]))))


colnames(coeff.frm) <- c("model", "coefficient", "sign")
coeff.frm$sign <- as.integer(as.character(coeff.frm$sign))


## Transforming it
dcast(coeff.frm, model ~ coefficient)
```

```
## Residual explorations ——

plot(acetic.model$fitted.values, acetic.model$residuals); abline(0,0)

plot(akph.model$fitted.values, akph.model$residuals); abline(0,0)

plot(furfurals.model$fitted.values, furfurals.model$residuals); abline(0,0)

plot(pyran.model$fitted.values, pyran.model$residuals); abline(0,0)


## Predict association panel for reasonable models ——

acetic.predict <- predict(acetic.model, newdata = sg.comp)

akph.predict <- predict(akph.model, newdata = sg.comp)

furfurals.predict <- predict(furfurals.model, newdata = sg.comp)

pyran.predict <- predict(pyran.model, newdata = sg.comp)


## Create prediction frame ——

pred.frame <- as.data.frame(cbind(plot = sg.comp$Plot,
                                  acetic = acetic.predict,
                                  akph = akph.predict,
                                  furfurals = furfurals.predict,
                                  pyran = pyran.predict))


## Changing the scale
```

```
pred.frame[,2:ncol(pred.frame)] <- pred.frame[,2:ncol(pred.frame)]/1000

m.predict <- melt(pred.frame, id.vars = 1)

## Let's combine furfurals with pyrans
pred.frame$furf.py <- pred.frame$furfurals + pred.frame$pyran

m.predict.combined <- melt(pred.frame[,c(1,2,3,6)], id.vars = 1)

## Compare observed & predicted data (xy) ——

## Vector of the plots we actually ran
run.plots <- as.numeric(c("2", "1055", "1095", "1129", "1347", "170", "207",
                          "217", "218", "246", "314", "325", "436", "452",
                          "464", "528", "67", "709", "790", "80", "90", "95"))

tc.lump$furf.py <- tc.lump$furfurals + tc.lump$pyran

observed <- melt(as.matrix(tc.lump))
colnames(observed) <- c("plot", "variable", "value")
observed$plot <- gsub("BS", "", observed$plot)
```

```
observed$plot <- gsub("000", "", observed$plot)

observed$plot <- as.numeric(observed$plot)

observed$value <- observed$value / 1000

observed$type <- "obs"


predicted.tests <- melt(pred.frame[which(pred.frame$plot %in% run.plots),],

                        id.vars = "plot")


predicted.tests$type <- "pred"


## Join these two together

compare.frame <- rbind.data.frame(predicted.tests, observed)

compare.frame <- dcast(compare.frame, plot + variable ~ type)

compare.frame <- compare.frame[complete.cases(compare.frame),]


# Basic plot object

p <- ggplot() + theme_bw(base_size = 24) +

    guides(color = guide_legend(override.aes = list(size=10,linetype=0)))


## Compare predicted vs. observed

p + geom_point(data = compare.frame, aes(y = pred, x = obs)) +
```

```
  geom_abline(data = data.frame(1), slope = 1, intercept = 0) +
  facet_wrap(~variable, scales = "free")


## Explore association panel for reasonable models ----


# m.predict[which(m.predict$plot %in% run.plots),]


## First: Simple histograms for values
## All models
p + geom_bar(data = m.predict, aes(value)) +
    facet_wrap(~variable, scales = "free") +
    geom_bar(data = subset(compare.frame, variable != "furf.py"),
             aes(obs, y=..count.. * 4), fill = "red") +
    ggtitle("Predicted Association Panel Yield Distributions\n
            All Values = (Normalized FID Area Count / 1000)\n
            Red bars are test set observations")


## Combined models
p + geom_bar(data = m.predict.combined, aes(value)) +
    facet_wrap(~variable, scales = "free") +
    ggtitle("Predicted Association Panel Combined Yield Distributions\n
```

```
        All Values = (Normalized FID Area Count / 1000)")


## Next: Three-variable showing everything.
p <- ggplot() + theme_bw(base_size = 16) +
    guides(color = guide_legend(override.aes = list(size=10,linetype=0)))


p + geom_point(data = pred.frame,
               aes(x = acetic, y = furf.py, fill = akph),
               pch = 21, alpha = 1, size = 4) +


  scale_fill_gradientn(colours=rainbow(6, start = 0, end = 0.8),
                       guide = guide_colorbar(guide_legend(title = "Alkyl\n
                                                           Phenols\n
                                                           Yield"))) +
  ggtitle("Predicted Association Panel Thermochemical Switchgrass Yields\n
          All Values = (Normalized FID Area Count / 1000)") +
  xlab("Acetic Acid Yield") + ylab("Furfurals + Pyrans Combined Yield") +
  ggsave("./Figure xx - Colorful torrefactin prediction.pdf",
         height = 5.9, width = 8.25)


## This one zooms in on the upper quadrant
```

```
quadset <- subset(pred.frame, acetic > 486)

quadset <- subset(quadset, furf.py > 228)

names(quadset)[1] <- "s.plot"


p + geom_point(data = quadset,

                aes(x = acetic, y = furf.py, fill = akph),

                pch = 21, alpha = 1, size = 4) +


  scale_fill_gradientn(colours=rainbow(6, start = 0, end = 0.8),

                       limits=range(pred.frame$akph),

                       guide = guide_colorbar(guide_legend(title = "Alkyl\n

                                                           Phenols\n

                                                           Yield"))) +

  ggtitle("Predicted Yields (Upper Right Quadrant)\n

          All Values = (Normalized FID Area Count / 1000)") +

  xlab("Acetic Acid Yield") + ylab("Furfurals + Pyrans Combined Yield") +

  ggsave("./Figure xx - Colorful UPPER QUAD.pdf", height = 5.9, width = 8.25)


## Exploration of labeled upper quadrant plot ----


## First we will dump out the distributions of the variables NOT in the coefficients
```

```
dist.sub <- sg.comp[,c(6,match(levels(coeff.frm$coefficient),
                               colnames(sg.comp)))]
m.dist.sub <- melt(dist.sub, id.vars = "Plot")

## Now show the distributions for the upper quadrant.
quad.comp <- sg.comp[match(quadset$s.plot, sg.comp$Plot),
                     c(6,match(levels(coeff.frm$coefficient),
                               colnames(sg.comp)))]

high.plots <- quadset[which(quadset$akph > mean(quadset$akph)),]$s.plot
low.plots <- quadset[which(quadset$akph < mean(quadset$akph)),]$s.plot

m.quad.comp <- melt(quad.comp, id.vars = "Plot")

m.quad.comp.high <- melt(quad.comp[match(high.plots,
                                         quad.comp$Plot),], id.vars = "Plot")
m.quad.comp.low <- melt(quad.comp[match(low.plots,
                                        quad.comp$Plot),], id.vars = "Plot")

p + geom_bar(data = m.dist.sub, aes(value), alpha = 0.5) +
  geom_bar(data = m.quad.comp.high, aes(y=..count.. * 2, value),
```

```
            fill = "#5000ff", alpha = 0.7) +
  geom_bar(data = m.quad.comp.low, aes(y=..count.. * 2, value),
            fill = "#40ff7c", alpha = 0.7) +
  facet_wrap(~variable, scales = "free") +
  ggtitle("Distributions of Model Coefficient Compositional Features")
ggsave("./Figure xx - upper quad distributions.pdf", height = 5, width = 8.25)


dcast(coeff.frm, model ~ coefficient)
```

Appendix D: Source code for pyrolysis models

```
# Copyright Statement:

# This work is licensed under a Creative Commons

# Attribution-NonCommercial-ShareAlike 3.0 Unported License.



# pyrolysis modeling for thesis.r

# Version 1.0

# April 1, 2016

# Christopher Waters

# cwaters@ou.edu

# University of Oklahoma

# College of Engineering

# School of Chemical, Biological, and Materials Engineering

# Center for Biofuels Refining

# Pyrolysis Lab Group



## Modeling of pyrolysis data for thesis written in the R programming language



rm(list = ls()) ## This clears out the global environment

library(stats)

library(randomForest)
```

```
library(caret)

library(MASS)

library(car)

library(ggplot2)


setwd("/Users/cwaters/Dropbox/Biomass Pyrolysis/Manuscripts/Switchgrass

modeling/Data/pyrolysis/")

source("pyrolysis data preparation.r")


## Prepare data for modeling ----

## These column numbers are hard-coded to sg.combined so be careful.

tc.columns <- 1:10 ## added levoglucosan

compdata.columns <- -c(11:16, 31:34, 36:38, 60:61, 66:68, 70, 72, 78:80,

                       82, 87:90, 92:94)

compdata.columns <- -c(-compdata.columns, 39, 40, 41, 45, 50, 51, 52, 53, 54,

                       55, 56, 57) ## Known linear combinations

## Toss out the forage inorganics data in favor of other data

compdata.columns <- -c(-compdata.columns, 62:65, 71)


## Collinearity checks ----

tclc <- findLinearCombos(sg.combined[,tc.columns])
```

```
tclc #no dependent variable suggested for removal


## Check full association panel for linear combos, dropping factors
combo.check <- sg.comp[, names(sg.comp)[sapply(sg.comp,class)!="factor"]]
combo.check <- combo.check[,-c(1:5, 86)] ## remove stupid stuff like plot number
full.panel.remove <- findLinearCombos(combo.check)
vars.to.remove <- colnames(combo.check)[full.panel.remove$remove]

compdata.columns <- -c(-compdata.columns, match(vars.to.remove,
                                                colnames(sg.combined)))
compdata.columns <- compdata.columns[!duplicated(compdata.columns)]

sg.combined <- sg.combined[,compdata.columns]

## Put model data together ------
modeldata.acetic <- sg.combined[,c(1, 11:ncol(sg.combined))]
modeldata.akph <- sg.combined[,c(2, 11:ncol(sg.combined))]
modeldata.furfurals <- sg.combined[,c(3, 11:ncol(sg.combined))]
modeldata.levogluc <- sg.combined[,c(4, 11:ncol(sg.combined))]
modeldata.lightgas <- sg.combined[,c(5, 11:ncol(sg.combined))]
modeldata.lightoxy <- sg.combined[,c(6, 11:ncol(sg.combined))]
```

```
modeldata.moxph <- sg.combined[,c(7, 11:ncol(sg.combined))]

modeldata.pyran <- sg.combined[,c(8, 11:ncol(sg.combined))]

modeldata.sugars <- sg.combined[,c(9, 11:ncol(sg.combined))]

modeldata.unknown <- sg.combined[,c(10, 11:ncol(sg.combined))]


phenolics <- (modeldata.akph[,1] + modeldata.moxph[,1])

modeldata.phenolics <- cbind.data.frame(phenolics,

                                        sg.combined[,11:ncol(sg.combined)])


## Clean up stuff from this section

rm(list = c("compdata.columns", "tc.columns", "tclc", "full.panel.remove",

            "vars.to.remove", "combo.check", "phenolics"))


## Modeling function ----


buildModel <- function(modeldata, plots = FALSE, rf.variables = 7,

                       return.model = "step", custom.call = NULL) {

  ## The final approach used was random forest -> Top 7 important variables ->

  ## OLS model -> forward / backward selection -> Test for collinearity -> Final


  ## Model 'call' built from the column names of the data passed in
```

```
## This makes it easier for our interpretation later
dependent <- paste(colnames(modeldata)[1], "~")
call.full <- paste(dependent, ".")

## Center and scale data
## This helps us assess  relative importance of each final model coefficient
## It also prevents us from calculating predicted values by hand later
## It's optional and can be commented out
# modeldata <- data.frame(scale(modeldata[,-1]), modeldata[1]) ## For scaling
modeldata <- data.frame(modeldata[,-1], modeldata[1]) ## For no scaling

## Generate the random forest object
set.seed(20)
model.RF <- randomForest(as.formula(call.full), data=modeldata,
                         importance=T, mtry=20, ntree=15000)

if(return.model == "RF"){
  return(model.RF)
}

## Pull the top 7 and make the final call for the OLS model
```

```
var.importance <- as.data.frame(model.RF$importance[,1])
important.variables <- rownames(head(var.importance[order(-var.importance),,
                                                    drop=FALSE], rf.variables))
independent <- paste(important.variables, collapse=" + ")
call.ols <- paste(dependent, independent)

rm(dependent)
rm(independent)

## Ordinary least squares  using the 7 automatically selected features
model.OLS <- lm(as.formula(call.ols), data=modeldata)

## This is here in case we want to try some custom calls
##  - e.g., after collinearity checks
if(is.null(custom.call) == FALSE) {
  model.OLS <- lm(as.formula(custom.call), data=modeldata)
}

if(return.model == "OLS"){
  return(model.OLS)
}
```

```
## Original model probably had insignificant features

## Now doing a stepwise selection to try to improve model

model.step <- step(model.OLS, direction="both", trace = 0)


## Variance inflation score check - prints out

print(paste(colnames(modeldata)[ncol(modeldata)],

            "step model variance inflation scores"))


if(class(try(vif(model.step)))=="try-error") {

  print("Only two variables: Nothing to do")

}

else{print(vif(model.step))}


if(plots == TRUE) {

  print("doing plots")


  predicted.var <- colnames(modeldata)[ncol(modeldata)]


  print(varImpPlot(model.RF,

                   main=paste("Random Forest", predicted.var,
```

```
                    "Importance Charts")))


## Value fitting
## Only use this if you're making PDF, otherwise plots won't run
par(mfrow=c(1,2))

print(plot(modeldata[,1] ~ model.OLS$fitted.values,
           main=paste("OLS prediction for", predicted.var),
           xlab = "Predicted", ylab = "Observed"))

print(plot(modeldata[,1] ~ model.step$fitted.values,
           main=paste("Stepwise Prediction for", predicted.var),
           xlab = "Predicted", ylab = "Observed"))

par(mfrow=c(1,1))

## Bar plots for discussion
##OLS Coefficients
barplot(rev(sort(model.OLS$coefficients[-1])),
        main=paste("OLS Coefficients for", predicted.var,
```

```
                    "\n (Scaled & Centered)"))


    ## Stepwise coefficients
    print(barplot(rev(sort(model.step$coefficients[-1])),
                  main=paste("Stepwise OLS Coefficients for", predicted.var,
                             "\n (Scaled & Centered)")))
  }


  return(model.step)
}


## Final, "Locked in" calls to avoid high VIF's ---
acetic.model <- buildModel(modeldata.acetic,
                           custom.call = "acetic ~ K + Ash + PENT2")
akph.model <- buildModel(modeldata.akph,
                         custom.call = "akph ~ pCA + glucose.nrel + K +
                                        Crude.Protein + Ash + SUC + PENT2")
furfurals.model <- buildModel(modeldata.furfurals,
                              custom.call = "furfurals ~ K + GLC + SSE")
levogluc.model <- buildModel(modeldata.levogluc)
lightgas.model <- buildModel(modeldata.lightgas)
```

```
lightoxy.model <- buildModel(modeldata.lightoxy)

moxph.model <- buildModel(modeldata.moxph,
                          custom.call = "moxph ~ ARA + lignin.nrel")

pyran.model <- buildModel(modeldata.pyran)

sugars.model <- buildModel(modeldata.sugars)

phenolics.model <- buildModel(modeldata.phenolics,
                              custom.call = "phenolics ~ FA + lignin.nrel +
                                  Ash + pCA")


## Print coefficients ----

acetic.model$coefficients

akph.model$coefficients

furfurals.model$coefficients

levogluc.model$coefficients

lightgas.model$coefficients

lightoxy.model$coefficients

moxph.model$coefficients

pyran.model$coefficientst

sugars.model$coefficients


## Dump out performance info ----
```

```
r2.values <- as.data.frame(rbind(cbind('acetic', summary(acetic.model)$r.squared),
                                 cbind('akph', summary(akph.model)$r.squared),
                                 cbind('furfurals',
summary(furfurals.model)$r.squared),
                                 cbind('levogluc',
summary(levogluc.model)$r.squared),
                                 cbind('lightgas',
summary(lightgas.model)$r.squared),
                                 cbind('lightoxy',
summary(lightoxy.model)$r.squared),
                                 cbind('moxph', summary(moxph.model)$r.squared),
                                 cbind('pyran', summary(pyran.model)$r.squared),
                                 cbind('sugars', summary(sugars.model)$r.squared),
                                 cbind('phenolics',
summary(phenolics.model)$r.squared)))

r2.values

## Transformed coefficient table ----

## Let's pull a dataset of all of the 'good' model coefficients together
```

```
coeff.frm <- data.frame()
coeff.frm <- rbind.data.frame(coeff.frm,
                              cbind("acetic",
                                    names(acetic.model$coefficients)[-1],
                                    as.numeric(sign(acetic.model$coefficients[-
1])))))
coeff.frm <- rbind.data.frame(coeff.frm,
                              cbind("akph",
                                    names(akph.model$coefficients)[-1],
                                    as.numeric(sign(akph.model$coefficients[-1]))))
coeff.frm <- rbind.data.frame(coeff.frm,
                              cbind("furfurals",
                                    names(furfurals.model$coefficients)[-1],
                                    as.numeric(sign(furfurals.model$coefficients[-
1])))))
coeff.frm <- rbind.data.frame(coeff.frm,
                              cbind("levogluc",
                                    names(levogluc.model$coefficients)[-1],
                                    as.numeric(sign(levogluc.model$coefficients[-
1])))))
coeff.frm <- rbind.data.frame(coeff.frm,
```

```
                                cbind("lightgas",
                                      names(lightgas.model$coefficients)[-1],
                                      as.numeric(sign(lightgas.model$coefficients[-
1])))))
coeff.frm <- rbind.data.frame(coeff.frm,
                                cbind("lightoxy",
                                      names(lightoxy.model$coefficients)[-1],
                                      as.numeric(sign(lightoxy.model$coefficients[-
1]))))
coeff.frm <- rbind.data.frame(coeff.frm,
                                cbind("moxph",
                                      names(moxph.model$coefficients)[-1],
                                      as.numeric(sign(moxph.model$coefficients[-1]))))
coeff.frm <- rbind.data.frame(coeff.frm,
                                cbind("pyran",
                                      names(pyran.model$coefficients)[-1],
                                      as.numeric(sign(pyran.model$coefficients[-1]))))
coeff.frm <- rbind.data.frame(coeff.frm,
                                cbind("sugars",
                                      names(sugars.model$coefficients)[-1],
```

```
                                        as.numeric(sign(sugars.model$coefficients[-
1])))))
coeff.frm <- rbind.data.frame(coeff.frm,
                                    cbind("phenolics",
                                        names(phenolics.model$coefficients)[-1],
                                        as.numeric(sign(phenolics.model$coefficients[-
1]))))

colnames(coeff.frm) <- c("model", "coefficient", "sign")

good.models <-
as.character(r2.values$V1[which(as.numeric(as.character(r2.values$V2)) > 0.41)])
coeff.frm <- coeff.frm[!is.na(match(as.character(coeff.frm$model), good.models)),]
coeff.frm$sign <- as.integer(as.character(coeff.frm$sign))
dcast(coeff.frm, model ~ coefficient)

## Residual explorations ----
plot(acetic.model, which = 1)
plot(akph.model, which = 1)
plot(levogluc.model, which = 1)
plot(moxph.model, which = 1)
```

```
plot(pyran.model, which = 1)

plot(phenolics.model, which = 1)


ncvTest(acetic.model)

ncvTest(akph.model)

ncvTest(levogluc.model)

ncvTest(moxph.model)

ncvTest(pyran.model)

ncvTest(phenolics.model)


plot(acetic.model, which = 5)

plot(akph.model, which = 5)

plot(levogluc.model, which = 5)

plot(moxph.model, which = 5)

plot(pyran.model, which = 5)

plot(phenolics.model, which = 5)


## Predict association panel for reasonable models ----

## Reasonable means r2 >~ 0.5 & at least somewhat interpretable

acetic.predict <- predict(acetic.model, newdata =sg.comp)

akph.predict <- predict(akph.model, newdata =sg.comp)
```

```
levogluc.predict <- predict(levogluc.model, newdata =sg.comp)

moxph.predict <- predict(moxph.model, newdata =sg.comp)

pyran.predict <- predict(pyran.model, newdata =sg.comp)

phenolics.predict <- predict(phenolics.model, newdata =sg.comp)


## Create prediction frame ----

pred.frm <- as.data.frame(cbind(plot = sg.comp$Plot,
                                acetic = acetic.predict,
                                akph = akph.predict,
                                moxph = moxph.predict))


pred.frm$phenolics <- pred.frm$akph + pred.frm$moxph


## Changing the scale, because mallinson

pred.frm[,2:ncol(pred.frm)] <- pred.frm[,2:ncol(pred.frm)] / 1000


m.predict <- melt(pred.frm, id.vars = 1)


## Compare observed & predicted data (xy) ----


## Vector of the plots we actually ran
```

```
run.plots <- as.numeric(c("2", "1055", "1095", "1129", "1347", "170", "207",
                          "217", "218", "246", "314", "325", "436", "452",
                          "464", "528", "67", "709", "790", "80", "90", "95"))


tc.lump$phenolics <- tc.lump$moxph + tc.lump$akph


observed <- melt(as.matrix(tc.lump))
colnames(observed) <- c("plot", "variable", "value")
observed$plot <- gsub("BS", "", observed$plot)
observed$plot <- gsub("000", "", observed$plot)
observed$plot <- as.numeric(observed$plot)
observed$value <- observed$value / 1000
observed$type <- "obs"


predicted.tests <- melt(pred.frm[which(pred.frm$plot %in% run.plots),],
                        id.vars = "plot")
predicted.tests$type <- "pred"


## Join these two together
compare.frame <- rbind.data.frame(predicted.tests, observed)
```

```
compare.frame <- dcast(compare.frame, plot + variable ~ type)
compare.frame <- compare.frame[complete.cases(compare.frame),]

# Basic plot object
p <- ggplot() + theme_bw(base_size = 24) +
     guides(color = guide_legend(override.aes = list(size=10,linetype=0)))

## Compare predicted vs. observed
p + geom_point(data = compare.frame, aes(y = pred, x = obs)) +
  geom_abline(data = data.frame(1), slope = 1, intercept = 0) +
  facet_wrap(~variable, scales = "free")

## Explore association panel for reasonable models ----
## Basic plot object
p <- ggplot() + theme_bw(base_size = 16) +
    guides(color = guide_legend(override.aes = list(size=10,linetype=0)))

## First: Simple histograms for values
## All models
p + geom_bar(data = subset(m.predict, variable != "phenolics"), aes(value)) +
    facet_wrap(~variable, scales = "free") +
```

```
    geom_bar(data = subset(compare.frame, variable != "phenolics"),
             aes(obs, y=..count.. * 4), fill = "red") +
    ggtitle("Predicted Association Panel Pyrolysis Yield Distributions\n
            All Values = (Normalized FID Area Count / 1000)\n
            Red bars are test set observations")


## Next: four-variable showing everything, Addding potassium as color


pred.frm$K <- sg.comp$K[match(pred.frm$plot, sg.comp$Plot)]


p + geom_point(data = pred.frm,
               aes(x = acetic, y = phenolics, fill = K),
               pch = 21, alpha = 0.8) +
  scale_fill_gradientn(colours=rainbow(6, start = 0, end = 0.8),
                       guide = guide_colorbar(guide_legend(title = "Potassium"))) +
  # guides(size = guide_legend(title="Levoglucosan\nYield")) +
  ggtitle("Predicted Association Panel Pyrolysis Switchgrass Yields\n
          All Values = (Normalized FID Area Count / 1000)") +
  ylab("Combined Phenolics Yield") + xlab("Acetic acid Yield") +
  ggsave("./Figure xx - pyrolysis predictions.pdf", height = 5.9, width = 8.25)
```