Analysis of gas chromatography/mass spectrometry data for catalytic lignin depolymerization using positive matrix factorization

Journal:	*Green Chemistry*
Manuscript ID	GC-ART-05-2018-001474.R1
Article Type:	Paper
Date Submitted by the Author:	19-Jul-2018
Complete List of Authors:	Gao, Yu; Washington University, Department of Energy, Environmental & Chemical Engineering Walker, Michael; Washginton University in Saint Louis Barrett, Jacob; University of California, Santa Barbara, Department of Chemistry and Biochemistry Hosseinaei, Omid; University of Tennessee Harper, David; The University of Tennessee, ; Ford, Peter; University of California, Santa Barbara, Department of Chemistry and Biochemistry Williams, Brent; Washington University in Saint Louis Foston, Marcus; Washington University, Department of Energy, Environmental & Chemical Engineering;

Green Chemistry

PAPER

Analysis of gas chromatography/mass spectrometry data for catalytic lignin depolymerization using positive matrix factorization†

Received 00th January 2018,
Accepted 00th January 2018

DOI: 10.1039/x0xx00000x

www.rsc.org/

Yu Gao,[‡a] Michael J. Walker,[‡a] Jacob A. Barrett,[b] Omid Hosseinaei,[c] David P. Harper,[d] Peter C. Ford,[b] Brent J. Williams[a] and Marcus B. Foston*[a]

Various catalytic technologies are being developed to efficiently convert lignin into renewable chemicals. However, due to its complexity, catalytic lignin depolymerization often generates a wide and complex distribution of product compounds. Gas chromatography/mass spectrometry (GC-MS) is a common analytical technique to profile the compounds that comprise lignin depolymerization products. GC-MS is applied not only to determine product composition, but also to develop an understanding of catalytic reaction pathways and of relationships among catalyst structure, reaction conditions, and the resulting compounds generated. Although a very useful tool, the analysis of lignin depolymerization products with GC-MS is limited by the quality and scope of available mass spectral libraries and the ability to correlate changes in GC-MS chromatograms to changes in lignin structure, catalyst structure, and reaction condition. In this study, GC-MS data of the depolymerization products generated from organosolv hybrid poplar lignin using a copper-doped porous metal oxide catalyst and a methanol / dimethyl carbonate co-solvent was analyzed by applying a factor analysis technique, positive matrix factorization (PMF). Several different solutions to the PMF model were explored. A 13-factor solution sufficiently explains the chemical changes occurring to lignin depolymerization products as a function of lignin, reaction time, catalyst, and solvent. Overall, seven factors were found to represent aromatic compounds, while one factor was defined by aliphatic compounds.

Introduction

Biorefineries have attracted widespread interest as a promising scheme for renewable energy, chemical, and material production.[1-4] If US second generation biofuel production targets are met in 2022, nearly 62 M dry tons/yr of lignin will be generated as a by-product, which is currently being under-utilized as low-grade fuel for heat and electricity.[5-8] Thus, technologies must be developed to efficiently use both carbohydrate and lignin fractions of biomass for biorefineries to be economical and have a minimal environmental footprint.

Lignin is one of three main plant cell wall components (i.e., lignin, cellulose, and hemicellulose), comprising ~15-30% of the dry weight of lignocellulosic biomass.[9] Described as a random copolymer, lignin is comprised of three major subunits [i.e., *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) monomers] linked by different types of inter-unit linkages (Figure 1). The native, aromatic-rich substructure of lignin makes it an ideal resource for the production of renewable aromatic and phenolic chemicals.[2, 10-12] Evolution has made lignin an integral component of the defensive and support structures within plant cell walls. Accordingly, lignin is highly

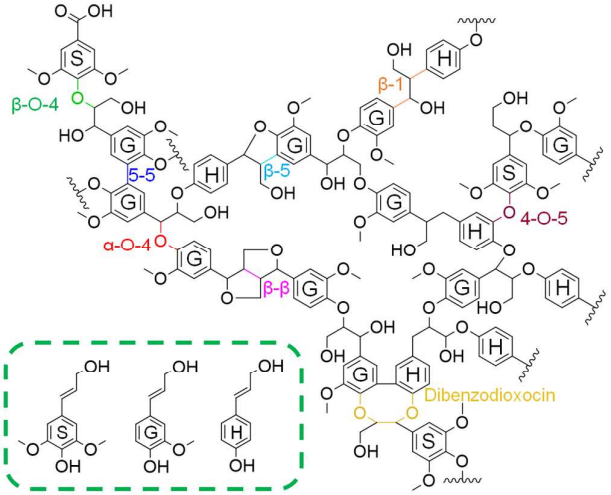


Figure 1. A structural representation of lignin, depicting various linkages and three monomers. In the bottom-left, the three monomers [i.e., coniferyl (G), sinapyl (S), and *p*-coumaryl (H) alcohol] are shown.

[a] Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis MO, 63130, USA.
*Email: mfoston@wustl.edu
[b] Department of Chemistry and Biochemistry, University of California, Santa Barbara, Santa Barbara CA, 93106-9510, USA.
[c] RISE Bioeconomy, PO Box 5604, 11486 Stockholm, Sweden
[d] Center for Renewable Carbon, University of Tennessee, Knoxville, Knoxville TN, 37996, USA
‡ Authors contributed equally
† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

resistant to biological and chemical deconstruction. Therefore, thermal depolymerization approaches require harsh reaction conditions and lead to wide product distributions that compromise downstream processing of any particular compound for chemical production.

In lignin, 50-75% of the inter-unit linkages are comprised of substructures that contain aryl-ether bonds.[13] Therefore, selective depolymerization of lignin into its monomers for chemical production at lower temperatures, which limits secondary reactions, is possible through catalytic systems that target the cleavage of aryl-ether bonds. One such promising catalytic route is hydrogenolysis which generates aromatic and phenolic derivatives as products.

Various catalysts have been evaluated for selective aryl-ether bond cleavage and lignin depolymerization.[14-19] For example, Song et al. reported using Ni catalyst on activated carbon, alumina, or porous silica to selectively convert birch wood lignin to GC-detectable phenolics in alcohols.[20] In another contribution, Ye et al. showed the selective production of 4-ethylphenolics from hydrogenolysis of lignin via noble metals (Pt, Pd, and Ru) on an activated carbon support.[21] Ford et al. studied a copper-doped porous metal oxide catalyst (CuPMO) for lignin depolymerization via hydrogenolysis in methanol (MeOH).[22-27] In addition to catalyzing aryl-ether hydrogenolysis, CuPMO promotes alcohol reforming to provide the necessary reducing equivalents for hydrogenolysis.

There are many catalytic lignin depolymerization conditions (e.g., biomass/lignin source, reaction temperature, reaction time, catalyst structure, catalyst loading, stirring speed, etc.) and phenomena that can affect lignin depolymerization reaction kinetics and networks, and thus the resulting product distribution. Therefore, controlling lignin-derived product selectivity and yield requires the ability to analyze lignin depolymerization products. More important than enabling product analysis would be the ability to leverage analytical results to develop a deeper understanding of lignin depolymerization reactions that facilitates the design of new catalysts and reaction systems for the conversion of lignin into desired products. However, due to the compositional heterogeneity and complexity of lignin and lignin depolymerization products, such analysis and its utilization is challenging.

Gas chromatography/mass spectrometry (GC-MS) has long been applied in many catalytic lignin depolymerization studies.[28-34] Even when applying selective catalysts that target aryl-ether bond cleavage, a complex distribution of compounds can still be generated. In this case, the resulting GC-MS chromatograms have a level of complexity reflecting the compositional complexity in the lignin depolymerization product. These complex GC-MS chromatograms generally consist of numerous chromatographic features (i.e., peaks), many of which are unresolved and associated with mass spectral fragmentation patterns that are not in available mass spectral libraries. The manual comparative analysis of GC-MS datasets for a small number of lignin depolymerization products, comparing their compositional distributions, can be performed.[35-38] However, when probing large numbers of products to explore the effect of several lignin depolymerization conditions, the quantity and complexity of these GC-MS datasets make human analyses difficult and time-consuming. Hence, computer-assisted signal processing can reduce GC-MS dataset complexity and transform GC-MS datasets into usable and actionable information.

The broad field of chemometrics generally aims to reduce the dimensionality of a given dataset to a more manageable number of components, clusters, or factors.[39, 40] Principal components analysis (PCA) is one of the more widely used chemometric techniques, and works by identifying the orthogonal (uncorrelated) components that can best recreate a dataset.[41] In addition to PCA, other factor analysis methods such as positive matrix factorization (PMF) and non-negative matrix factorization (NMF), have been developed to aid the analysis of complex datasets.[42, 43] Previous studies comparing PMF and PCA have found that PCA is prone to influence by low signal-to-noise measurements.[44-46] PMF and NMF analysis techniques both constrain solutions to include only positive values, but use different algorithms to reach a solution and NMF does not utilize uncertainty weighting. Based on previous studies[44, 45], uncertainty weighting (used only in PMF) limits

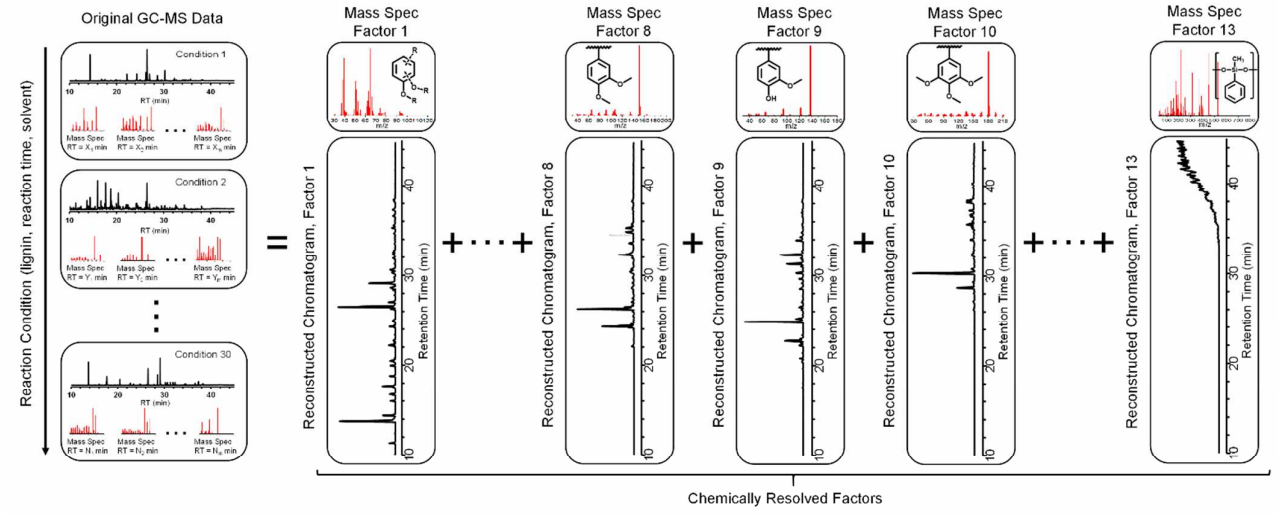


Figure 2. A graphical illustration of PMF analysis. Original lignin depolymerized GC-MS datasets (on the left of the equal sign) were analyzed by grouping lignin depolymerization products, or more specifically their chromatographic features, according to mass spectral similarity. The reconstructed chromatograms (on the right of the equal sign) represent factors that best recreate the original datasets.

the impacts of outliers and any questionable data on the solution, and thus is key to obtaining solutions that can fully recreate the input data. PMF provides solutions to a bilinear, unmixing model in which a dataset matrix is assumed to be comprised of the linear combination of factors with constant profiles that have varying contributions across the dataset.[47-49]

This study focuses on applying a PMF technique to analyze a large GC-MS dataset by grouping lignin depolymerization products, or more specifically their chromatographic features, according to mass spectral (i.e., ion fragmentation) similarity. Thus, all GC-MS detected products, including products attributed to unresolved chromatographic features or chromatographic features that have mass spectral fragmentation patterns not present in the mass spectral library, can be characterized by their chemical and structural similarities, defining factors that often represent chemical homologs (see Figure 2). PMF attempts to reduce the complexity of GC-MS datasets, and includes mechanisms that provide chemically-relevant solutions.[44] PMF has been widely used in the atmospheric chemistry community to analyze bulk MS measurements,[50, 51] and has recently been extended to more chemically-resolved GC-MS measurements of organic aerosol composition.[52, 53]

The present study was initiated to demonstrate the power of the application of PMF analysis to GC-MS datasets in the analysis of the complex reaction networks and product mixtures characteristic of lignin depolymerization. In total, 30 different reaction conditions were applied, from which, were collected 30 different sets of lignin depolymerization products. Comparing the products from 30 such samples via traditional GC-MS analysis, and more importantly, extracting a meaningful understanding of how key reaction conditions affect lignin depolymerization pathways would be very difficult. To our knowledge, this is one of the first studies that use PMF analysis of GC-MS datasets to inform catalytic reaction pathways, specifically, for lignin catalysis. This study uses a methanol-soluble and methanol-insoluble fraction of lignin extracted from a hybrid poplar biomass source, respectively denoted as MS and MIS lignin. Depolymerization reactions were conducted at 300 °C using three reaction conditions: 1) without catalyst in MeOH (non-catalyzed), 2) with CuPMO in MeOH (MeOH), and 3) with CuPMO in a MeOH and dimethyl carbonate (DMC) mixture (MeOH/DMC). DMC was applied as a co-solvent with MeOH because in our previous work[26], we found that DMC would O-methylate phenolic intermediates generated from catalytic hydrogenolysis. More importantly, we also found that the resulting aromatic methyl ether products from that O-methylation are much less susceptible to aromatic ring hydrogenation pathways than their phenolic counterparts. This study was done in a time-resolved fashion, collecting products at reaction times of 1, 2, 3, 6, and 9 h. Gaseous products of MeOH reforming and lignin depolymerization were analyzed by GC with a thermal conductivity detector (GC-TCD). Solid residues remaining after lignin depolymerization (i.e., lignin that has undergone chemical modification such that it is methanol-insoluble and/or char) were analyzed by dioxane extraction, nitric acid digestion, and thermogravimetric analysis (TGA). Liquid products (i.e., methanol-soluble lignin remaining after reaction that may or may not have undergone chemical modification and GC-detectable lignin depolymerization products) were analyzed by gel permeation chromatography (GPC) and GC-MS (see Supplemental Information for details on methods). Primarily though, GC-MS data is the main focus of this report. PMF analysis of the GC-MS data was applied to understand how lignin source (i.e., MS and MIS lignin), reaction time, and the presence of CuPMO and/or DMC affects the GC-detectable (i.e., low-molecular weight and volatile) product distribution of lignin depolymerization. We found that, compared to traditional (manual) peak integration and assignment analysis for GC-MS datasets, our PMF analysis significantly reduces the time required to complete data processing, allows compounds to be chemically classified that are not in a MS library, and facilitates the analysis of the unresolved complex mixtures (UCMs).

Table 1. Summary of the chemical assignment for individual factors with their major characteristic fragment ions.

	Defined Factor	Major Characteristic *m/z*
Factor 1	less polar and/or more volatile aromatics	39, 50, 65, 74, 93
Factor 2	air and other light contaminates	31, 40, 44
Factor 3	less polar and/or more volatile aromatics	39, 51, 63, 77, 91, 107
Factor 4	aliphatics	41, 55, 69, 83, 97, 111
Factor 5	carboxylics and carbonates	31, 47, 75
Factor 6	benzoates	39, 65, 93, 121
Factor 7	more polar and/or less volatile aromatics	39, 53, 67, 79, 95, 109, 123
Factor 8	dimethoxy benzyls	39, 65, 79, 91, 107, 119, 135, 151
Factor 9	methoxy phenyls	39, 55, 65, 79, 94, 105, 122, 137
Factor 10	trimethoxy benzyls	39, 45, 52, 66, 79, 92, 105, 120, 136, 148, 167, 181
Factor 11	unresolved compounds	31, 41, 55, 65, 77, 91, 105, 115, 135, 149, 165, 179, 191
Factor 12		150, 165, 195, 253, 315, 393, 408, 451
Factor 13	Column residues and heavier contaminants	135, 156, 179, 197, 218, 239, 255, 315, 373, 393, 451, 529

Results and discussion

Factor analysis techniques.

As applied herein, PMF analysis is for grouping chromatographic features into classes of compounds (i.e., factors) based on similarities in MS fragmentation patterns, or more succinctly, similarities in chemical structure. Our use of PMF analysis also involves a chromatographic binning technique that sums together sequential MS scans to reduce computational time and minimize the effects of retention time shifting. GC-MS chromatographic signal intensities therefore undergo pre-processing (i.e., retention time shift correction, background subtraction and internal standard normalization) and chromatographic binning before PMF analysis.

The combined chromatographic binning and PMF analysis of the GC-MS data from the 30 lignin depolymerization samples results in a set of PMF solutions where the number of factors within a solution is specified by the user. The Q/Q_{exp} value for a given solution represents the extent to which the residual data (i.e., the data from the input data matrix that was not recreated through any of the factors) can be explained by the user-provided uncertainty estimates. However, this measure alone cannot ultimately inform which solution is the "correct" solution, as an mathematical criteria would be arbitrary.[54] A subjective choice by the user is still ultimately required to conclude the number of factors that best fits the data and its interpretation.[50]

To settle upon a final solution and set of factors, solutions where we defined the number of factors as 2-18 were considered. The Q/Q_{exp} values decreased monotonically with the addition of a new factor to the solution (Figure S1). Chemically distinct classes of compounds were identified in factors for solutions up to the 13-factor solution, but following that point factor splitting resulted in redundant factors that did not add additional chemical insight. Such factor splitting often results in factors that are composed of a single compound, which taken to its extreme would produce a factor for each compound detected and defeat the purpose of performing factor analysis. For solutions that are fewer than the optimal number of factors, resolvable factors are presumably superimposed. In our case, the 13-factor solution generated a set of factors that mostly did not display factor splitting yet yielded factors which provided the maximum chemical insight. Thus, a 13-factor solution (summarized in Table 1) was chosen and is the subject of further discussion. The mass spectrum for a given factor consists of a set of co-varying fragment ions that best recreates the input dataset upon a linear combination with the other factors in the solution. The electron ionization (EI) used in MS analysis produces ions from GC-separated compounds that fragment in a reproducible way, generating similar fragmentation patterns for compounds that have similar structural moieties. Therefore, the factor mass spectrum is useful in the identification of a homologous series of compounds that defines the factor.

The retention time-series data of a given factor upon further post-processing can be used to reconstruct a factor average chromatogram for a given factor eluting from the column. Factor average chromatograms can be reconstructed for either any sub-set of PMF input samples (e.g., different catalyst, lignin, and/or reaction conditions) or for all of the samples. For example, Figure S2A shows the factor average chromatograms constructed in this manner for product samples from catalyzed reactions in MeOH/DMC (green), catalyzed reactions in MeOH (blue), and non-catalyzed reactions in MeOH (black). Figure S2D shows the factor average chromatogram for the combination of all reaction conditions. The binned abundances displayed in a factor average

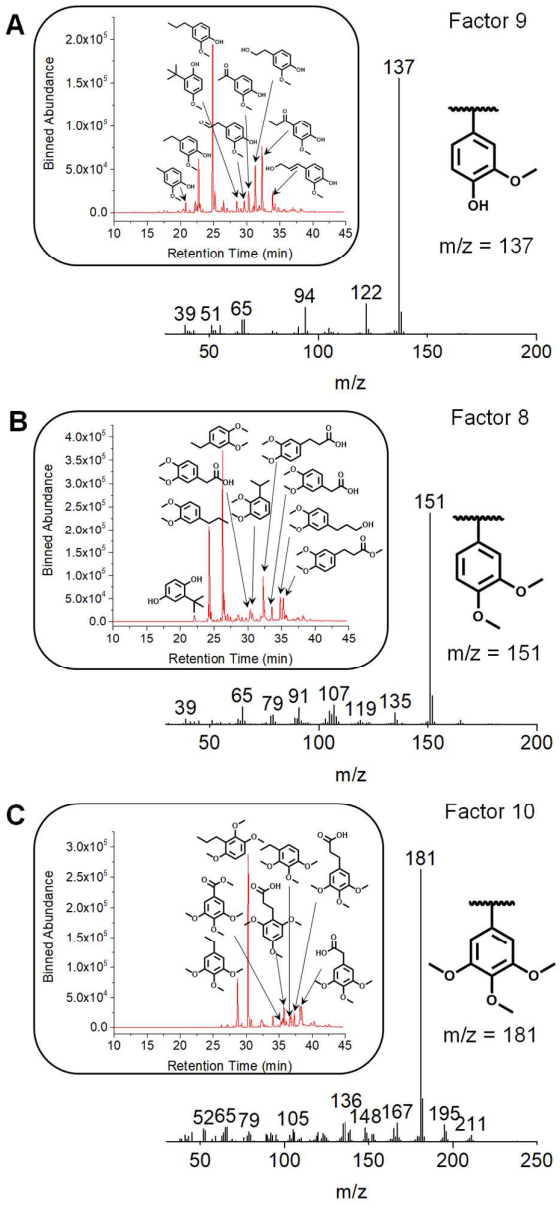


Figure 3. Factor average chromatograms and mass spectra for Factor 8 (dimethoxy benzylic), Factor 9 (methoxy phenolic), and Factor 10 (trimethoxy benzylic) of the 13-factor PMF solution for GC-MS datasets of lignin depoylmerization samples from MS/MIS lignin that have udergone depolymerization for 1-9 h using non-catalyzed, MeOH, and MeOH/DMC conditions.

chromatogram corresponds to the amount of the signal (i.e., output from the mass spectrometer) that the model apportions to that factor at that retention time. The factor abundance (Figure S2C) of one (or multiple) factor(s) can be calculated by integrating the entire area of the factor chromatogram for a given sample (or set of samples). Due to the complexity of the products across the samples and differences in detector responses for different compounds, these factor abundances cannot quantitatively be thought of as the mass of compounds comprising different factors. The reported factor abundance is therefore in an arbitrary unit. However, considered in a more qualitative manner, a much higher abundance for a specific factor would suggest that the higher abundance factor is compositionally favored in the lignin depolymerization sample. More useful, when comparing two different lignin depolymerization samples, an increase in factor abundance for a specific factor when comparing one sample to another would suggest that one sample compositionally favors that factor relative to the other sample. Again, in that case, the intensity difference between factors is not quantitative, but suggestive in nature to the degree of difference. Thus, the comparisons of factor abundances can provide useful insights into the different processes occurring for different samples and under different reaction conditions.

In combination, the factor mass spectrum and average chromatogram are used to assign the chemical identity of a given factor. For example, in the 13-factor solution, the Factor 1 mass spectrum contains fragment ions with *m/z* values of 39, 50, 65, 74, and 93 (Table 1). These fragment ions match common diagnostic ions that originate from and represent fragments of compounds that contain aromatic/phenolic moieties. The factor average chromatogram for Factor 1 indicates that compounds assigned to Factor 1 elute at early retention times, which corresponds to aromatics that are less polar and/or more volatile (due to the GC column type and GC oven heating ramp program). To verify the factor assignments, peaks in the factor average chromatograms were assigned by identifying compounds eluting at the same retention time for GC-MS datasets from the 30 lignin depolymerization samples. Due to the limited number of known compounds in the mass spectral library and the number of unresolved peaks, not all peaks could be assigned with a high level of certainty. However, the majority of identified peaks suggested that the classifications of the factors based on the factor mass spectrum are reliable. A complete list of MS library-identified compounds from the GC chromatogram of all 30 samples is provided in Table S1.

The 13 identified factors include both resolved and unresolved GC-detectable products. Due to the reliance on differences in mass spectral fragmentation patterns in separating the factors, it is important to note that a single compound may contribute to more than a single factor. Table 1 defines each factor's chemical identity based on the characteristic *m/z* values in the factor mass spectrum and the identified compounds associated with the factor. Table S2 provides further details about fragment ions and their chemical assignment. Compounds in Factors 1 and 3 are primarily low polarity aromatic/phenolic compounds (Table 1, Figure S2, and Figure S4). Factor 3 compounds generate fragment ions that are associated with benzyl moieties (Table 1 and Figure S4). The differences between Factors 1 and 3 are driven by the association of fragments related to phenol and methyl 4-hydroxybenzoate with Factor 1, which are absent from Factor 3. Similarly, Factor 7 compounds are also aromatic but tend to be more polar and/or less volatile than compounds in Factors 1 and 3 (Table 1 and Figure S8). Factor 6 compounds have fragment ions that are associated with benzoate moieties (Table 1 and Figure S7). The Factor 6 average chromatogram is dominated by two compounds: methyl 4-hydroxybenzoate and methyl 4-hydroxy-3-methoxybenzoate. Other classes of aromatic compounds are found in Factors 8, 9, and 10. Specifically, Factors 8, 9, and 10 contain compounds that have fragment ions associated with dimethoxy benzylic, methoxy phenolic, and trimethoxy benzylic moieties, respectively (Table 1, Figure S9, Figure S10, and Figure S11). The mass spectra for Factors 8, 9, and 10 are dominated by fragment ions at *m/z* 151, 137, and 181, respectively (Figure 3). Similarities in Factors 1 and 3 due to their association with fragments related to phenol and Factors 1 and 6 due to their association with fragments related to methyl 4-hydroxybenzoate suggest these factors may be the result of splitting. However, the 13-factor solution was the first solution to generate separate factors (i.e., Factors 8, 9, and 10) related to compounds that resemble lignin monomers. Though not done, we could have also visualized the aggregated factor abundances of Factors 1 and 3; Factors 1 and 6; and Factors 1, 3, and 6 to account for this splitting and further detect trends that inform our understanding of lignin depolymerization.

The mass spectrum of Factor 5 is dominated by a fragment ion at *m/z* 75, suggesting compounds in Factor 5 contain moieties that have a carboxylate and two additional carbons (e.g., methyl acetate, ethyl formate, propionate, etc.). Factors 4 and 11 contain unresolved chromatographic features within their factor average chromatograms. Factor 4 has primarily a low-retention time (low polarity and/or more volatile compounds) unresolved complex mixture (UCM) with mass spectral features indicative of substituted aliphatics (Table 1 and Figure S5). Conversely, Factor 11 features a high-retention time UCM (higher polarity and/or less volatile compounds) that includes mass spectral features indicative of both aliphatics and aromatics (Table 1 and Figure S12). UCMs appear as a hump or background feature in a chromatogram and represent a large number of co-eluting compounds. UCMs are commonly observed for petroleum or biomass-derived pyrolysis oils.[55, 56] Factors 4 and 11 suggest that our lignin depolymerization samples contain both polar and non-polar UCMs. The remaining factors can be assigned as measurement artifacts, which persist even after the pre-processing. Factor 2 comes from air within the GC-MS, with *m/z* 32, 40, and 44 attributed to oxygen, argon, and carbon dioxide, respectively. The variation in abundance of Factor 2 across the different samples is largely driven by the scaling differences introduced with the internal standard normalization. The column bleed from the GC column is the defining feature of Factors 12 and 13, and the variation in abundance across samples is driven by how similar the subtracted blank sample was to a given sample. Details

of the factor mass spectra and average chromatograms for all 13 solution factors are given in Figures S2-S14.

Analysis of trends in GC-detectable products as grouped by PMF.

The overall GC-detectable aromatic compound production was monitored by combining the aromatic factors (Factors 1, 3, 6, 7, 8, 9, and 10). Figure 4A is the factor average chromatogram for the combined aromatic factors and the aggregated factor abundance for these aromatic factors across reaction conditions are in Figure 4B. Similarly, the aliphatic factor average chromatogram (Factor 4) is in Figure 4C with the factor abundance for the aliphatic factor across reaction conditions in Figure 4D.

As shown in Figure 4, catalyzed depolymerizations (MeOH and MeOH/DMC) generated higher factor abundances of GC-detectable products (i.e., aromatic and aliphatic) compared to non-catalyzed depolymerizations. Additionally, the overall production of aromatic compounds in catalyzed depolymerizations was significantly higher than that without catalyst. This increase for catalyzed depolymerizations was due to the catalytic activity[26] of CuPMO producing H_2 (Figure S18) and performing aryl-ether hydrogenolysis. Over the course of the entire reaction, the production of GC-detectable products remained relatively low for depolymerizations without catalyst, which indicated that most of the depolymerized lignin fragments in liquid products are larger molecular weight species (i.e., non-GC-detectable) that were unreacted or subject to condensation reactions.

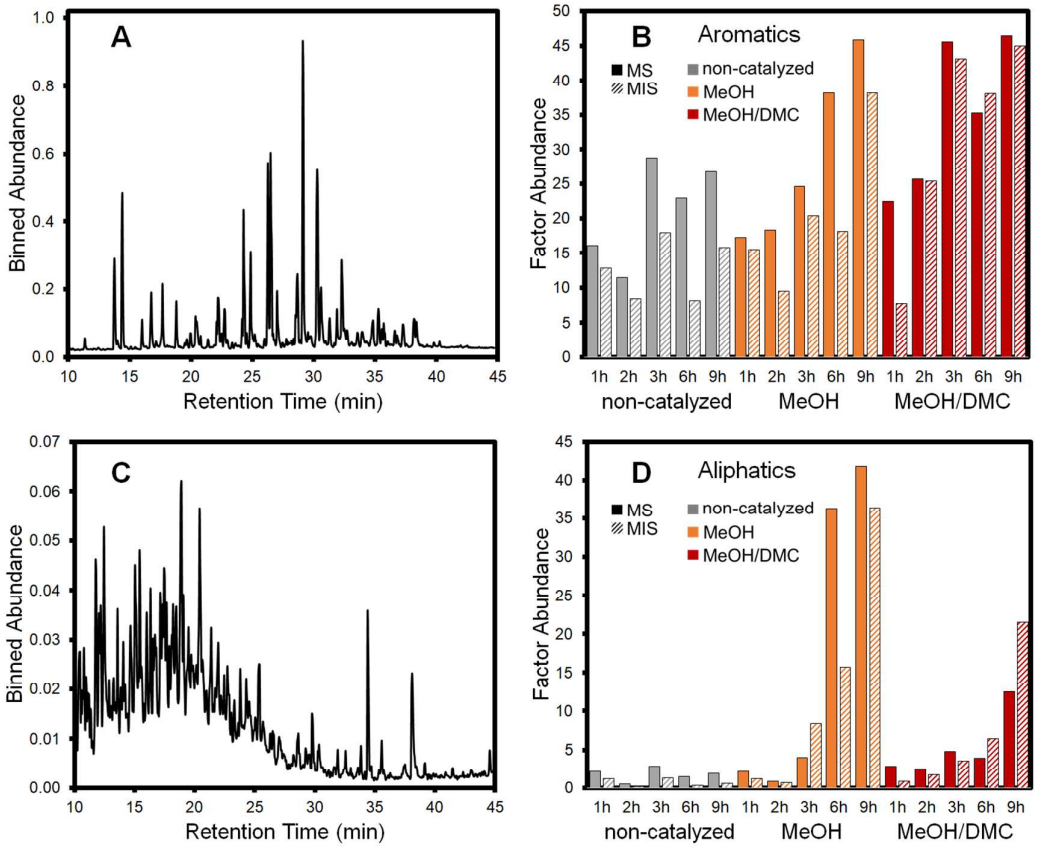


Figure 4. A) Factor average chromatograms of the combined aromatic factors (i.e., Factors 1, 3, 6, 7, 8, 9 and 10); B) combined aromatic factor abundance; C) factor average chromatograms of the aliphatic factor (i.e., Factor 4); and D) aliphatic factor abundance for lignin depoylmerization samples from MS/MIS lignin that have udergone depolymerization for 1-9 h using non-catalyzed, MeOH, and MeOH/DMC conditions.

Aside from our work[26], Bernt et al. also found that anisole and ethoxybenzene are much less reactive (i.e., slower rate of conversion) over CuPMO in MeOH than are phenol or guaiacol.[24] Phenol conversion was attributed primarily to reaction pathways: 1) reduction to cyclohexanol (k_{obs} ~0.5 h^{-1}), 2) methylation of the aromatic ring to give cresols (k_{obs} ~0.1 h^{-1}), and 3) O-methylation at phenolic alcohols to form anisole (k_{obs} ~0.1 h^{-1}). The primary product of anisole over CuPMO in MeOH was benzene, a very stable product, which resulted from (C_{aryl}-$O_{methoxy}$) hydrogenolysis of the methoxyl group; although, at longer reaction times some ring hydrogenation to methoxy-cyclohexane occurred. These results suggest that DMC and O-methylation of fragments from lignin depolymerization increases pathways to products that are less susceptible to certain reactions, such as ring hydrogenation reactions, when compared to their phenolic counterparts. We consider hydrogenation of phenolics as undesirable due to the loss of aromaticity and broadening of product distribution. With increasing reaction time, the abundance of the aliphatic factor increased for MeOH samples (Figure 4D), while a relatively lower abundance of the aliphatic factor was detected for MeOH/DMC samples. Observed relative increases for the abundances of Factors 8 and 10, which represent increases in O-methylated aromatics, are the highest for MeOH/DMC samples.

Factors 1 and 3 represent low polarity aromatic compounds (Figure S2 and S4), while Factor 11 compounds are associated with a more polar and/or less volatile UCM (Figure S12). The abundance of both Factors 3 and 11 in non-catalyzed samples are relatively lower when compared to their abundance in MeOH and MeOH/DMC samples. Whereas, the abundance for Factors 3 and 11 increase with reaction time for MeOH and MeOH/DMC samples.

Factors 5, 6, and 7 represent compounds that contain carbonyl substituents that have been correlated with the production of reactive compounds similar to those that require stabilization and upgrading in lignin pyrolysis oil.[4, 57] Factor 5 compounds are aromatic and aliphatic compounds that generate carboxylate moieties upon EI, and are much more abundant in non-catalyzed samples than MeOH or MeOH/DMC samples (Figure S6). The average factor chromatogram of Factor 5 for non-catalyzed samples shows two major chromatographic features at retention times of 32.33 min [3-(3,4-dimethoxyphenyl)propanoic acid] and 38.23 min [methyl octadecenoate or methyl stearate]. Factor 6 compounds, defined as having benzoate moieties, are more abundant in non-catalyzed samples than MeOH samples and then more abundant in MeOH samples than MeOH/DMC samples. The average factor chromatogram of Factor 6 for non-catalyzed and MeOH samples shows two major chromatographic features at retention times of 29.11 min [methyl 4-hydroxybenzoate] and 30.61 min [methyl 4-hydroxy-3-methoxybenzoate], which are not present in the DMC/MeOH samples (Figure S7). The average chromatogram of Factor 6 for MeOH/DMC samples suggests that the DMC shifts reaction pathways in such a fashion that the production of compounds with benzoate moieties is reduced, in particular, methyl 4-hydroxybenzoate and methyl 4-hydroxy-3-methoxybenzoate. Factor 7 compounds are most prevalent in MeOH samples at reaction times of 6

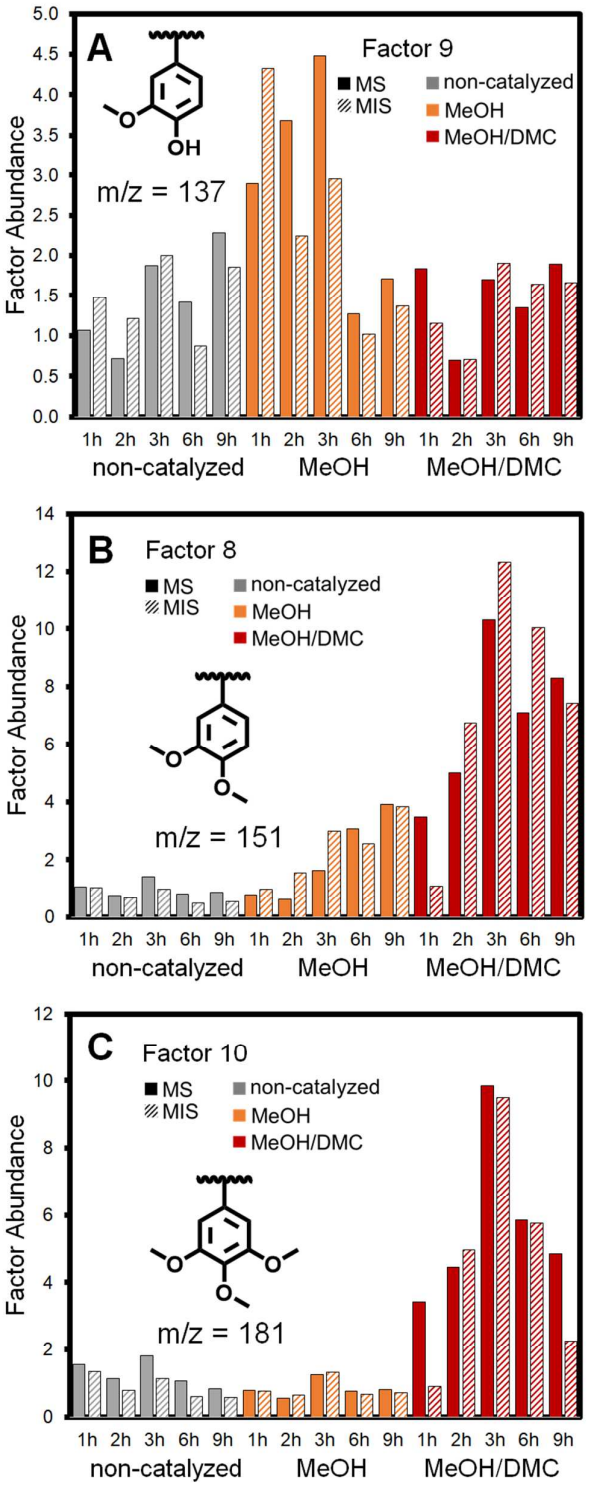


Figure 5. A) Factor 8 (dimethoxy benzylic); B) Factor 9 (methoxy phenolic); and C) Factor 10 (trimethoxy benzylic) abundance for lignin depoylmerization samples from MS/MIS that have udergone depolymerization for 1-9 h using non-catalyzed, MeOH, and MeOH/DMC conditions.

and 9 h and MeOH/DMC samples at a reaction time of 9 h. For Factor 7, resolved compounds begin to elute from the column after 22 min that are primarily associated with non-catalyzed and MeOH/DMC samples, whereas ~ 5 – 32 min the average factor chromatogram presents as a UCM as a result of contributions from MeOH samples (Figure S8). The 2D ^{13}C-^{1}H heteronuclear single quantum correlation (HSQC) nuclear magnetic resonance (NMR) spectra of MS and MIS lignin substrates show they contain α-oxidized syringyl, α-oxidized guaiacyl and *p*-hydroxybenzoate monolignols (Figure S16). Factor 6 and 7 compound production could be, in part, linked to the release those monomers.

Factors 8, 9, and 10 most resemble the monomeric substructures of lignin (Figure 3 and 5). Factors 8 and 10 are individually defined by characteristic features of di- and tri-methoxylated compounds, which most likely result from O-methylation of phenolic intermediates from G and S lignin monomers. Consequently, these two factors have higher abundance in MeOH/DMC samples (Figure 5). The average chromatograms of Factors 8 and 10 for MeOH and MeOH/DMC samples each have two major chromatographic peaks at retention times of 24.27 min [4-ethyl-1,2-dimethoxybenzene] and 26.26 min [1,2-dimethoxy-4-propylbenzene] for MeOH and as 28.68 min [5-ethyl-1,2,3-trimethoxybenzene] and 30.29 min [1,2,3-trimethoxy-5-propylbenzene] for MeOH/DMC (Figure S9 and S11). Bernt el al. showed that 1,2-dimethoxybenzene was more reactive than anisole over CuPMO in MeOH; however, the primary product of 1,2-dimethoxybenzene was anisole resulting from the (C_{aryl}-$O_{methoxy}$) hydrogenolysis of a methoxyl group.[24] As a result, we expect that the di- and tri-methoxybenzene compounds resulting from DMC mediated lignin depolymerization would similarly be susceptible to methoxyl group hydrogenolysis. The average chromatograms for Factors 8 and 10 suggest that the presence of the catalyst does promote phenolic O-methylation producing di- and tri-methoxylated aromatic compounds but that, even without catalyst, some phenolic O-methylation occurs. Notably, non-catalyzed reactions tend to produce a wide distribution of di- and tri-methoxylated compounds that are more polar than 1,2-dimethoxy-4-propylbenzene and 1,2,3-trimethoxy-5-propylbenzene due to the presence of various of functionalities on the fragment at the monolignol propyl substituent. The abundances for Factors 8 and 10 suggest that at longer reaction times (> 3 h), di- and tri-methoxylated aromatic compounds are converted into some other species. Conversely, Factor 9 represents characteristic features of compounds that contain aromatic rings with both hydroxyl and methoxyl substituents, which are generated from G lignin monomers that did not undergo O-methylation. The average chromatogram of Factor 9 for MeOH and MeOH/DMC samples each have a common major chromatographic peak at retention times of 24.87 min [2-methoxy-4-propylphenol]. Compounds related to Factor 9 show significantly higher abundance in the MeOH samples at short reaction times between 1-3 h than other samples. The decrease in Factor 9 abundance for MeOH samples beyond a reaction time of 3 h is likely due to reaction pathways shown to be prevalent for phenol conversion over CuPMO. The average factor chromatogram of Factor 9 for non-catalyzed samples has three additional major chromatographic peaks not observed in MeOH or MeOH/DMC sample at retention times of 29.64 min [2-(4-hydroxy-3-methoxyphenyl)acetaldehyde], 31.31 min [4-(2-hydroxyethyl)-2-methoxyphenol], and 32.33 min [1-(4-hydroxy-3-methoxyphenyl)propan-1-one] (Figure S10).

Comparing the lignin depolymerization products generated from MS and MIS lignin, overall GC-detectable aromatic compound production as well as Factors 8 and 10 abundances for MS and MIS samples were similar. This suggests there was little effect on product distributions caused by room temperature solubility or by chemical differences that existed for MS and MIS, which are highlighted in Table S3, Table S4, and Figure S16. However, Factor 6 abundances suggest that under non-catalyzed and MeOH conditions MS-derived samples contain more benzoates compounds, while Factor 4 abundances suggest that at reaction times of 6 and 9 h in MeOH, MS is more susceptible to aliphatic forming reaction pathways. Factor 5 abundances suggest that under non-catalyzed conditions and at a reaction times of 1 and 2 h, MIS-derived samples contain more carboxylate compounds.

PMF and NMF comparison.

In an effort to assess the effectiveness of PMF analysis on these GC-MS datasets, a 13-factor NMF solution was also obtained using the same input data matrix. As expected, computation time was much faster for NMF (approximately 14 seconds to converge) compared to PMF (approximately 600 seconds to converge). However, the NMF solution with the same number of factors had a much larger residual component, with 71% of the input data left unexplained whereas only 21% of the signal was unexplained by the PMF solution. While the residual would likely decrease with additional factors in the NMF solution, the same should be true for PMF. The average abundance for all samples obtained from the summation of the NMF factors only better recreates the input data for a single GC peak compared to PMF (Figure S15). Given these observation and our previous efforts in source apportionment of atmospheric organics, PMF seems to be a better method for this dataset and may better lend itself to similar chromatographic datasets of complex mixtures with resolved and unresolved features.

Non-GC-detectable products.

Lignin oligomers are both intermediates and products of lignin depolymerization; however, these lignin oligomers are not detectable by GC-MS. Hence, the production of lignin oligomers was

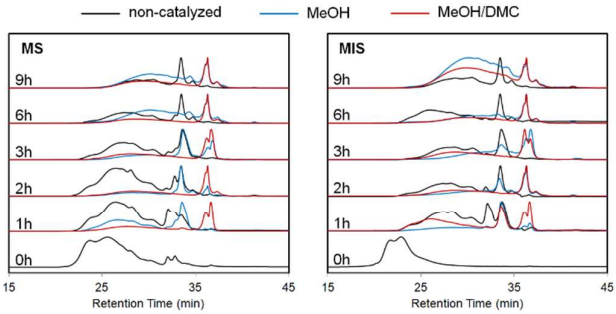


Figure 6. GPC chromatograms for untreated lignin and for product samples from MS and MIS liignin that have udergone depolymerization for 1-9 h using non-catalyzed, MeOH, and MeOH/DMC conditions.

examined by GPC analysis. Untreated lignin and their depolymerized liquid products were directly injected into the GPC. Relative molecular weight values, including number-average molecular weight (M_n), weight-average molecular weight (M_w), and polydispersity index (PDI = M_w/M_n), were determined based on GPC retention times and a polystyrene standard calibration curve (Table S5 and S6). A higher PDI means a broader distribution of molecular weights. Lignin depolymerization, as signified by chromatographic shifts toward longer retention times, occurred faster for catalyzed depolymerizations. After reaction times of 2 h for catalyzed depolymerizations, the broad peak from ~22-30 min representing untreated lignin almost completely disappears (Figure 6). However, this peak persists or even shifts toward shorter retention times for non-catalyzed depolymerizations at long reaction times. This suggest either that effective depolymerization is not occurring or that re-condensation of GC-detectable compounds is occurring. High abundance of Factors 5 and 6 seem to correlate with samples that are susceptible to GPC chromatographic shifts toward longer retention times; however, well-defined, co-varying trends between GPC behavior and factor abundance were difficult to extract.

Solid products.

Raw solid residues were also analyzed to study the leftover lignin and char formation as shown in Figure S17. Solid residues from each reaction were first separated from liquid products by filtration. Dioxane was used to extract dioxane-soluble solid products after lignin depolymerization. Note that both MS and MIS lignin are soluble in dioxane at room temperature. For the reactions without catalyst, the leftover solids after dioxane extraction were composed of dioxane-insoluble lignin and char. In this case, dioxane-insoluble solids were treated with nitric acid to determine char yields. Char formation increased with reaction time and was higher for MIS. For the reactions with catalyst, char formation was never observed and dioxane-insoluble solid content were higher than reactions without catalyst. In addition, dioxane-insoluble solid content decreased as a function of increasing reaction time. Dioxane-insoluble solids formation most likely result from reactions that modify the chemical structure of the lignin such that it is no longer soluble in dioxane. Dioxane-insoluble solids contents were higher for MIS lignin depolymerization samples when compared to MS lignin depolymerization samples.

Gas products.

Gas products were collected and analyzed to track the formation of H_2 production (Figure S18). No gas or H_2 formed for non-catalyzed depolymerizations. Gas products in both MeOH and MeOH/DMC catalyzed depolymerizations are mainly composed of H_2 and CO_2, with small amounts of CO and CH_4. The production of H_2 corresponds well with the increase overall aromatic compound production monitored by aromatic factors (Factors 1, 3, 6, 7, 8, 9, and 10) and the aliphatic compound production monitored by Factor 4.

Conclusions

In summary, upon PMF analysis of GC-MS datasets from 30 different lignin depolymerization products that were depolymerized as a function of lignin, reaction time, catalyst, and solvent, we determined a 13-factor solution sufficiently explains the chemical changes occurring. These 13 factors represent various classes of compounds based on similarities in chemical structure that best reconstruct the original lignin depolymerization PMF inputs. Factors include low and high polarity aromatic compounds, compounds with carbonyl moieties, compounds that resemble lignin monomers, and aliphatic compounds. Overall catalyzed depolymerizations generated higher factor abundances of GC-detectable products compared to non-catalyzed depolymerizations. In addition, we found that with increasing reaction time, the abundance of the aliphatic factor increased for MeOH samples while MeOH/DMC samples remained at a relatively low abundance. Thus, catalyzed depolymerization in the latter medium was superior at preserving product aromaticity. The products generated by reaction in MeOH in the absence of catalyst seemed to contain more compounds with carbonyl substituents. Lastly, we determined that there was little discernable difference in the GC-detectable products generated from MS and MIS lignin.

The complexity of lignin makes conducting fundamental research into its catalytic reactivity difficult. As a result, the development of analytical tools that can effectively capture this complexity and data processing methods that can interpret those analytical results is critical to the success of such fundamental research. Our results show that PMF analysis, as a computer-assisted signal processing tool to reduce GC-MS dataset complexity, can be applied to GC-MS datasets not only for the purposes of understanding lignin depolymerization but also a broad range of other chemical processes that involve complex reaction networks and product distributions.

Experimental

Materials.

Analytical grade methanol, n-decane, and reagent grade dimethyl carbonate (DMC) were used as purchased from Sigma-Aldrich. Lignin used in this work was extracted from *Populus spp*. biomass (see Supplementary Information). The CuPMO catalyst used in this work was synthesized by following the same procedure as reported by Ford et al. (see Supplementary Information).[26]

Lignin depolymerization.

The lignin depolymerization reactions have been conducted in stainless steel bomb reactors with internal volume of ~ 10 mL. The bomb reactor design used was from previous work by Ford et al.[22, 23, 26, 58, 59] Each reactor was charged with 100 mg of lignin (MS or MIS) and 100 mg of CuPMO catalyst. Either methanol (3 mL) only or pre-mixed methanol and dimethyl

carbonate (2:1 ratio, 3 mL) solution with n-decane (1.76 µL) as internal standard was added into the reactor as solvent. The reactor was heated in an isotherm muffle furnace (Thermo Fisher Scientific) at 300 °C for reaction times of 1, 2, 3, 6, and 9 h. A series of reactions on the same lignin substrates were conducted without catalyst in methanol. Reactors were quenched within an ice water bath. Gas products were collected by an inverted graduate cylinder, which is pre-filled with water. The volume of gas products was measured by the displacement of water by gas collected in the cylinder. Gas composition was determined by GC-TCD. Solid residues and liquid products were separated by a vacuum filtration apparatus with a 0.45 µm nylon membrane filter. Solids were further washed by analytical grade methanol portion by portion until the total liquid products volume is 20 mL. Liquid products were collected for GPC and GC-MS analysis. Solid residues were further washed by dioxane (5 mL) three times to extract leftover untreated lignin (that was soluble in dioxane) from dioxane-insoluble lignin, char, and catalyst. The solids following dioxane extraction were subjected to TGA and nitric acid digestions to determine the amount of catalyst and char present.

Product characterization.

GC-MS was used to characterize the GC-detectable products from lignin depolymerization. GC-MS samples (1 µL) were injected on an Agilent GC system 7890A coupled with an Agilent 5975C mass spectroscopy with triple-axis detector. Triplicate injections were performed for the 6h MeOH/DMC samples derived from MS and MIS lignin. GC analysis was performed using a Restek fused silica RTX-50 capillary column (ID: 0.25 mm, film thickness: 0.5 µm, and length: 30 m) with the following program: 2 min at 35 °C and then ramped at 5 °C/min up to 300 °C for 5 min with helium as a carrier gas (splitting ratio: 10:1). The mass spectral scan rate was 1.6 scans/s with acquisition from *m/z* 30-600. GC-MS data was exported and analyzed through ChemStation Software. Identification of the compounds was carried out by comparing the mass spectra obtained with these from Palisade Complete Mass Spectral Database (600 K edition, Palisade Mass Spectrometry, Ithaca, NY). Non-GC-detectable products (gas, solid, and non-GC-detectable liquid products) characterization details are described further in the Supplementary Information section.

PMF and NMF analysis.

Positive matrix factorization (PMF) takes an input data matrix, $\boldsymbol{X}$(n x m), and separates the data into a time series matrix, $\boldsymbol{G}$ (n x p), and factor profile matrix, $\boldsymbol{F}$ (p x m), where p is the user-specified number of factors in the solution. A residual matrix, $\boldsymbol{E}$ (n x m), consists of the portion of the input data that cannot be captured by the factors to ensure mathematical continuity (equation 1).

$$\boldsymbol{X} = \boldsymbol{GF} + \boldsymbol{E} \quad [1]$$

The determination of the factors is achieved through the minimization of a function, $\boldsymbol{Q}$, which is the sum of uncertainty-weighted squared residuals:

$$Q = \sum_{n}^{i=1} \sum_{m}^{j=1} \frac{e_{ij}^2}{\sigma_{ij}^2}, \text{such that } g_{ik} \geq 0 \; and \; f_{kj} \geq 0 \quad [2]$$

where e_{ij} is the residual for a given value x_{ij}, and this is weighted by σ_{ij}, which corresponds to the uncertainty in the measured value. Constraining g_{ik} and $f_{kj}$to positive values ensures that nonsensical, negative solutions are not obtained. An idealized Q value (Q_{exp}), in which the degrees of freedom for the input are fully accounted for by the uncertainty in the measurements can be calculated as:

$$Q_{exp} = mn - p(m+n) \quad [3]$$

and the ratio of Q/Q_{exp} should approach one if both an appropriate number of factors are selected for the solution and the uncertainty estimates are appropriate.[60]

Prior to PMF analysis, preprocessing of the GC-MS data is required, and was carried out in a custom software package developed within Igor Pro (version 6.37, Wavemetrics, Inc.), which is available upon request and will be more broadly available following further development. The pre-processing package includes: 1) loading in the GC-MS datasets, 2) correcting for shifts in chromatographic retention times across samples, 3) subtracting mass spectral background contributions identified by blank samples, 4) binning sequential MS scans, 5) generating uncertainty estimates for each bin, and 6) scaling a given sample's abundance by a user-defined factor. For the retention time shift corrections, a relatively simple, linear shift was applied to each based upon the change in retention time of the n-decane internal standard. Blank samples, both with and without DMC in addition to MeOH solvent with internal standard, were used to attempt to remove the influence of instrumental artifacts. However, this approach is often insufficient to remove all artifacts (e.g. the presence of air or column bleed), due to sample-to-sample variation.[52] A chromatogram binning approach, described in detail previously,[52] was used to decrease the computational burden of solving the PMF model, and bins were composed of 5 sequential mass spectral scans. In total, 667 bins for each of the 30 reaction conditions were constructed from 3335 mass spectral scans, corresponding to retention times of 8.97 - 44.76 minutes for each sample. The included mass spectra, which comprise the columns of the input data matrix, ranged from 30-600 Th.

One of the most challenging aspects of conducting PMF analysis of an entire chromatogram's dataset is deriving appropriate uncertainty estimates (σ_{ij}) for all input data values. Building upon previous efforts to use PMF on datasets from GC-MS work,[52, 61] the uncertainties were calculated as:

$$\sigma_{ij} = \begin{cases} 2 \times MDL_{ij}, if \; x_{ij} < MDL_{ij} \\ \sqrt{\left(x_{ij} \times precision\right)^2 + \left(MDL_{ij}\right)^2}, if \; x_{ij} > MDL_{ij} \end{cases} \quad [4]$$

where MDL_{ij} is a retention time and mass-to-charge (*m/z*) ratio dependent detection limit estimate based on a sample blank chromatogram, and precision is an estimate of the reproducibility of the instrument (10% for this study).

Finally, the data for each sample was scaled based upon the integrated abundance of the n-decane internal standard. The peak integrations were performed in the Igor-based "TERN" software (version 2.1.6), which utilizes a peak fitting approach for quantification.[62] The PMF calculations were carried out in another custom software package (PMF Evaluation Tool version 3.00A) within Igor Pro, which utilizes the PMF2 solver.[51] To prevent an oversized impact from low abundance data within the matrix, *m/z* values with a signal-to-noise ratio (SNR) of less than 2 had their uncertainty values increased by a factor of 2, and values with SNR < 0.2 were excluded from the analysis entirely, as has been reported previously.[44] While the various Igor packages used are freely available, both the Igor Pro software and PMF2 solver require licenses for use.

The same input data matrix was entered into the NMF package within Matlab (MATLAB R2018a, MathWorks, Inc.), and a 13 factor solution was solved for with termination tolerances for both the residual and matrix elements of $1 \cdot e^{-6}$. No corresponding matrix containing the uncertainty estimates, as is used in PMF, is required for NMF.

Acknowledgements

Funding for this research was provided by the National Science Foundation Catalysis Program under award number CBET-1603692 and by the USDA National Institute of Food and Agriculture, Hatch project 1012359. We would also like to thank the Washington University's Engineering Communication Center for their help in revising this manuscript.

References

1 A. J. Ragauskas, C. K. Williams, B. H. Davison, G. Britovsek, J. Cairney, C. A. Eckert, W. J. Frederick, J. P. Hallett, D. J. Leak and C. L. Liotta, *Science*, 2006, **311**, 484-489.

2 A. J. Ragauskas, G. T. Beckham, M. J. Biddy, R. Chandra, F. Chen, M. F. Davis, B. H. Davison, R. A. Dixon, P. Gilna and M. Keller, *Science*, 2014, **344**, 1246843.

3 M. P. Pandey and C. S. Kim, *Chemical Engineering & Technology*, 2011, **34**, 29-41.

4 Y. Gao, M. Beganovic and M. B. Foston, in *Valorization of Lignocellulosic Biomass in a Biorefinery: From Logistics to Environmental and Performance Impact*, eds. R. Kumar, S. Singh and V. Balan, Nova Science Publishers, Inc., USA, 2016, ch. 8, pp. pp. 245-292.

5 W. O. Doherty, P. Mousavioun and C. M. Fellows, *Industrial Crops and Products*, 2011, **33**, 259-276.

6 C. A. Gasser, G. Hommes, A. Schäffer and P. F.-X. Corvini, *Applied Microbiology and Biotechnology*, 2012, **95**, 1115-1134.

7 G. W. Huber, S. Iborra and A. Corma, *Chemical Reviews*, 2006, **106**, 4044-4098.

8 A. G. Vishtal and A. Kraslawski, *BioResources*, 2011, **6**, 3547-3568.

9 W. Boerjan, J. Ralph and M. Baucher, *Annual review of plant biology*, 2003, **54**, 519-546.

10 M. Fache, B. Boutevin and S. Caillol, *ACS Sustainable Chemistry & Engineering*, 2016, **4**, 35-46.

11 C. Li, X. Zhao, A. Wang, G. W. Huber and T. Zhang, *Chemical Reviews*, 2015, **115**, 11559-11624.

12 R. Rinaldi, R. Jastrzebski, M. T. Clough, J. Ralph, M. Kennema, P. C. Bruijnincx and B. M. Weckhuysen, *Angewandte Chemie International Edition*, 2016, **55**, 8164-8215.

13 R. B. Santos, P. Hart, H. Jameel and H.-m. Chang, *Wood Based Lignin Reactions Important to the Biorefinery and Pulp and Paper Industries*, 2013.

14 M. Zaheer and R. Kempe, *ACS Catalysis*, 2015, **5**, 1675-1684.

15 P. J. Deuss and K. Barta, *Coordination Chemistry Reviews*, 2016, **306**, 510-532.

16 J. Zhang, H. Asakura, J. van Rijn, J. Yang, P. Duchesne, B. Zhang, X. Chen, P. Zhang, M. Saeys and N. Yan, *Green Chemistry*, 2014, **16**, 2432-2437.

17 P. J. Deuss, M. Scott, F. Tran, N. J. Westwood, J. G. de Vries and K. Barta, *Journal of the American Chemical Society*, 2015, **137**, 7456-7467.

18 L.-P. Xiao, S. Wang, H. Li, Z. Li, Z.-J. Shi, L. Xiao, R.-C. Sun, Y. Fang and G. Song, *ACS Catalysis*, 2017, **7**, 7535-7542.

19 J. Zhang, J. Teo, X. Chen, H. Asakura, T. Tanaka, K. Teramura and N. Yan, *ACS Catalysis*, 2014, **4**, 1574-1583.

20 Q. Song, F. Wang, J. Cai, Y. Wang, J. Zhang, W. Yu and J. Xu, *Energy & Environmental Science*, 2013, **6**, 994-1007.

21 Y. Ye, Y. Zhang, J. Fan and J. Chang, *Bioresource Technology*, 2012, **118**, 648-651.

22 K. Barta, T. D. Matson, M. L. Fettig, S. L. Scott, A. V. Iretskii and P. C. Ford, *Green Chemistry*, 2010, **12**, 1640-1647.

23 T. D. Matson, K. Barta, A. V. Iretskii and P. C. Ford, *Journal of the American Chemical Society*, 2011, **133**, 14090-14097.

24 C. M. Bernt, G. Bottari, J. A. Barrett, S. L. Scott, K. Barta and P. C. Ford, *Catalysis Science & Technology*, 2016, **6**, 2984-2994.

25 C. M. Bernt, H. Manesewan, M. Chui, M. Boscolo and P. C. Ford, *ACS Sustainable Chemistry & Engineering*, 2018, **6**, 2510-2516.

26 J. A. Barrett, Y. Gao, C. M. Bernt, M. Chui, A. T. Tran, M. B. Foston and P. C. Ford, *ACS Sustainable Chemistry & Engineering*, 2016, **4**, 6877-6886.

27 M. Chui, G. Metzker, C. M. Bernt, A. T. Tran, A. C. B. Burtoloso and P. C. Ford, *ACS Sustainable Chemistry & Engineering*, 2017, **5**, 3158-3169.

28 P. Chen, Q. Zhang, R. Shu, Y. Xu, L. Ma and T. Wang, *Bioresource technology*, 2017, **226**, 125-131.

29 A. D. Brittain, N. J. Chrisandina, R. E. Cooper, M. Buchanan, J. R. Cort, M. V. Olarte and C. Sievers, *Catalysis Today*, 2018, **302**, 180-189.

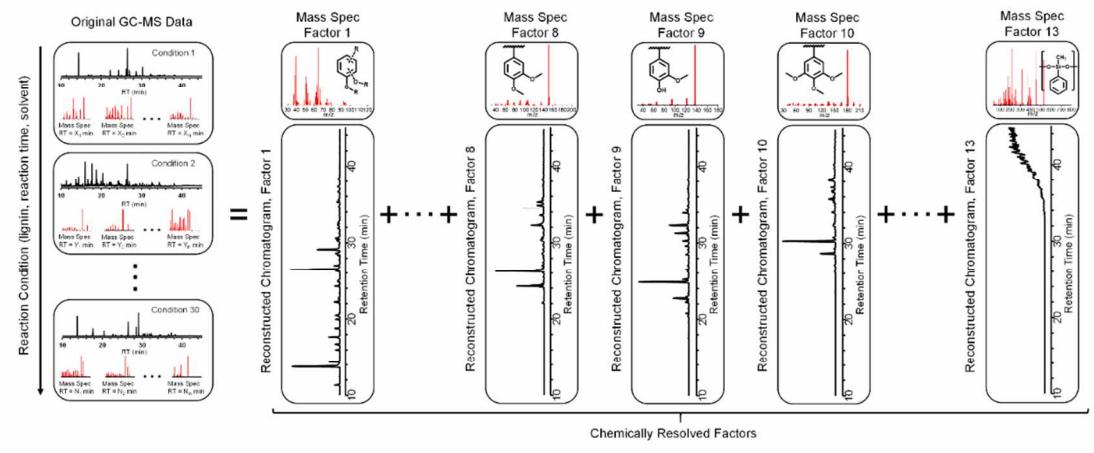
30 R. Shu, Y. Xu, L. Ma, Q. Zhang, C. Wang and Y. Chen, *Chemical Engineering Journal*, 2018.

31 J. Long, Y. Xu, T. Wang, Z. Yuan, R. Shu, Q. Zhang and L. Ma, *Applied Energy*, 2015, **141**, 70-79.

32 S. Van den Bosch, W. Schutyser, S.-F. Koelewijn, T. Renders, C. Courtin and B. Sels, *Chemical Communications*, 2015, **51**, 13158-13161.

33 W. Wanmolee, N. Laosiripojana, P. Daorattanachai, L. Moghaddam, J. Rencoret, J. C. del Río and W. O. Doherty, *ACS Sustainable Chemistry & Engineering*, 2018, **6**, 3010-3018.

34 H. Luo and M. M. Abu-Omar, *Green Chemistry*, 2018.
35 I. Klein, B. Saha and M. M. Abu-Omar, *Catalysis Science & Technology*, 2015, **5**, 3242-3245.
36 T. Ouyang, L. Wang, F. Cheng, Y. Hu and X. Zhao, *BioResources*, 2018, **13**, 3880-3891.
37 I. Kumaniaev, E. Subbotina, J. Sävmarker, M. Larhed, M. V. Galkin and J. S. Samec, *Green Chemistry*, 2017, **19**, 5767-5771.
38 J. Long, R. Shu, Z. Yuan, T. Wang, Y. Xu, X. Zhang, Q. Zhang and L. Ma, *Applied Energy*, 2015, **157**, 540-545.
39 A. Gredilla, S. Fdez-Ortiz de Vallejuelo, N. Elejoste, A. de Diego and J. M. Madariaga, *TrAC Trends in Analytical Chemistry*, 2016, **76**, 30-39.
40 G. Lubes and M. Goodarzi, *Chemical Reviews*, 2017, **117**, 6399-6422.
41 J. Lever, M. Krzywinski and N. Altman, *Nature Methods*, 2017, **14**, 641.
42 D. D. Lee and H. S. Seung, *Nature*, 1999, **401**, 788.
43 P. Paatero and U. Tapper, *Environmetrics*, 1994, **5**, 111-126.
44 P. Paatero and P. K. Hopke, *Analytica Chimica Acta*, 2003, **490**, 277-289.
45 R. Tauler, M. Viana, X. Querol, A. Alastuey, R. M. Flight, P. D. Wentzell and P. K. Hopke, *Atmospheric Environment*, 2009, **43**, 3989-3997.
46 J. Deng, Y. Zhang, Y. Qiu, H. Zhang, W. Du, L. Xu, Y. Hong, Y. Chen and J. Chen, *Atmospheric Research*, 2018, **202**, 23-32.
47 Y. Zushi and S. Hashimoto, *Analytical Chemistry*, 2018, **90**, 3819-3825.
48 M. R. L. Paine, J. Kim, R. V. Bennett, R. M. Parry, D. A. Gaul, M. D. Wang, M. M. Matzuk and F. M. Fernández, *PLOS ONE*, 2016, **11**, e0154837.
49 S. Tikole, V. Jaravine, V. Rogov, V. Dötsch and P. Güntert, *BMC Bioinformatics*, 2014, **15**, 46.
50 I. M. Ulbrich, M. R. Canagaratna, Q. Zhang, D. R. Worsnop and J. L. Jimenez, *Atmos. Chem. Phys.*, 2009, **9**, 2891-2918.
51 Q. Zhang, J. L. Jimenez, M. R. Canagaratna, I. M. Ulbrich, N. L. Ng, D. R. Worsnop and Y. Sun, *Analytical and Bioanalytical Chemistry*, 2011, **401**, 3045-3067.
52 Y. Zhang, B. J. Williams, A. H. Goldstein, K. Docherty, I. M. Ulbrich and J. L. Jimenez, *Aerosol Science and Technology*, 2014, **48**, 1166-1182.
53 Y. Zhang, B. J. Williams, A. H. Goldstein, K. S. Docherty and J. L. Jimenez, *Atmos. Meas. Tech.*, 2016, **9**, 5637-5653.
54 P. Paatero, S. Eberly, S. Brown and G. Norris, *Atmospheric Measurement Techniques*, 2014, **7**, 781-797.
55 G. S. Frysinger, R. B. Gaines, L. Xu and C. M. Reddy, *Environmental Science & Technology*, 2003, **37**, 1653-1662.
56 C. Lindfors, E. Kuoppala, A. Oasmaa, Y. Solantausta and V. Arpiainen, *Energy & Fuels*, 2014, **28**, 5785-5791.
57 Q. Bu, H. Lei, A. H. Zacher, L. Wang, S. Ren, J. Liang, Y. Wei, Y. Liu, J. Tang, Q. Zhang and R. Ruan, *Bioresource Technology*, 2012, **124**, 470-477.
58 G. S. Macala, T. D. Matson, C. L. Johnson, R. S. Lewis, A. V. Iretskii and P. C. Ford, *ChemSusChem*, 2009, **2**, 215-217.
59 T. S. Hansen, K. Barta, P. T. Anastas, P. C. Ford and A. Riisager, *Green Chemistry*, 2012, **14**, 2457-2461.
60 P. Paatero, P. K. Hopke, X.-H. Song and Z. Ramadan, *Chemometrics and Intelligent Laboratory Systems*, 2002, **60**, 253-264.
61 B. J. Williams, A. H. Goldstein, N. M. Kreisberg, S. V. Hering, D. R. Worsnop, I. M. Ulbrich, K. S. Docherty and J. L. Jimenez, *Atmos. Chem. Phys.*, 2010, **10**, 11577-11603.
62 G. Isaacman-VanWertz, D. T. Sueper, K. C. Aikin, B. M. Lerner, J. B. Gilman, J. A. de Gouw, D. R. Worsnop and A. H. Goldstein, *Journal of Chromatography A*, 2017, **1529**, 81-92.



Positive matrix factorization analysis significantly enables the use of gas chromatography-mass spectrometry to elucidate complex catalytic reaction networks, specifically, for lignin catalysis.