Understanding the Chemistry of Conversion of Model Compounds and Biomass by Hydrous Pyrolysis Based on Spectroscopic Data Using Data Fusion, Data Mining, and Chemometrics

by

Fereshteh Sattari

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

Doctor of Philosophy

Process Control

Department of Chemical and Materials Engineering

University of Alberta

Abstract

Encoding the presence of multispecies in a complex system, difficulty in characterizing the physical constituents of the products with analytical instruments along with developing the causality or modeling between these groups are some major challenges in a complex system such as biomass conversion. To overcome these challenges and provide insights into the potential chemistry and reaction mechanism of the hydrous pyrolysis (HTL) of biomass and its main products (cellulose and lignin), this research employed some machine learning methods such as data mining and data fusion techniques. Therefore, Levoglucason, 2-Phenoxyethyl benzene (representing cellulose and lignin, respectively), a physical mixture of these model components, and Monterey pine whole biomass underwent 108 HTL reactions in the presence of hot water and catalysts under different conditions. For characterization of the produced bio-oil two spectroscopic techniques were used, Fourier transform-infrared (FTIR) and Proton nuclear magnetic resonance (^{1}H NMR). In the process of knowledge discovery from hidden interesting patterns in the large data sets provided by these spectroscopic techniques, this research employed data fusion. The aim of data fusion is to develop experimentally and computationally sensible models from spectroscopic data with the advantages of a consistent combination of absorbance across wavenumbers (variables) with demonstrable improvement in the reaction network structure for integrating multiple data sets provided by FTIR and ^{1}H NMR. Developed model has the advantage of decreasing the error while processing large-scale reactions along with increasing the model performance by factoring in complementary information. The final fused data set was used for data clustering by using the Bayesian hierarchical clustering (BHC). In a large data set, while the traditional hierarchical clustering algorithms have the difficulties of deciding which distance metric to choose, BHC has the advantage of computing the marginal

likelihoods in order to decide which clusters to merge and to avoid overfitting. After grouping the wavenumbers into different clusters, the Bayesian network learning approach (BN) was applied to develop the optimal reaction network. To identify the optimal structure of the network, three different optimization approaches are applied: two greedy search-and-score algorithms called tabu and hill climbing, and a hybrid algorithm called the max-min hill climbing (MMHC).

In spite of the fact that spectroscopic techniques such as FTIR can provide useful information relating to the structure of the compound, it has a weakness of having a high dimensional space of wavenumbers which is sometimes difficult to be interpreted or analyzed. To resolve this issue, chemometric methods can be applied. In these methods, statistical or mathematical techniques have been used to collect the required information regarding the objects of interest in the data. Self-modeling multivariate curve resolution (SMCR) is a popular example of a chemometric technique. The reason to employ this method is to obtain a set of pseudo-components and their spectra and to use them to develop the reaction network. SMCR is a very useful tool for the elucidation of the multi-component phenomena in complex chemical systems such as biomass conversion. Developed algorithm can be applied for real-time analysis of many complex reacting systems and mixtures because it provides quantitative tracking of changes in the process and can be used for compositional control. In addition, it also acts as a screening method to propose hypotheses about reaction mechanisms in complex reacting mixtures. Moreover, for online monitoring of species conversion in these kinds of complex reactions, this research computed the concentrations of these pseudocomponents over the number of samples. The application of this trace makes it useful in the online monitoring of species conversion by integrating it with a suitable control strategy that adjusts process conditions to maximize the yield of the desired product.

Preface

Chapter 2 of the thesis is a paper manuscript to be submitted as Fereshteh Sattari, Dereje Tefera, Kaushik Sivaramakrishnan, Arno de Klerk and Vinay Prasad, "Postulating Pseudo-reaction Networks for the Conversion of Levoglucosan in Hydrous Pyrolysis Using Spectroscopic Data & Self-Modeling Multivariate Curve Resolution".

Chapter 3 of the thesis is a paper manuscript to be submitted as Fereshteh Sattari, Arno de Klerk and Vinay Prasad, "Application of Data Combination and Data Mining Techniques to Investigate the Chemistry of Cellulose and Lignin Derivatives in Hydrous Pyrolysis".

Chapter 4 of the thesis is a paper manuscript to be submitted as Fereshteh Sattari, Arno de Klerk and Vinay Prasad, "Developing the Most Probable Reaction Networks for Thermochemical Conversion of a Physical Mixture of Levoglucosan and 2-Phenoxyethyl benzene by Using Data Fusion and Chemometric Techniques (SMCR-ALS)".

Chapter 5 of the thesis is a paper manuscript to be submitted as Fereshteh Sattari, Dereje Tefera, Arno de Klerk and Vinay Prasad, "The Application of Data Fusion and Chemometric Techniques for Understanding the Conversion of Biomass by Hydrous Pyrolysis Based on Spectroscopic Data".

Model development and computer programming for data fusion, analysis and manuscript composition as well as the other chapters of the thesis were completed by myself with the supervision of Prof. Vinay Prasad. For manuscript composition and concept information for chapters 2-5, Prof. Arno de Klerk was also supervised. Models development for the Bayesian Network and SMCR which were initially done by Dereje Tefera were improved by myself. All the experiments were performed by myself under Prof. Vinay Prasad and Prof. Arno de Klerk supervision. For performing the experiments, I was trained by Kaushik Sivaramakrishnan in Prof. Arno de Klerk's chemical labs.

Acknowledgements

In connection with this work, I wish to express my sincere gratitude to Dr. Prasad for providing me the opportunity to conduct this work and pursue my doctorate program. His guidance, advice, and patience during my study period has been truly appreciated. I genuinely would like to appreciate Dr. de Klerk who carefully guided and supported me through my research.

My special thanks and appreciations also go to Dr. Lianne Lefsrud, Dr. Samir H. Mushrif, and Lily Laser who guided and believed in me during my doctorate program,

I am grateful to all of my colleagues in the Dr. Vinay Prasad's and Dr.de klerk's groups who have willingly helped me out with their abilities.

I also wish to thank Mohammad, my husband, whose love and supportive is with me in whatever I pursue, and Amirreza, my son, who provides me his unending inspiration.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

LIST OF ABBRIVIATION

- Acetaldehyde AA
- Alternative least square ALS
- Bayesian hierarchical clustering BHC
- Bayesian information criterion BIC
- Bayesian network BN
- Broad B
- Density Based Spatial Clustering of Applications with Noise DBSCAN
- Directed acyclic graph DAG
- Dirichlet process mixture model DPM
- Evolving factor analysis EFA
- Fourier Transform-Infrared FTIR
- Full weight at half maximum FWHM
- Furfural FF
- Gas chromatography-mass spectroscopy GC-MS
- Glucose GP
- Hill climbing HC
- Hot compressed water HCW
- Hydrous pyrolysis/Hydrothermal liquefaction HTL
- Hydroxyacetic acid vinyl ester HAVE
- Indicator function IND
- Lack of fit LOF
- Levogluccosan LG
- Max-min hill climbing MMHC
- Medium M
- Minute min
- Narrow N
- Near-infrared NIR
- Ordering points to identify the clustering structure OPTICS
- Orthogonal projection analysis OPA
- Part per billion ppb
- Part per million ppm
- Percentage of explained variance R^2
- Physical mixture PM
- Propanedialdehyde PD
- Proton nuclear magnetic resonance ^{1}H NMR
- Root mean square RMS
- Self-modeling multivariate curve resolution SMCR
- Sharp Sh
- Strong S

• Sum square	SSQ
• Temperature	T
• Time	t
• Ultraviolet	UV
• Weak	W
• 1,2-dihydroxyethene	1,2-DE
• 1-pentene3,4-dione	1-P-3,4-D
• 2,3-dihydroxypropanal	2,3-DP
• 2-Phenoxyethyl benzene	PEB
• 5-hydroxymethylfurfural	5-HMF

LIST OF SYMBOLS

- Concentration C
- Data D
- Directed acyclic graph G
- Error E
- Mean intensity μ
- Normal distribution N
- Random variables V
- Random variables Φ
- Spectral profile S
- Transpose T

1. INTRODUCTION

1.1 BACKGROUND

Biomass can be used in several ways for the production of electricity, heat, and steam; and it can be converted into synthesis gases and liquid fuels by undergoing biochemical or thermochemical processing.[1] Any materials from a biological origin with organic components such as agricultural crops, trees, and grasses would be considered as biomass.[2] Biomass has a lot of availability and its renewability makes it a nearly universal feedstock. As far back as the 19th century, biomass was the world's most important energy supply,[3] until oil took over. In modern times our culture is coming full circle for many reasons. The impact of fossil oil on our environment is undeniable and biomass is being considered the preferred energy source by many researchers. [4, 5] Using biofuels reduces greenhouse gases considerably since its carbon dioxide production compared to crude oil is hard to ignore. This is due to the fact that biomass is carbon dioxide neutral, how much it accrued during its lifespan is equal to what it puts out.[6] The environmental benefits along with the increase in demand make biomass in a great place to make a comeback.[7] For all significant potential benefits of biomass, the US Department of Energy created an outline in 1998 on using renewable resources like crop plants for future energy suppliers.[8] The goal was that 10% of the basic chemical building blocks by 2020, and 50 % by 2050 should be provided from renewable sources such as biomass (Figure 1-1). This outline proposed the major opportunities to increase the use of renewable resources.

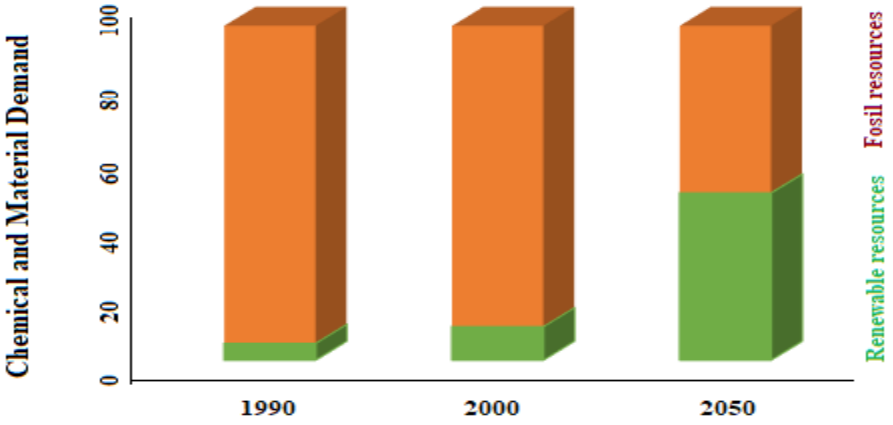


Figure 1-1. Renewable targets set by the U.S. Department of Energy.[8]

From the chemical point of view, 99% of biomass is made of cellulose, hemicelluloses and lignin. These components are created from the polymerization of monosaccharides such as fructose, galactose and glucose (from plant leaves).[9]

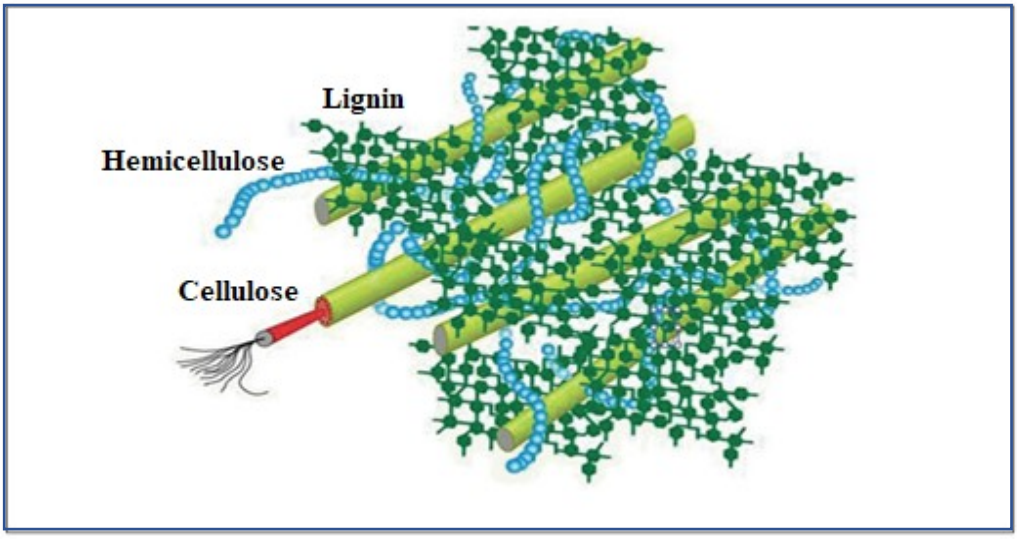


Figure 1-2. Biomass components (adapted from Wang et al). [2]

Among different processes for biomass conversion into fuels, energy, and chemicals, thermochemical and biochemical processes are the most important ones (Figure 1-3).[10] This dissertation investigates a specific kind of thermochemical process which is called hydrothermal liquidation (HTL). This process utilizes an aqueous medium (subcritical water at high temperature) as both solvent and reactant mostly to overcome the activation energy needed for the chemical disintegration of biomass macrostructures.

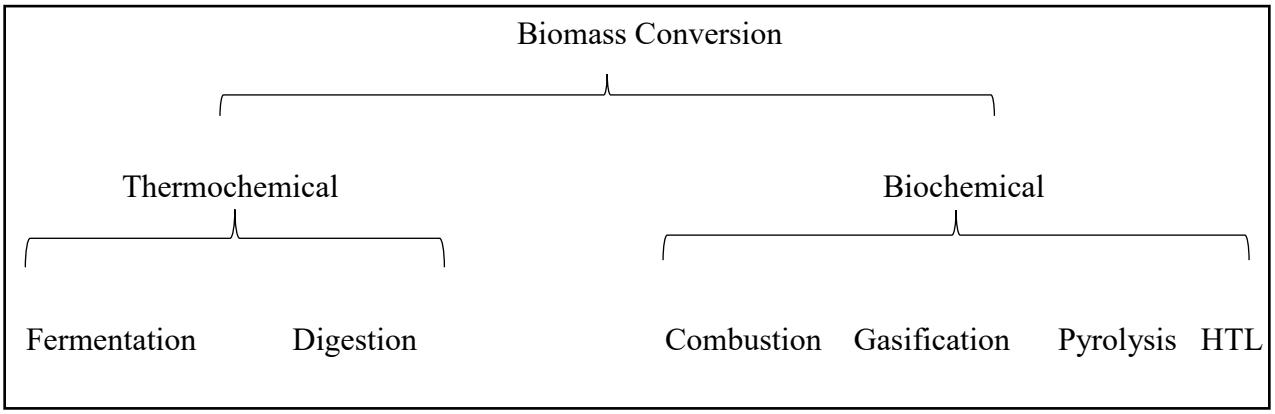


Figure 1-3. Common biomass conversion processes.

Generally, the basic process of HTL of biomass consists of the decomposition of biomass's main components, hydrolysis of large molecules, and reformation of produced molecules.[2] For the past few decades, research has been acquiring a broad knowledge of these processes even though the exact mechanism, reaction modeling, and real-time monitoring of these complicated processes are still unknown.[11, 12]

There are usually two classical methods for reaction modeling of a system: Empirical and Mechanistic modeling techniques. [13] While these methods can be considered reasonable for a simple reaction, their applications in a complex system such as biomass conversion are impractical. Empirical models require prior assumption and lots of data while mechanistic models are computationally expensive and need comprehensive domain knowledge about the process which is difficult to obtain in a complex process. In other words, in biomass conversion, there is still a major challenge related to developing an algorithm that could combine expert knowledge and experimental data to develop the most likely reaction network.[14, 15]

Regarding quantitative analysis of chemical species and real-time online monitoring of a system, recently, spectroscopy methods such Fourier transform infrared (FTIR) spectroscopy,[16] Ultraviolet (UV) spectroscopy,[17] Near-infrared (NIR) spectroscopy,[18] and Proton nuclear magnetic resonance (^{1}H NMR) spectroscopy [19] have been widely used. In spite of the fact that these spectroscopic techniques can provide useful information relating to the structure of the compound, they have a weakness of having a high dimensional space of wavenumbers which is sometimes difficult to be interpreted or analyzed.[20]

1.2. RESEARCH OBJECTIVES

To overcome the challenge related to develop the reaction modeling and mechanism for biomass conversion, this research has employed two powerful machine learning techniques: data fusion and data mining. A dataset from a single instrument may be limited in providing a comprehensive analysis of the investigated sample. To work with these limitations, data fusion based on complementary techniques is a dependable method.[21] Integrating or fusing data from multiple instruments can provide more accurate information with less error for classification or less uncertainty for predictions compared to the results provided from a single technique. [22] Therefore, the aim of the dissertation is to develop experimentally and computationally sensible

models from spectroscopic data with the advantages of developing an algorithm for integrating multiple data sets provided by FTIR and ^{1}H NMR.

Moreover, to understand the large and complex data set provided by data fusion, data mining was also explored. In general, data mining was employed for prediction and description. [23] The first data mining technique used was Bayesian hierarchical clustering (BHC)[24] whose application was for data clustering of the spectral information. In a large data set, for the traditional hierarchical clustering algorithm, difficulties arise when they must decide which distance metric to choose. On the other hand, BHC has the advantage of computing the marginal likelihoods in order to decide which clusters to merge and to avoid overfitting.[25] After grouping the wavenumbers into different clusters, the Bayesian network learning approach (BN)[26] was applied to develop the optimal reaction network. As indicated by Heckerman, to develop the reaction network an increase in the understanding of relationships and patterns between variables is essential to predict the future because of which the BN learning approach is the most effective method.[27] This technique has the advantage of integrating the observed data, even in the case of missing data, with the prior knowledge in order to generate the most probable network.

To resolve the issue related to interpretation of high data sets provided by spectroscopy methods, chemometric methods seem reasonable. In these methods, statistical or mathematical techniques have been used to collect the required information regarding the objects of interest in the data. Self-modeling multivariate curve resolution (SMCR) [28] is a popular example of a chemometric technique. The reason to employ this method is to obtain a set of pseudo-components and their spectra and to use them to develop the reaction network. SMCR is a very useful tool for the elucidation of the multi-component phenomena in complex chemical systems such as biomass conversion.[29] The algorithms developed can be applied for real-time analysis of many complex reacting systems and mixtures because it provides quantitative tracking of changes in the process and can be used for compositional control.[30] In addition, it also acts as a screening method to propose hypotheses about reaction mechanisms in complex reacting mixtures.

Along with a deep understanding of a complex reaction chemistry, there are some other objectives that this study has tried to achieve. One objective could be beneficial for process optimization (by predicting the effect of interaction qualitatively and quantitatively for real time analysis of the process), and the next one is applicable for upgrading processes (by focusing on

selectivity towards product compounds). As a result, the aim of this research is to provide a theoretical basis to handle uncertainty that provides insights into the potential chemistry and reaction mechanism of the biomass conversion process.[31]

1.3. THESIS OUTLINE

This thesis includes six chapters, and the main goals of chapters 2-5 are the applications of data fusion and data mining to develop the most probable reaction networks for bio-oils provided from HTL of levoglucosan (chapter 2), 2-Phenoxyethyl benzene (chapter 3), A physical mixture of model components representing cellulose and lignin (chapter 4), and Monterey pine whole biomass (chapter 5). Furthermore, these chapters (2-5) will include developed SMCR-ALS algorithms in order to acquire the spectral profiles of existing pseudocomponents and their concentrations, which can be used to hypothesize a reaction network, and for real time analysis of many complex reacting systems and mixtures. Chapter 6 summarizes the finding of the research.

1.4. REFERENCES

1. Wang Y, He T, Liu K, Wu J, Fang Y. From biomass to advanced bio-fuel by catalytic pyrolysis/hydro-processing: hydrodeoxygenation of bio-oil derived from biomass catalytic pyrolysis. *Bioresour Technol*. 2012;108:280-284.
2. Wang S, Zhongyang L. *Pyrolysis of Biomass*. Science Press Beijing; 2017.
3. Crocker M. Thermochemical Conversion of Biomass to Liquid Fuels and Chemicals. *Cambridge R Soc Chem*. 2010.
4. Pang S. Advances in thermochemical conversion of woody biomass to energy, fuels and chemicals. *Biotechnol Adv*. 2018;(November):0-1. doi:10.1016/j.biotechadv.2018.11.004
5. Libra JA, Ro KS, Kammann C, et al. Hydrothermal carbonization of biomass residuals : a comparative review of the chemistry , processes and applications of wet and dry pyrolysis Hydrothermal carbonization of biomass residuals : a comparative review of the chemistry , processes and applicati. 2014;7269. doi:10.4155/bfs.10.81
6. E.J. Soltes TJE. Pyrolysis, in: I.S. Goldstein (Ed.), Organic Chemicals from Biomass. *CRC Press Boca Rat*. 1981:63-99.
7. Lopez Barreiro D, Prins W, Ronsse F BW. Hydrothermalliquefaction (HTL) of microalgae for biofuel production: state of the artreview and future prospeects. *Biomass Bioenergy*. 2013;53:113–127.
8. McLaren J. *The Technology Roadmap for Plant/Crop-Based Renewable Resources 2020*. DIANE Publishing; 1999.
9. Stefanidis SD, Kalogiannis KG, Iliopoulou EF, Michailof CM, Pilavachi PA, Lappas AA. A study of lignocellulosic biomass pyrolysis via the pyrolysis of cellulose, hemicellulose and lignin. *J Anal Appl Pyrolysis*. 2014;105:143-150. doi:10.1016/j.jaap.2013.10.013
10. Kruse A, Gawlik A. Biomass conversion in water at 330-410 °C and 30-50 MPa. Identification of key compounds for indicating different chemical reaction pathways. *Ind Eng Chem Res*. 2003;42(2):267-279. doi:10.1021/ie0202773
11. Anca-Couce A. Reaction mechanisms and multi-scale modelling of lignocellulosic biomass pyrolysis. *Prog Energy Combust Sci*. 2016;53(2016):41-79. doi:10.1016/j.pecs.2015.10.002
12. Peters JF, Banks SW, Bridgwater A V., Dufour J. A kinetic reaction model for biomass pyrolysis processes in Aspen Plus. *Appl Energy*. 2017;188:595-603. doi:10.1016/j.apenergy.2016.12.030
13. Stagni A, Cuoci A, Frassoldati A, Faravelli T, Ranzi E. Lumping and Reduction of Detailed Kinetic Schemes: An Effective Coupling. *Ind Eng Chem Res*. 2014;53(22):9004-9016.
14. Tefera DT, Yañez Jaramillo LM, Ranjan R, Li C, De Klerk A, Prasad V. A Bayesian Learning Approach to Modeling Pseudoreaction Networks for Complex Reacting Systems: Application to the Mild Visbreaking of Bitumen. *Ind Eng Chem Res*.

2017;56(8):1961-1970. doi:10.1021/acs.iecr.6b04437

15. Walker TW, Motagamwala AH, Dumesic JA, Huber GW. Fundamental catalytic challenges to design improved biomass conversion technologies. *J Catal*. 2018;369:518-525. doi:10.1016/j.jcat.2018.11.028

16. Bassbasi, M.; De Luca, M.; Ioele, G.; Oussama, A.; Ragno G. Prediction of the Geographical Origin of Butters by Partial Least Square Discriminant Analysis (PLS-DA) Applied to Infrared Spectroscopy (FTIR) Data. *J Food Compos Anal*. 2014;33(2):210-215.

17. Li, Y.; Zhang, J.; Liu, H. G.; Jin, H.; Wang, Y. Z.; Li T. Discrimination of Storage Periods for Macrocybe Gigantea (Massee) Pegler & Lodge Using UV Spectral Fingerprints. *Czech J Food Sci*. 2015;33(5):441-448.

18. Shen, T. T.; Zou, X. B.; Shi, J. Y.; Li, Z. H.; Huang, X. W.; Xu, Y. W.; Chen W. Determination Geographical Origin and Flavonoids Content of Goji Berry Using Near-Infrared Spectroscopy and Chemometrics. *Food Anal Methods*. 2016;9(1):68–79.

19. Fernández C, Callao MP, Larrechi MS. Talanta UV – visible-DAD and 1 H-NMR spectroscopy data fusion for studying the photodegradation process of azo-dyes using MCR-ALS. *Talanta*. 2013;117:75-80. doi:10.1016/j.talanta.2013.08.004

20. Wong KC. Review of Spectrometric Identification of Organic Compounds. *J Chem Educ*. 2015;92(10):1602-1603.

21. Li, Y.; Zhang, J.; Li, T.; Liu, H. G.; Wang YZA. Comprehensive and Comparative Study of Wolfiporia Extensa Cultivation Regions by Fourier Transform Infrared Spectroscopy and Ultra-Fast Liquid Chromatography. *PLoS One*. 2016;11(12).

22. Borràs, E.; Ferré, J.; Boqué, R.; Mestres, M.; Aceña, L.; Busto O. Data Fusion Methodologies for Food and Beverage Authentication and Quality assessment-A Review. *Anal Chim Acta*. 2015;891:1-14.

23. Kantardzic M. *Data Mining : Concepts, Models, Methods, and Algorithms.* Wiley-IEEE Press; 2011.

24. Nia VP, Davison AC. High-Dimensional Bayesian Clustering with Variable Selection: The R Package bclust. *J Stat Softw*. 2012;47(5):1-22. doi:http://dx.doi.org/10.18637/jss.v047.i05

25. Heller K GZ. Bayesian hierarchical clustering. . *ICML '05 Proc 22nd Int Conf Mach Learn*. 2005:297–304.

26. Friedman, Nir, Nachman, Iftach. Peer D. Learning Bayesian Network Structure from Massive Datasets: The "Sparse Candidate" Algorithm. In: *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*. ; 1999:206-215.

27. Heckerman D. A tutorial on learning with Bayesian networks, in Learning in Graphical, M.J. Models. *MIT Press Cambridge*. 1997;1:79-119.

28. Rahimdoust Mojdehi N, Sawall M, Neymeyr K, Abdollahi H. Investigating the effect of

flexible constraints on the accuracy of self-modeling curve resolution methods in the presence of perturbations. *J Chemom*. 2016;30(5):252-267. doi:10.1002/cem.2787

29. Jiang J, Ozaki Y. SELF-MODELING CURVE RESOLUTION (SMCR): PRINCIPLES , TECHNIQUES , AND APPLICATIONS. 2007;4928. doi:10.1081/ASR-120014359

30. Awa K, Okumura T, Shinzawa H, Otsuka M, Ozaki Y. Self-modeling curve resolution (SMCR) analysis of near-infrared (NIR) imaging data of pharmaceutical tablets. *Anal Chim Acta*. 2008;619(1):81-86. doi:10.1016/j.aca.2008.02.033

31. Zhang S, Ding HUA, Wang X. Research and Application of Structure Learning Algorithim. In: *Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an,.* ; 2003:2-5.

2. POSTULATING PSEUDO-REACTION NETWORKS FOR THE CONVERSION OF LEVOGLUCOSAN IN HYDROUS PYROLYSIS USING SPECTROSCOPIC DATA AND SELF-MODELING MULTIVARIATE CURVE RESOLUTION

A version of this chapter will be submitted as; Fereshteh Sattari, Dereje Tefera, Kaushik Sivaramakrishnan, Arno de Klerk and Vinay Prasad, “Postulating Pseudo-reaction Networks for the Conversion of Levoglucosan in Hydrous Pyrolysis Using Spectroscopic Data & Self-Modeling Multivariate Curve Resolution”.

2.1. BACKGROUND

While fossil fuels have been the main source of energy for the world,[1] concerns about the environmental impact of their use,[2] including greenhouse gas emissions, has led to an interest in renewable resources such as biomass. Biomass conversion is usually pursued through thermochemical or biochemical processes.[3] The thermochemical route involving pyrolysis produces bio-oil, which can potentially be used as fuel after upgrading. Biomass pyrolysis is complex and is influenced by many factors,[4] and presents challenges for process monitoring and optimization. Researchers have represented the conversion of biomass through pyrolysis in terms of the conversion of its three major components: hemicellulose, cellulose, and lignin.[5]

Depending on the source, cellulose can constitute more than 50% of biomass; therefore, the chemistry of its pyrolysis has gained much attention from researchers, with many reaction mechanisms and associated kinetic models having been proposed.[6,7] Despite these efforts, there are still many challenges in developing a clear and complete understanding of the mechanism.[8] Using model compounds such as glucose or levoglucosan can be helpful in the study of cellulose pyrolysis mechanisms.[9] Note that levoglucosan as a model compound does not represent cellulose, but is an important intermediate in one of the pathways of cellulose decomposition.[10] Even using model compounds, however, does not always provide an unambiguous picture of the reaction mechanism. Hosoya et al. propose pathways for the pyrolysis of levoglucosan in two environments, but do not provide direct evidence to prove these pathways.[11] Rass and his coworkers also believe that a key challenge for chemists in the twenty-first century is predicting a reaction network for the pyrolysis and catalytic conversion of biomass;[12] which would enable optimal design, online monitoring, optimal dynamic operation and control. While multiple analytical techniques are often used to probe the chemistry, developing reaction networks is often seen and treated as an art. However, data analytics and machine learning may be able to provide a systematic framework to probe reaction chemistry and online monitoring.

The focus of this research is to explore the ability of machine learning and chemometric methods to uncover the reaction chemistry in pyrolysis based purely on spectroscopic data that could potentially be obtained in real-time. This would serve two purposes: one, to hypothesize meaningful reaction networks to aid understanding of the process, and two, be usable developing a real-time monitoring and control scheme for the process. In this work, we outline the

development of a reaction network based solely on Fourier Transform Infrared (FTIR) spectroscopic data. The development of a systematic approach to accomplish this is the main contribution of this work, and the system chosen to demonstrate this is the hydrous pyrolysis of levoglucosan. The methods used include principal component analysis (PCA), Bayesian hierarchical clustering (BHC), Bayesian networks (BN) and Self-modeling curve resolution (SMCR).

Despite the fact that spectroscopic techniques such as FTIR can provide useful information relating to chemical species, the high dimensionality of the data and the presence of overlaps in the signatures of different species makes it difficult to be interpreted or analyzed in many cases.[13] Chemometric methods are used to extract information about important features in the data, and SMCR is one such technique that we have explored in this work. The advantages of using this technique are that it requires a smaller amount of quantitative data to be effective and that it does not require vast prior knowledge of the system.[14] Therefore, to predict the reaction chemistry of converting biomass or related model compounds into products, SMCR can be used to characterize the reacting mixture in terms of pseudocomponents. Alternately, BHC can be used to identify clusters of peaks (that can be related to specific functional groups) which can also be viewed as pseudocomponents.

After applying BHC and SMCR, the next step is developing a reaction network between clusters or pseudocomponents. A Bayesian network (BN) is a directed graph that can be applied to develop a model representing the quantitative strength of the connections between clusters or functional groups found in the previous steps.[15,16] The literature indicates that when the data length is short with a small number of samples, the Bayesian method is an effective approach to creating a reaction network.[17]

2.2. MATERIALS

Levoglucosan (LG) was selected as the model compound in this study because it is an important intermediate in cellulose conversion, and because it has a similar molecular structure and almost the same density. LG's structure contains a pyranose ring, including one oxygen and five carbon atoms. This ring is attached to one extra bridge of the anhydro ring. Many researchers have also confirmed that LG is the most abundant product in the primary reaction from cellulose pyrolysis

as it undergoes decomposition reactions to produce lighter compounds.[7,18] Sulfuric acid (0.05 M) and sodium hydroxide (1 M) were used as catalysts in the hydrous pyrolysis process.

2.3. METHODS

2.3.1. Hydrous Pyrolysis

Pyrolysis of biomass to bio-oil has received significant attention.[19,20] Available technologies in this field include enzymatic hydrolysis, diluted and concentrated acid pyrolysis, and alkaline hydrolysis.[21] In the past two decades, several studies and developments have also focused on biomass hydrolysis in hot compressed water (HCW) conditions.[22,23] HCW can be used as a solvent and reactant in many processes for biomass use, including hydrothermal degradation for bio-oil and subcritical water hydrolysis.[24] We employ this process for the thermal conversion of LG, using subcritical water as the solvent and conducting the reactions in a batch micro-reactor. In all, twenty-seven different conditions varying temperature, catalyst and reaction time were used.

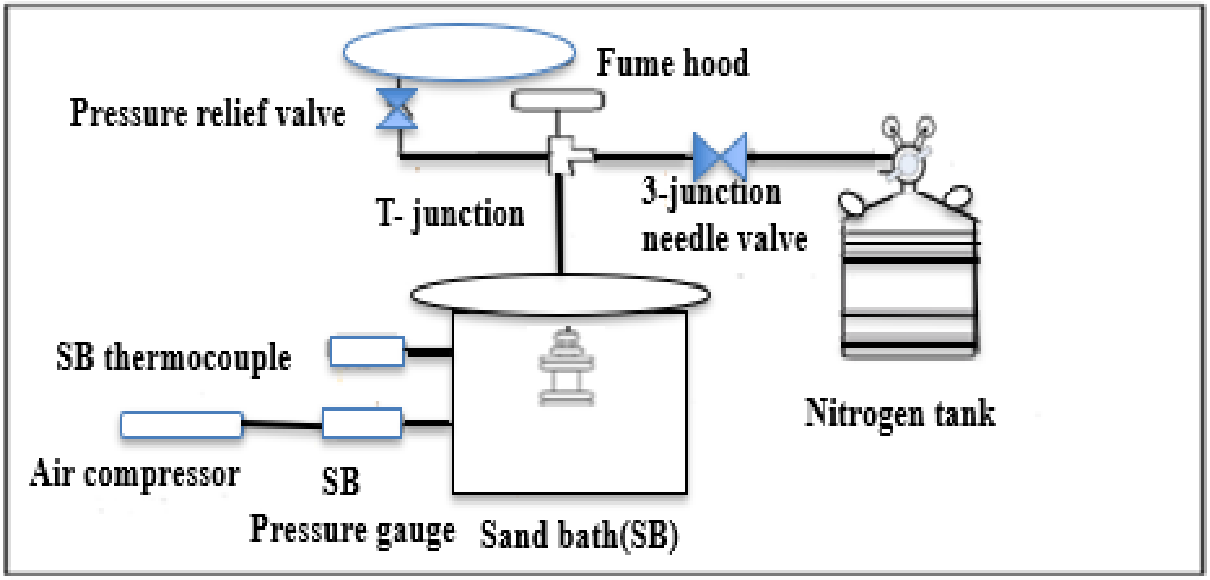


Figure 2-1. Schematic of the batch micro-reactor set-up for the hydrous pyrolysis of LG.

Figure 2-1 presents the experimental setup for the hydrous pyrolysis of LG. The initial pressure was kept constant at 0.1 MPa by closing the pressure relieve valve. The stainless-steel micro

batch reactor (24 cm long and 2.1 cm in diameter) was equipped with a thermocouple and heated to the requisite temperature; the temperature was maintained for the desired reaction time, after which the reactor was allowed to cool to 100°C before quenching to room temperature with water. Ranges of temperature and residence time were selected based on the literature, and we focused on their effect on obtaining the maximum liquid with the desired characteristics,[25] the operating conditions used are listed in Table 2-1. The volume ratio of LG to the medium was 1:10, and the products were measured in a glass beaker and stored in a glass vial for later analysis (applying FTIR and ^{1}HNMR techniques).

Table 2-1. Operating conditions for temperature and residence time for LG conversion in a stainless-steel micro batch-reactor.

Medium	T (°c)	t1(min)	t2(min)	t3(min)
Subcritical water	150	15	35	55
Subcritical water	200	15	35	55
Subcritical water	250	15	35	55
Sulfuric acid	150	10	20	30
Sulfuric acid	180	10	20	30
Sulfuric acid	240	10	20	30
Sodium hydroxide	100	10	60	110
Sodium hydroxide	150	10	60	110
Sodium hydroxide	200	10	60	110

T=Temperature, t=Time, min=minutes.

2.3.2. Fourier Transform Infrared Spectroscopy (FTIR)

After the hydrous pyrolysis reaction was completed, samples were collected for characterization. The infrared spectra of the product samples were collected using an ABB MB 3000 FTIR spectrometer. The spectra were collected at 2 cm^{-1} resolutions over the spectral range 4000 to 600 cm^{-1}. The analyses were performed on the neat samples using a Pike MIRacle ^{TM} attenuated total reflectance attachment, and the spectra for the 27 samples are shown in Figure 2-2. The hand book of spectroscopic data was used for preliminary identification of the functional groups.[26]

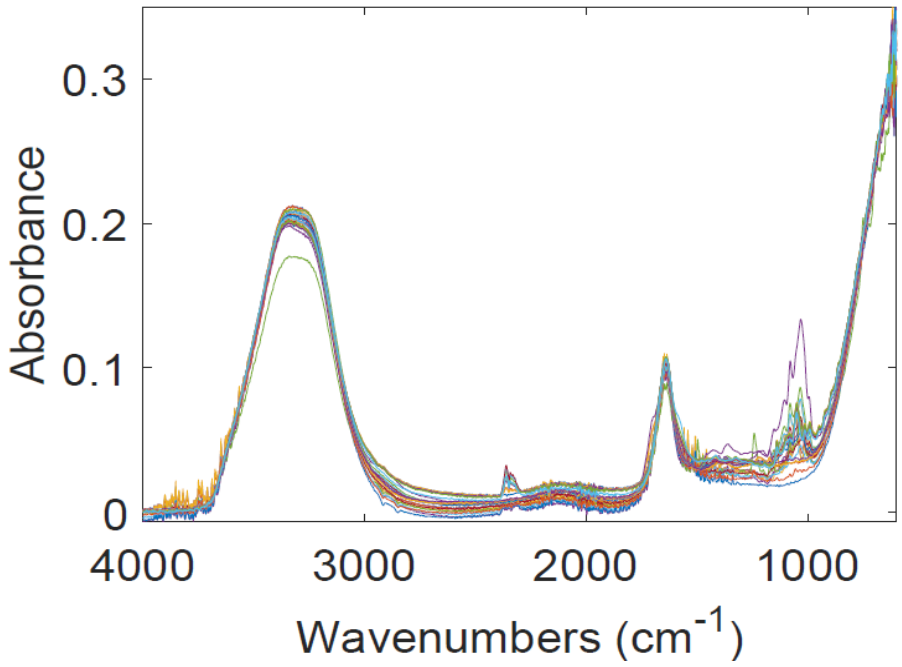


Figure 2-2. Twenty-seven FTIR spectra of the products of HTL of LG in various conditions.

2.3.3. Proton Nuclear Magnetic Resonance Spectroscopy (^{1}H NMR)

In order to have a better understanding about the composition and structure of the products, another spectroscopic method (^{1}H NMR) was used. NMReady at the frequency of 60MHz with a resolution of FWHM<1.0 Hz (20 ppb) was used for all ^{1}H NMR measurements and data acquisition. Figure 2-3 reports the spectra for LG conversion in the presence of SCW at nine different conditions. Note that the ^{1}H NMR data was not used in the chemometric analysis.

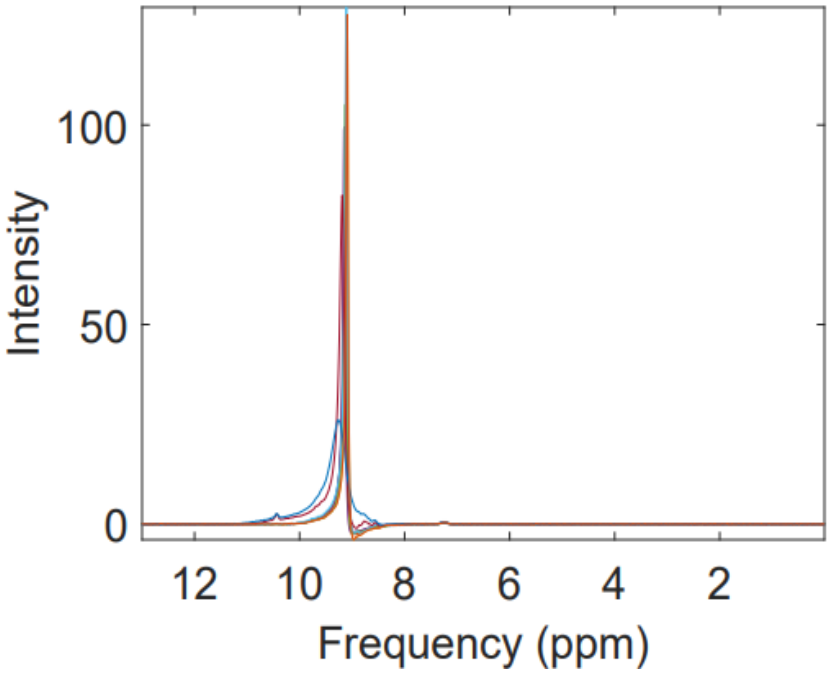


Figure 2-3. ^{1}H NMR results for the products of HTL of cellulose and supercritical water.

2.3.4. Bayesian Hierarchical Clustering (BHC)

In a data set, classification or cluster analysis methods that are mostly based on deterministic[27,28] or model-based techniques[29,30] can be used to create distinct subgroups in a data set. The BHC method, which is based on the Dirichlet process mixture model (DPM), has many advantages over other model-based methods. [31]This technique increases the quality of clustering by using Bayesian model selection rather than an ad hoc distance metric, which is important in dealing with high-dimensional data sets such as the data collected with FTIR in this research. [32,33]

The BHC technique, instead of dealing with point estimates and their variances, focuses on calculating the posterior distributions of the unknown quantities given data, x, which can be interpreted as:

Posterior $\propto$ Prior $\times$ Likelihood

Suppose a data allocation K groups all observations into K clusters, and T is the total size of all individual clusters T_1, T_2, …., T_K. To determine the Multinomial-Dirichlet distribution, the following formula can be considered:

$$f(K) \propto \frac{(K-1)!T1!,\ldots,TK!}{\mathrm{T!(T+K-1)!}} \quad (2\text{-}1)$$

Then, the clustering posterior can be represented as:

$$f(K|x)=\frac{f(x|K)f(K)}{\beta} \quad (2\text{-}2)$$

Based on this, the individual samples can be grouped into K clusters where the marginal density of the data is $f(x|K)$ with a normalization constant larger than zero with no role in clustering (β).[34]

2.3.5. Self-Modeling Multivariate Curve Resolution (SMCR)

Spectroscopic techniques that provide high-dimensional data consisting of one spectral and two spatial dimensions are usually difficult to interpret and analyze. To resolve this issue, chemometric methods can be applied. In these methods, statistical or mathematical techniques have been used to collect the required information regarding the objects of interest in the data.[35] Principal component analysis (PCA) is an example of this class of methods.[36] Another example is SMCR, which was first introduced by Lawton and Sylvester in 1971 [35] and has since been applied successfully in data analysis of spectroscopic and chromatographic data. The idea behind SMCR is to deconvolve unresolved two-way signals into single signals of pseudocomponents by applying suitable mathematical techniques.[30] The advantages are that it requires a small amount of quantitative data to be effective and that it does not require vast prior knowledge of the system.[14] Many SMCR techniques exist;[30,36] however, techniques that deal with two-way resolution are usually divided into two groups. In the first group, the mathematical approaches are used to define the pure variables to ultimately find a unique resolution. Evolving factor analysis (EFA) is an example of a technique used routinely for unique reslolution.[37]

The second type of SMCR technique is more applicable when there is not enough prior knowledge about multi-mixture systems such as in our case. Orthogonal projection analysis (OPA)[38] and alternating least squares (ALS)[39] are two common approaches for these so-called rational resolution techniques; this is the approach adopted in this work. To develop the SMCR-ALS algorithm, the first step (after denoising and background correction) is calculating the error based on the following formula:

$$Error(l) = \sum_{i=1}^{m} \sum_{j=1}^{n} (Data - Datareconstructed\ using\ l\ PC's)^2 \quad (2\text{-}3)$$

Next, the IND function, which is the square root of the reconstructed error variance, is calculated. It is representative of the wavenumbers of pseudocomponents whose composition changes in the reaction.

$$IND(L) = \sqrt{\frac{var(Error(1{:}l))}{\mathrm{n}-1}} \quad (2\text{-}4)$$

The ratio of the second and third derivatives (ROD) is calculated to determine the number of chemical species changing during the reaction, i.e. the chemical rank:

$$\mathrm{ROD} = \frac{IND(L-2)-IND(L-1)}{\mathrm{IND(L-1)-IND(L)}} \quad (2\text{-}5)$$

After finding the chemical rank, EFA is used to determine the initial guess for concentration of the pseudocomponents as the smallest of the forward and backward singular values.[40] Concentration of pseudocomponents is estimated over each process run. The initial concentration estimate is then used in an alternating least squares (ALS) process to decompose the data matrix, D, into its resolved components or chemical rank by applying equation 2-5. These initial concentrations will be also used later to develop BN.

$$D = CS^T + E \quad (2\text{-}6)$$

For a mixture with A components, where D is a matrix of m spectra at n various wavelengths, C is an m × A matrix of pure concentration profiles, S is an n × A matrix of pure spectral profiles, and E is an m × n noise matrix.

After resolving the spectra, D, the following two equations are solved iteratively to convergence. The threshold for the residual is set to 10^{-4}.

$$\min_{S} ||D - CS^T||^2, S \geq 0 \quad (2\text{-}7)$$

$$\min_{C} ||D - CS^T||^2, C \geq 0 \quad (2\text{-}8)$$

2.3.6. The Bayesian Network (BN) Approach

After cluster analysis was performed, Bayesian learning networks were used to develop the most probable reaction network based on the data. The Bayesian network can represent the quantitative strength of the connections between clusters or functional groups found in the previous steps. A Bayesian network, B, is a joint probability distribution including a series of random variables (V). This network is characterized by a pair B = (G, Φ), in which G presents a directed acyclic graph whose nodes X_1, X_2, ….. , X_n, and Φ represent random variables, V, and edges represent the direct dependencies between these variables. The graph, G, encodes independence assumptions by which each variable, X_i, is independent of its non-descendants given its parents in G.[41]

The Bayesian information criterion (BIC), [42] which is commonly used in the score-based approach to develop Bayesian networks, is given by:

$$BIC = \sum_{i=1}^{n} \text{logP xi (Xi } |\Pi \text{ xi)} - \frac{d}{2} \log(n) \qquad (2\text{-}9)$$

Where d is the number of parameters of the global distribution, n is the number of samples, and the last term is introduced to reduce the effect of overfitting.

Developing a Bayesian network includes two steps: (a) learning the structure of the network or graph, and (b) parameter estimation, which determine or estimate the parameters of the global distribution provided in the first step.[41] More than one learning approach is often used to test for optimality. In our work, we consider the solution to be likely to be optimal if the same structure is obtained applying two or more learning approaches.[68] All the algorithms described above have been implemented in MATLAB version 2018b and R version 3.5.1.

2.4. RESULTS AND DISCUSIONS

2.4.1. Cluster Analysis Using BHC

In this work, the basis for clustering is the similarity of the functional groups and their wave numbers. Before cluster analysis, due to the large size of each data set (1764 variables or

wavenumbers), principal component analysis (PCA) was applied for dimensionality reduction. The first two PCs, which explain 95.5% of the variance in the data, were used.

The BHC approach was then used for cluster analysis; the dendrogram is presented in Figure 2-4. The literature indicates that in the pyrolysis of cellulose and related compounds, the main products of this process are mostly acids, aldehydes, ketones, gases, phenols, and alcohols.[43,44] To keep the number of clusters in this range of possible compound classes, the number of clusters to be extracted from the dendrogram were chosen to be 3-6. Rather than use a clustering metric like the silhouette coefficient to identify the optimal number of clusters, we evaluate the most probable number of clusters based on the chemical significance of the BN developed with each of the numbers of clusters studied. Also, handbooks related to spectrometric identification of organic compounds were consulted to identify the most likely compounds or functional groups in each cluster (Table 2-2).[45]

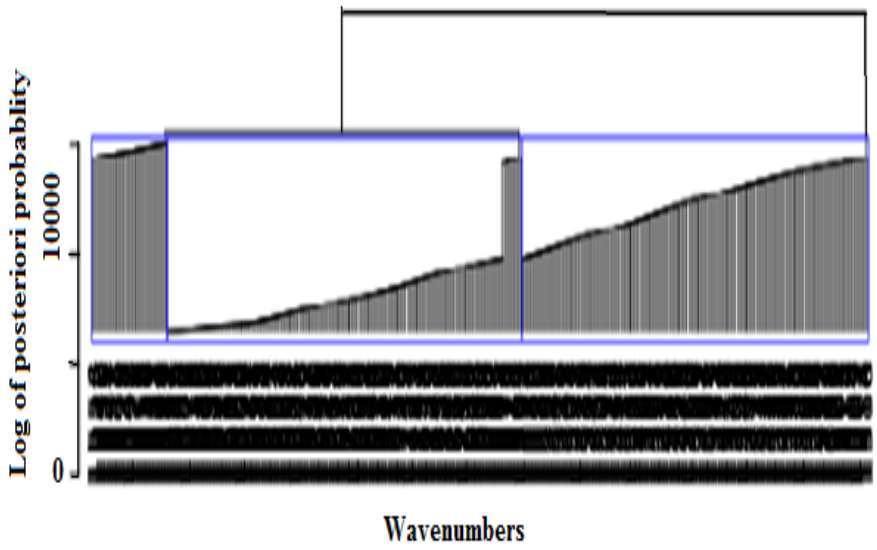


Figure 2-4. Dendrogram obtained by applying BHC cluster analysis result (3-cluster).

To develop the dendrogram, we assumed that the algorithm generated cluster probabilities for each observation rather than allocating an observation to a cluster based on distance metric. Classical clustering methods like k-means, when used to cluster high dimensional data, are sensitive to the presence of outliers.[46] This is overcome by using the BHC which accounts for the outliers while performing the clustering.

Table2- 2. The main functional groups from pyrolysis of cellulose (adapted from45).

Wave number(cm^{-1})	Functional groups	Compounds
3500-3200(s, b)	O-H stretch, H-bonded	Alcohols, Phenols
3300-3250(m)	O-H stretch	Carboxylic acids
3100-3000(s)	C-H stretch	Aromatics
3000-2850 (m)	C-H stretch	Alkanes
1760-1665(s)	C=O stretch	Carbonyls(general)
1760-1690(s)	C=O stretch	Carboxylic acids
1740-1720(s)	C=O stretch	Aldehydes
1715((s)	C=O stretch	Ketones, Saturated aliphatic
1500-1400(m)	C-C stretch(in-ring)	Aromatics
1410-1310	C-O stretch	Phenols
1370-1350(m)	C-H rock	Alkanes
1320-1000(s)	C-O stretch	Alcohols, carboxylic acids, esters, ethers
1000-650(s)	=C-H bend	Alkenes
725-720(m)	C-H rock	Alkanes

s=strong, m=medium, w=weak, n=narrow, b=broad, sh=sharp

2.4.2. Applying the BN Approaches to Develop a Reaction Network

Three different approaches were applied to the task of learning the structure of a reaction network for the hydrous pyrolysis of LG. Two greedy search-and-score algorithms (tabu and hill climbing) and a hybrid structure called the max-min hill climbing method (MMHC) were applied. While the tabu search starts with a feasible initial solution and picks the next best option that can increase the score function, HC starts with an initial guess for a solution and iteratively makes local changes to it until two conditions are met: the solution is found, or the heuristic gets stuck in a local maximum.[47] Unlike the first two methods, the last approach is a hybrid structure learning algorithm, which is a combination of constraint and score-based algorithms and includes two steps called restriction and maximization.[48]

Figures 2-(5& 7-9) depict the proposed reaction networks for hydrous pyrolysis of cellulose with 3-6 clusters using R programing version 3.5.1. The reaction network with three clusters was the only one where all three approaches reported the same network.

2.4.2.1. The BN for Hydrous Pyrolysis of Cellulose (Three Clusters)

The interesting point after application of the BN is that all three methods provided the same reaction network when three clusters were chosen, thus providing confidence in the validity of

the reaction network. Figure 2-5 shows that in the investigated reaction conditions, the final product can be obtained from two pathways: direct conversion of raw material or through an intermediate product. Appendices 2($A_{1\text{-}1}$, $A_{1\text{-}2}$, and $A_{1\text{-}3}$) display the wavenumbers grouped in each cluster. Based on this data and identification of the relevant functional groups,[49,26] we concluded that the first and second clusters consist mostly of alcohols and ethers. The third cluster contains carbonyl groups (aldehydes, ketones, and acids), normal alkanes, alkenes, alkynes, alcohols, and aromatics. Therefore, we can argue that by moving from cluster 1 to 3, the number of compounds with carbonyl groups and aliphatic groups increases. This result has a good agreement with the literature.[50,51] The implication is that aldehydes/acids can be produced directly from cellulose or through an intermediate product, most likely glucose. After we performed ^{1}H NMR spectroscopy, the result confirmed that under the investigated conditions, there is a singlet peak around 9.60 ppm confirming that the product can be an aldehyde with no neighboring carbon. This description has a good agreement with formaldehyde (Figure 2-3).

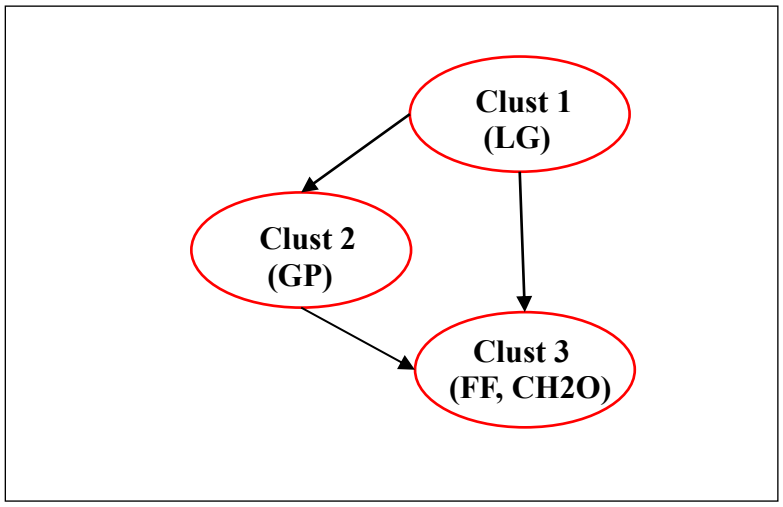


Figure 2-5. Reaction network derived by the tabu, HC, and MMHC Bayesian network.

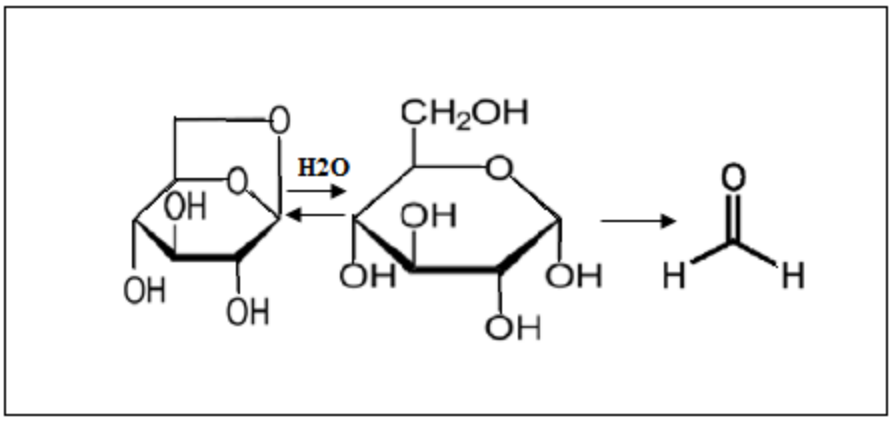


Figure 2-6. The proposed chemical pathways for the decomposition of LG.

Table 2-3 lists the strength of each arc in the tabu, HC, and MMHC search-based BN. These values present the dependency and probable pathways between the clusters in a network (the strongest dependency belongs to the highest negative number). Based on this, the highest belief is for the arc from cluster 1 to cluster 2 (94.36).

Table 2-3. Arc strength calculated by tabu, HC, and MMHC search methods.

From	To	Arc strength
Cluster 1	Cluster 2	-94.36
Cluster 1	Cluster 3	-29.62
Cluster 2	Cluster 3	-20.14

Figure 2-5 shows that all three methods identify the same proposed reaction network, and the arc strengths shown in Table 2-3 are the same for all three methods.

After developing the BN (DAG Structure) which is responsible for describing the directed causal map between clusters, there is a need for parameter learning to quantify the uncertainty about the model. This step calculates the quantitative parameters of the model (Θ) in a way that the conditional probability of each child (in a single path or arc, a child is a descendant of its parent in the sequence of the ordered nodes) depends only on its parents (Markovian property).[17] To do so, a scalar value representing the intensity of the cluster has to be estimated by using the conditional probability distribution of that group; we choose the RMS (root mean square) value of the absorbance intensities.

To define Θ, we can consider X_i the intensity value of the i_{th} variable (i= 1,2,3) and μ_i the mean value of X_i. Considering these conditions, the conditional probability distribution of each group has been shown in the pseudokinetic equations 2(10-12), and the model describing mean intensity can be found in the pseudokinetic equations 2(13-15).

$$P(X_1) \sim N(\mu_1, 23.7^2) \quad (2\text{-}10)$$
$$P(X_2 | X_1) \sim N(\mu_2, 0.39^2) \quad (2\text{-}11)$$
$$P(X_3 | X_1, X_2) \sim N(\mu_3, 0.19^2) \quad (2\text{-}12)$$
$$\mu_1 = 13.82 \quad (2\text{-}13)$$
$$\mu_2 = 21.32 + 0.58\mu_1 \quad (2\text{-}14)$$
$$\mu_3 = 7.81 - 0.53\mu_2 + 1.38\mu_1 \quad (2\text{-}15)$$

Equations 2(13-15) demonstrate to what extent the mean value of the probability distribution of each cluster is related to others in the BN. This pseudokinetic information can be useful to monitor an online process in real time analysis, since it provides (real-time) estimates of (pseudo) reaction rates.

2.4.2.2. The BN for Hydrous Pyrolysis of Cellulose (Four Clusters)

For the case of four clusters, tabu, HC, and MMHC search algorithms resulted in different networks. However, we selected the network provided by tabu search since this technique has the advantage of guiding the local search algorithms in search of optimal/near optimal solutions by developing or updating a tabu list of forbidden moves of neighboring solutions. In other words, tabu search, in opposite with other methods, uses memory to avoid looping or forbidding already visited solutions by imposing some tabu to the search.[52]

Figure 2-7 provides the Bayesian reaction network using tabu search for the investigated reaction with four clusters. According to this reaction network, the first cluster is the raw material, or levoglucosan. This compound can be converted to final products (cluster 4) such as acids (3300-2500 cm^{-1}) or aldehydes (1740-1720 cm^{-1}) in two pathways. First, a dehydration reaction can produce furfural (cluster 3, supported by ^1H NMR data) and then go through bond breaking to produce the final products. Piskorz et al. also confirmed that two important $C_{5\text{-}6}$ ring compounds from the cellulose pyrolysis are furfural and 5-hydroxymethyl-furfural (furan derivatives).[53] Secondly, bond breaking produces glucose (cluster 2) as an intermediate product, which can take two pathways: dehydration, resulting in furfural, or bond breaking, producing the final products.

This reaction network has a good agreement with the ones proposed by Lee. et al.[54] and Shen et al.[51]

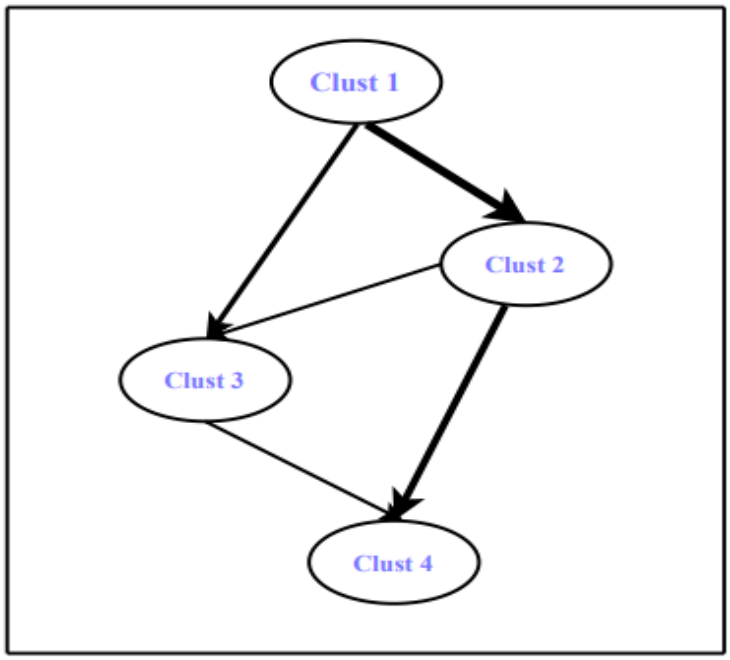


Figure 2-7. Bayesian network using tabu search (4 clusters).

In pyrolysis of LG, three mechanisms are reported in the literature for the decomposition processes. Bond breaking (C-C and C-O) and LG dehydration or desorption of water molecule are endothermic reactions. Pouwels et al. showed through gas chromatography-mass spectroscopy (GC-MS) analysis that acetaldehyde (AA), 1-pentene3,4-dione (1-P-3,4-D), 2,3-dihydroxypropanal (2,3-DP), and propanedialdehyde (PD), CO, and H_2O are the main products of C-O bond breaking.[55] Hydroxyacetic acid vinyl ester (HAVE) and 1,2-dihydroxyethene (1,2-DE) are the main products of C-C bond breaking.[7]

Table 2-4. The strength values of each arc in the tabu search-based BN.

From	To	Arc strength
Cluster 1	Cluster 2	-91.53
Cluster 1	Cluster 3	-20.88
Cluster 2	Cluster 3	-31.30
Cluster 2	Cluster 4	-49.57
Cluster 3	Cluster 4	-19.98

From Table 2-4 (the arc strength for this reaction network), it can be concluded that between three pathways from cluster one to the final products (cluster four), the one directed from LG to glucose (cluster two, arc strength -91.53) and then headed to cluster four (arc strength -49.57) has the highest dependency. This is in agreement with the BN with three clusters. Equations 2(16-19) describe the conditional probability of each group, and equations 2(20-23) are the models for the mean intensity of each group for the pseudokinetics associated with this BN structure.

$P(X_1) \sim N(\mu_1, 25.76^2)$ (2-16)
$P(X_2 | X_1) \sim N(\mu_2, 0.44^2)$ (2-17)
$P(X_3 | X_1, X_2) \sim N(\mu_3, 0.18^2)$ (2-18)
$P(X_4 | X_2, X_3) \sim N(\mu_4, 0.30^2)$ (2-19)
$\mu_1 = 9.85$ (2-20)
$\mu_2 = 24.07 + 0.53\mu_1$ (2-21)
$\mu_3 = 7.36 - 0.44\,\mu_1 + 1.30\,\mu_2$ (2-22)
$\mu_4 = 10.74 + 2.24\,\mu_2 - 1.45\,\mu_3$ (2-23)

2.4.2.3. The BN for Hydrous Pyrolysis of Cellulose (Five Clusters)

The tabu and hill climbing methods provide the same reaction networks for this reaction considering 5 clusters (Figure 2-8).

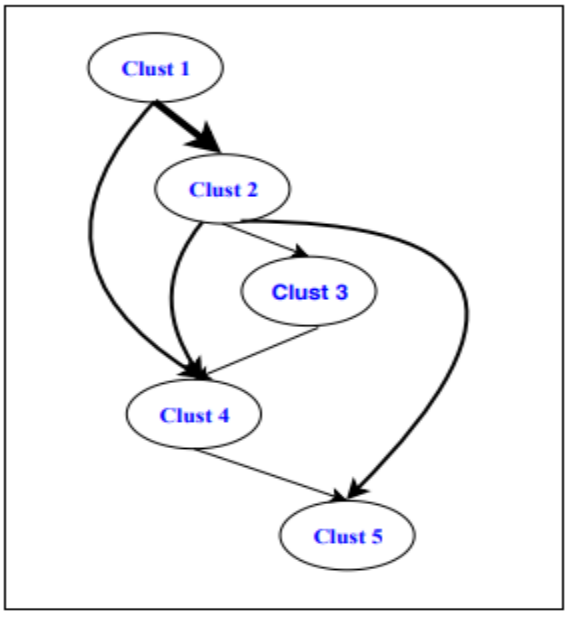


Figure 2-8. Bayesian network using tabu and hill climbing search (5-cluster).

Based on the proposed BN, multiple possible chemical pathways exist for primary decomposition of levoglucosan (cluster 1) to the final products (gases): (1) LG can undergo bond breaking to produce carboxyl groups such as acids/aldehydes (cluster 4), and these compounds can be broken to produce gases; (2) cluster 2, which contains glucose, can be broken into acids/aldehydes and then converted to the final products; (3) glucose also can be dehydrated to produce cluster 3 (furfural), which can be broken to acids/aldehydes; (4) glucose can be considered an intermediate compound to be converted to cluster 5. The BN with five clusters also finds support in the literature.[7] Shafizadeh also proposed such a scheme that through cellulose pyrolysis, low molecular weight compounds such as furfural, aldehydes, acetic acid, formic acid, and light species can be formed.[56]

Table 2-5. The strength values of each arc in the tabu search-based BN.

From	To	Arc strength
Cluster 2	Cluster 5	-51.63
Cluster 1	Cluster 4	-53.68
Cluster 2	Cluster 3	-32.43
Cluster 3	Cluster 4	-9.25
Cluster 4	Cluster 5	-0.36
Cluster 2	Cluster 4	-94.69
Cluster 1	Cluster 2	-98.92

Noticeable high strength value for arcs (Table 2-5) are the ones from group 1 (LG) to group 2 (glucose) and group 2 to group 4 (acids/aldehydes). This finding provides another confirmation that it is most likely to produce glucose from LG and acids/aldehydes from glucose.

The conditional probability distribution of each group is described in equations 2(24-28), and the model for the groups' mean intensities is presented in equations 2(29-33).

$$P(X_1) \sim N(\mu_1, 13.71^{\,2}) \quad (2\text{-}24)$$
$$P(X_2 | X_1) \sim N(\mu_2, 0.47^{\,2}) \quad (2\text{-}25)$$
$$P(X_3 | X_2) \sim N(\mu_3, 0.65^{\,2}) \quad (2\text{-}26)$$
$$P(X_4 | X_1, X_2, X_3) \sim N(\mu_4, 0.074^{\,2}) \quad (2\text{-}27)$$
$$P(X_5 | X_2, X_4) \sim N(\mu_5, 0.34^{\,2}) \quad (2\text{-}28)$$
$$\mu_1 = 29.31 \quad (2\text{-}29)$$
$$\mu_2 = -19.78 + 1.38\,\mu_1 \quad (2\text{-}30)$$
$$\mu_3 = -9.58 + 1.38\,\mu_2 \quad (2\text{-}31)$$
$$\mu_4 = 9.56 + 1.63\,\mu_1 - 069\,\mu_2 - 0.12\,\mu_3 \quad (2\text{-}32)$$
$$\mu_5 = -10.06 + 1.4\,\mu_2 - 0.24\,\mu_4 \quad (2\text{-}33)$$

2.4.2.4. The BN for Hydrous Pyrolysis of Cellulose (Six Clusters)

Tabu search provides the following BN for this reaction considering six clusters (Figure 2-9): Although LG can be fragmented and produce gases directly, it can also produce furfural as an intermediate compound (cluster 2). Many researchers confirmed that about 90% of total gas products by weight from cellulose products are CO and CO_2.[57,58] Another pathway for converting LG to gases exists through producing glucose (cluster 3) and then converting it to aldehydes/acids (cluster 5). Glucose and furfural can also produce phenols (cluster 4) through bond breaking and dehydration.[43]

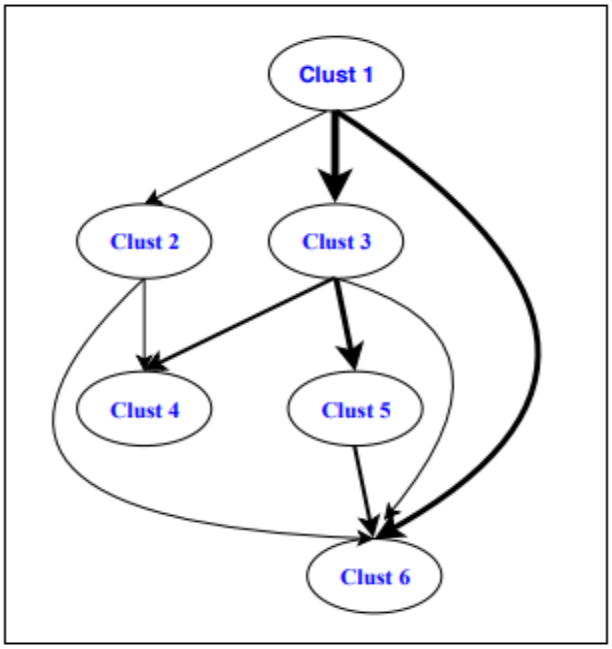


Figure 2-9. Bayesian network using tabu search (6-cluster).

Table 2-6. The strength values of each arc in the tabu search-based BN.

From	To	Arc strength
Cluster 3	Cluster 4	-87.41
Cluster 1	Cluster 6	-58.73
Cluster 1	Cluster 2	-39.85
Cluster 3	Cluster 6	-39.33
Cluster 2	Cluster 4	-12.13
Cluster 2	Cluster 6	-21.01
Cluster 5	Cluster 6	-11.46
Cluster 3	Cluster 5	-94.69
Cluster 1	Cluster 3	-98.92

It can be concluded (from Table 2-6) that there is no dependency from group 4 to group 6 meaning it is not likely to convert the phenolic groups directly to gases. This conclusion has a good agreement with the literature.[50]

The conditional probability distribution of each group is described in equations 2(34-39), and the model for the groups' mean intensities is presented in equations 2(40-45).

$$P(X_1) \sim N(\mu_1, 13.71^2) \quad (2\text{-}34)$$
$$P(X_2 | X_1) \sim N(\mu_2, 0.50^2) \quad (2\text{-}35)$$
$$P(X_3 | X_1) \sim N(\mu_3, 0.47^2) \quad (2\text{-}36)$$
$$P(X_4 | X_2, X_3) \sim N(\mu_4, 0.23^2) \quad (2\text{-}37)$$
$$P(X_5 | X_3) \sim N(\mu_5, 0.65^2) \quad (2\text{-}38)$$
$$P(X_6 | X_1, X_2, X_3, X_5) \sim N(\mu_5, 0.083^2) \quad (2\text{-}39)$$
$$\mu_1 = 29.31 \quad (2\text{-}40)$$
$$\mu_2 = 42.59 + 0.16\,\mu_1 \quad (2\text{-}41)$$
$$\mu_3 = -19.78 + 1.38\,\mu_1 \quad (2\text{-}42)$$
$$\mu_4 = 6.21 - 0.55\,\mu_2 + 1.4\,\mu_3 \quad (2\text{-}43)$$
$$\mu_5 = -9.58 + 1.18\,\mu_3 \quad (2\text{-}44)$$
$$\mu_6 = 10.14 + 2.3\,\mu_1 - 0.34\,\mu_2 - 1.0\,\mu_3 - 0.16\,\mu_5 \quad (2\text{-}45)$$

Next, we consider self-modeling multivariate curve resolution (SMCR) as an alternative to clustering for creating pseudocomponents for developing the reaction network. [59]

2.4.3. Self-Modeling Multivariate Curve Resolution (SMCR-ALS) & Bayesian Network

Figure 2-10 illustrates the denoised and background-corrected signals for the FTIR spectra shown in Figure (2-2) by using the Savitsky Golay filtering algorithm, along with the residual noise removed by the procedure.

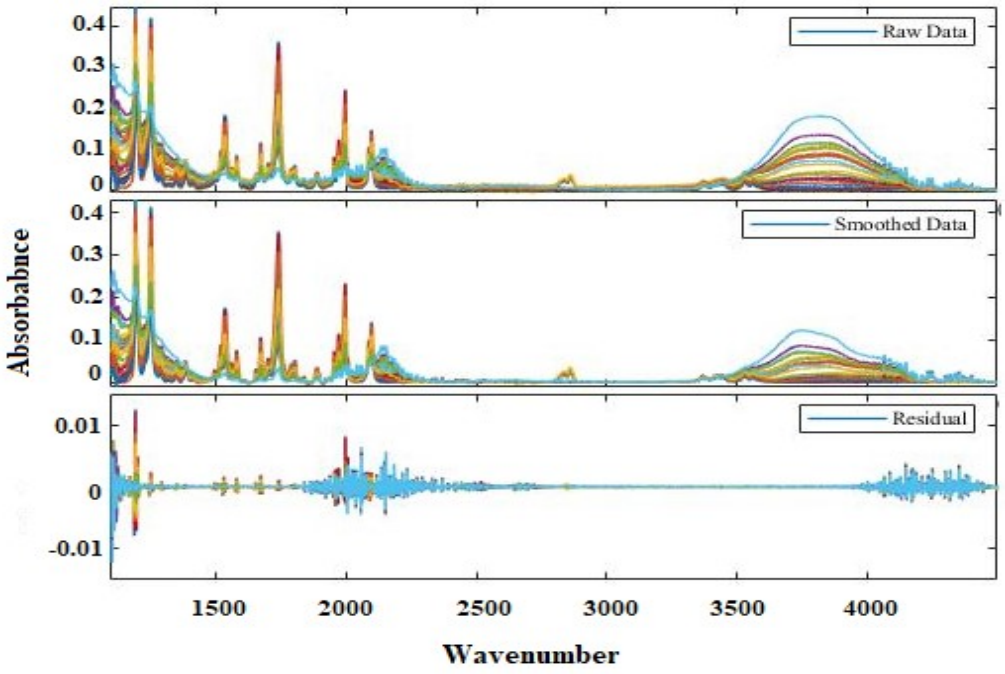


Figure 2-10. Raw and smoothed FTIR spectra and removed residual data.

To develop the SMCR-ALS algorithm, after calculating the error (equation 2-3), the sum of square of errors over the number of principal components (PCs) was estimated and plotted (Fig. 2- 11). The result confirms that reconstructing the data from an increasing number of PCs decreases appreciably from PC_1 to PC_3, indicating that most of the variance in the data is captured in the first three PCs.

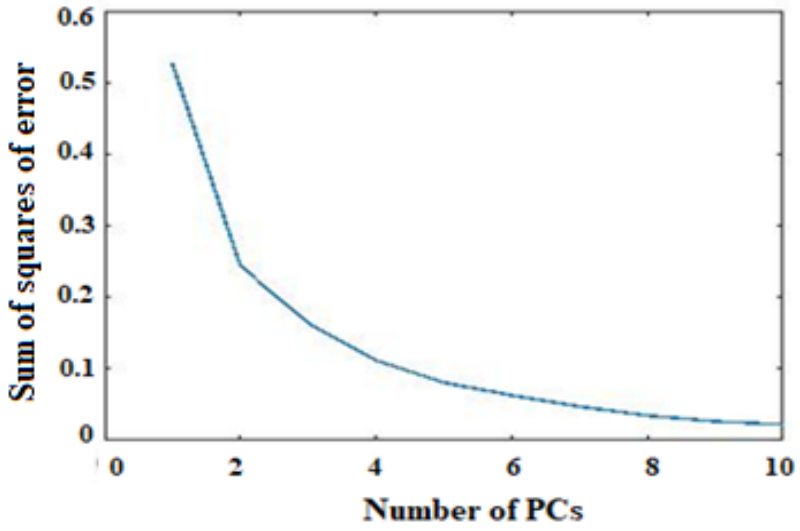


Figure 2-11. Sum of squares of error from the reconstruction of data using principal components.

The next important step was to determine the chemical rank. To do this, the ROD was calculated (equations 2-4 and 2-5) and is plotted as shown in Figure 2-12. The ROD reaches a maximum value of 3, which is the value of the chemical rank.

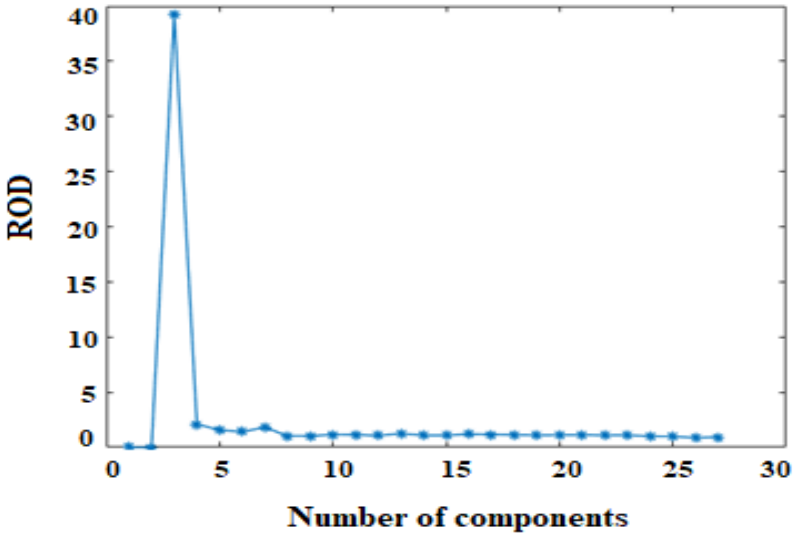


Figure 2-12. Calculating chemical rank using ROD.

After finding the chemical rank, in order to develop the BN, the initial concentration was calculated by applying EFA (equations 2(6-8)), after which ALS was used to identify pseudocomponents (equations 2(7-8)). Figure 2-13 shows the convergence of the ALS algorithm; less than 10 iterations lead to acceptable convergence.

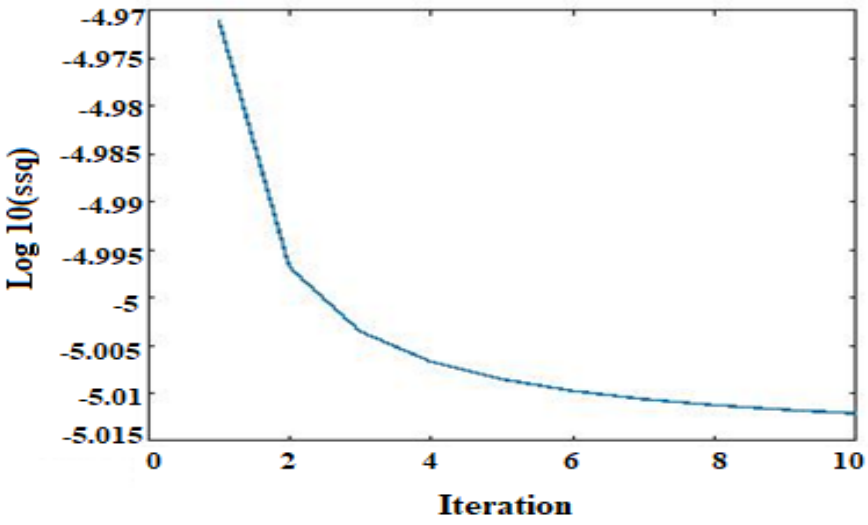


Figure 2-13. The rate of convergence of ALS.

When we want to have an appropriate interpretation regarding the spectra collected from the above algorithm, we should keep in mind that A_1, A_2, and A_3 (representing the first, second, and third pseducomponents) contain spectral structures which can represent the actual molecules. While Figure 2-14(a) demonstrates the overall resolved spectra for the pseudo components (A_1, A_2, and A_3) for the whole region (400-4000 cm^{-1}), Figures 2-14(b-d) illustrate the regions containing the major peaks of these pseudocomponents. For the frequency range shown in Figure 2-14(b), the significant peaks belong to A_1 and A_2. A_2 mostly represents primary and secondary alcohols because its major peaks are mainly located in 1075-1010 cm^{-1}, related to primary alcohols, and 1120-1100 cm^{-1}, representing secondary alcohols. The spectrum for A_1 illustrates the presence of monosubstituted alkene ($RCH=CH_2$) at 995-985 cm^{-1}, Alkene (C=C=C) at 1060 cm^{-1}, and aliphatic ethers (R-O-R) at 1150-1070 cm^{-1}. [26,60]

On the other hand, Figures 2-14 (c-d) show more peaks related to the pseudocomponent A_3. The paired absorptions at 1640 cm^{-1} and the broad absorption covering the range 3600-3000 cm^{-1} are typical of (C=O)-CH=(C-OH) (i.e., a ketone and alcohol one carbon apart) where the second pair carbons is olefinic (sp^2 hybridized). The carbonyl group can be a ketone or an aldehyde because the absorption at 1640 cm^{-1} is due to the conjugation and intramolecular hydrogen bonding.[44] In other words, the result can be obtained that pseudo components A_1 and A_2 are willing to convert to pseudo component A_3 in the regions with higher wave numbers.

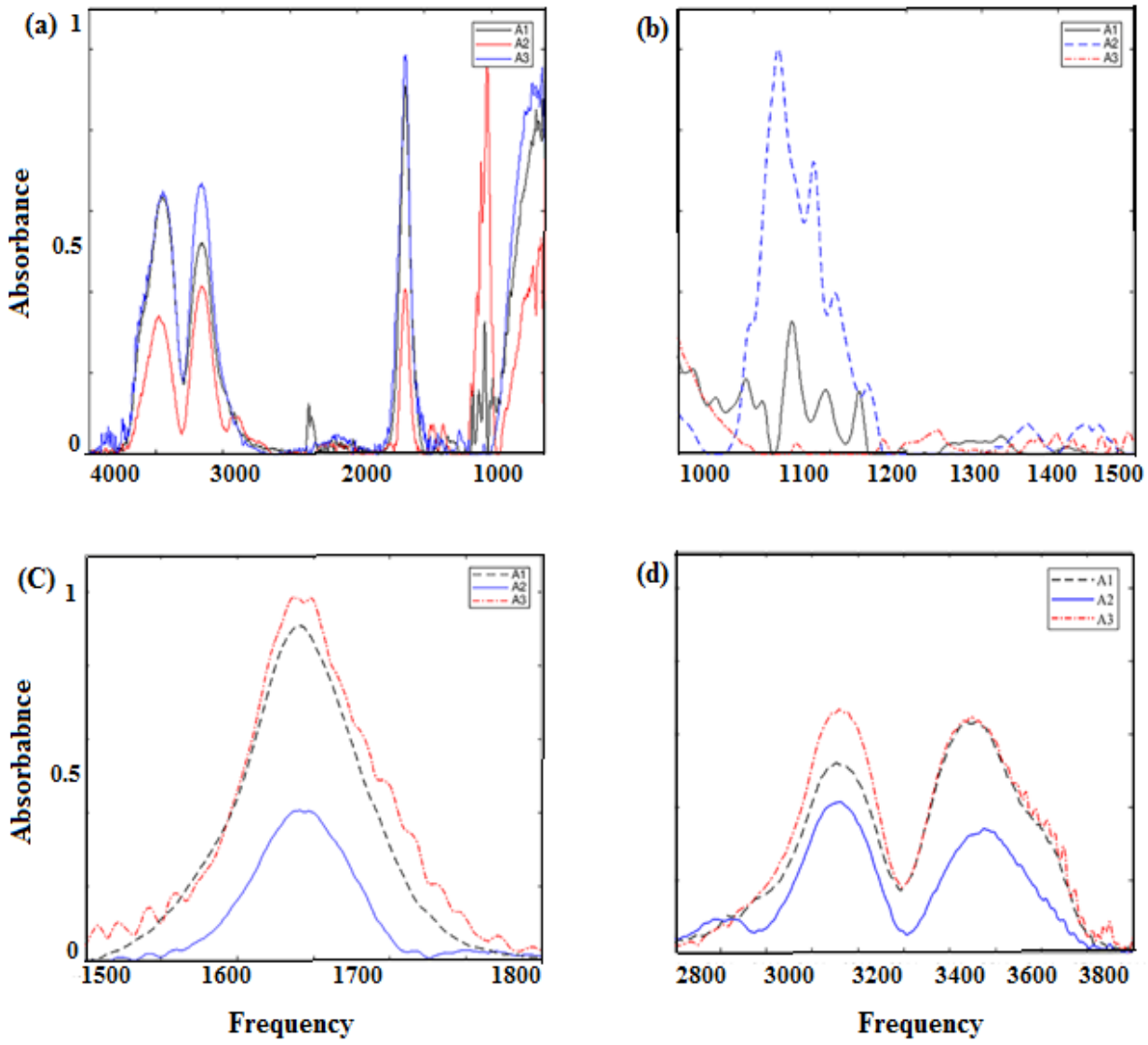


Figure 2-14. (a) Resolved spectra for the pseudo components after applying SMCR-ALS; and (b, c, and d) resolved spectra for A_1, A_2, and A_3 by focusing on the major peaks.

To develop the reaction network, both greedy research methods (tabu and hill climbing) were used, and both methods provided the same results (Figure 2-15). The one explanation could be that pseudo compounds having (-OH), (-C=C-), or (R-O-R') groups can be converted to ortho-hydroxy aryl ketones (C=O)-CH=(C-OH) by alkyl and hydrogen transfer and oxidation reactions.[26]

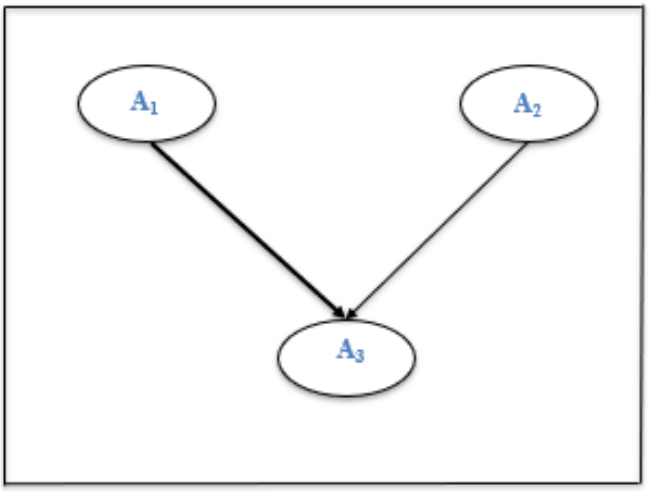


Figure 2-15. tabu and HC Bayesian network structure for the pseudo components

Table 2-7. The strength values of each arc in the tabu and HC search-based BN.

From	To	Arc strength
Cluster 1	Cluster 3	-78.31
Cluster 2	Cluster 3	-46.29

To better understand the second part of BN, which is parameter learning, equations 2(46-48) present each group's conditional probability, and the models for these groups' mean intensities are shown in equations 2(49-51). This represents the pseudokinetics associated with this BN.

$P(X_1) \sim N(\mu_1, 0.37^2)$ (2-46)
$P(X_2) \sim N(\mu_2, 0.21\ s^2)$ (2-47)
$P(X_3 | X_1, X_2) \sim N(\mu_3, 0.019^2)$ (2-48)
$\mu_1 = 0.45$ (2-49)
$\mu_2 = 0.26$ (2-50)
$\mu_3 = 1.002 - 0.92\mu_1 - 0.51\mu_2$ (2-51)

2.5. PROPOSED REACTION MECHANISM FOR LG DECOMPOSITION

Developing a reaction network provides the ability to predict and monitor the mechanism and pathways of the process. The goal of this research was to determine the most probable reaction network for hydrous pyrolysis of levoglucosan at relatively low temperature (< 300 °C). To do that, three different Bayesian network approaches and parameter learning were used considering three to six clusters due to the fact that the major products which can be established in good

yields (under the investigated conditions) are: levoglucosan, furfural, glucose, acids, aldehydes, and gases. [9] Based on the results provided earlier, the most probable pyrolytic pathway for this reaction is a causal map with three clusters which can be explained as converting LG to formaldehyde directly or by producing an indeterminate compound such as glucose. In other words, by applying BHC method and grouping wavenumbers, a BN with three clusters allow us to link LG, glucose, and formaldehydes, to clusters 1, 2, and 3, respectively. The findings are in good agreement with previously published data. According to Wang et al, the presence of LG can be validated by tracing the typical infrared carbohydrate peaks around 3356, 3273, 2906, 1400, 1131, and 1042 cm^{-1}. [61] In another study done by Nybacka, the fundamental frequencies of glucose are: 3350, 2920, 2850, 1450, and 1035 cm^{-1}. [62] Finally, the existence of formaldehyde (a major component of the third group) can be confirmed by the existence of peaks around 2785, 1750, 1485, and 1250 cm^{-1}.[63] All these wavenumbers can be found in clusters 1, 2, and 3 (Appendices 2 (A_1- A_3)), respectively. As a result, this study proposes the following reaction map for LG decomposition.

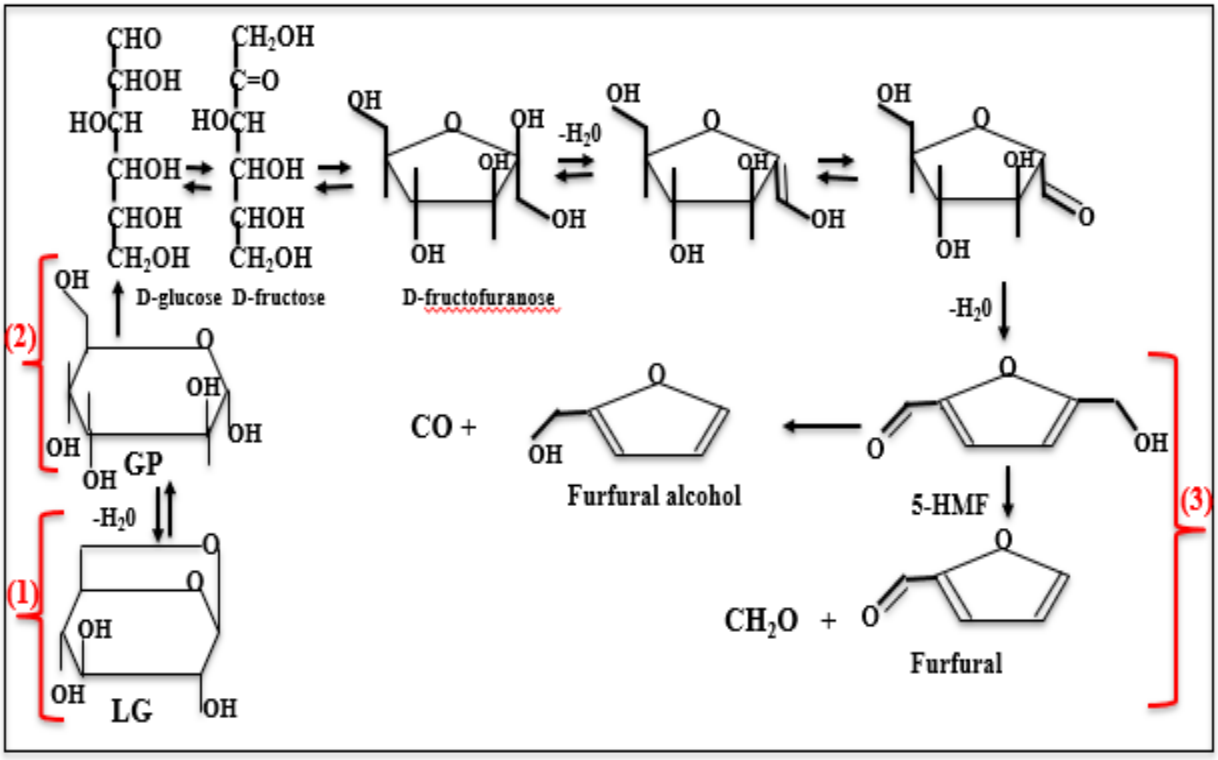


Figure 2-16. Proposed pathway for hydrous pyrolysis of LG.[64]

Based on this pathway, the presence of the hexose chain structure is the result of the cleavage of pyran-ring. This initial step will cause the production of 5-hydroxymethylfurfural (5-HMF) through the dehydration of the hydroxyl groups(C-2). Finally, the formation of formaldehyde (CH_2O) and furfural (FF) is the result of the dehydroxymethylation reaction of the side chain of the furan-ring.

Both the graphical representation (Figure 2-5) of the causal relationship between the variables (which in this case are wavenumbers), and the quantitative strength of the connections between these variables (Table 2-3) provided the same results, which are good agreement with the literature.[43] The arc from group 1 to group 2 has the highest probability (-94.36) representing the strongest dependency while the lowest dependency belongs to an arc pointing out from group 2 to group 3 (-20.14). It should be noted that the associated pseudokinetics developed can be valuable for online monitoring, but they assume Gaussian probability distributions and cannot represent nonlinearity effectively.

A remarkable result from all developed BNs is that (regardless of number of clusters) the highest strength value for arcs belongs to the one that produce glucose from LG and acids/aldehydes from glucose. This outcome is another confirmation that a BN with three clusters is the most probable pyrolytic pathway for the investigated reactions.

Another interesting finding comes from the six-cluster BN (Figure2- 9). This reaction network shows no dependency between clusters four and six, confirming no direct conversion of phenols to gases between 100°C and 250°C. This temperature range is consistent with other studies done by Xu et al.[65] and DiLeo et al. [66] which report similar findings at temperatures below 600° C.

Since high-dimensional data usually provides data with overlapping issues, in the second part of this research, a chemometric method such as SMCR was used to predict the number of pseudo compounds presenting in these reactions. The result confirmed the presence of three groups of changing compounds in these reactions. Later, BN was applied to develop a map or causality between these pseudo components. It can be concluded that in hydrous pyrolysis of cellulose under the investigated conditions, pseudo components A_1 and A_2, which can represent molecules such as alcohols, akenes and ethers, have a tendency to produce compounds with longer chains such as ortho-hydroxy aryl ketones, represented by A_3. Developing SMCR algorithms can be

applied as a useful technique for monitoring online and controlling of multi-mixture reactions such as thermochemical conversion of biomass.

Figure 2-17 illustrates the concentrations related to A_1, A_2 and A_3 over the number of samples. These concentration profiles obtained from SMCR corroborate with the inference made from the BN where A_1 and A_2 are seen to react to give A_3, whose concentration is seen to increase at the expense of the other pseudocomponents as the processing time is increased. The application of this plot makes it useful in the online monitoring of species conversion by integrating it with a suitable control strategy that adjusts process conditions to maximize the yield of the desired product.

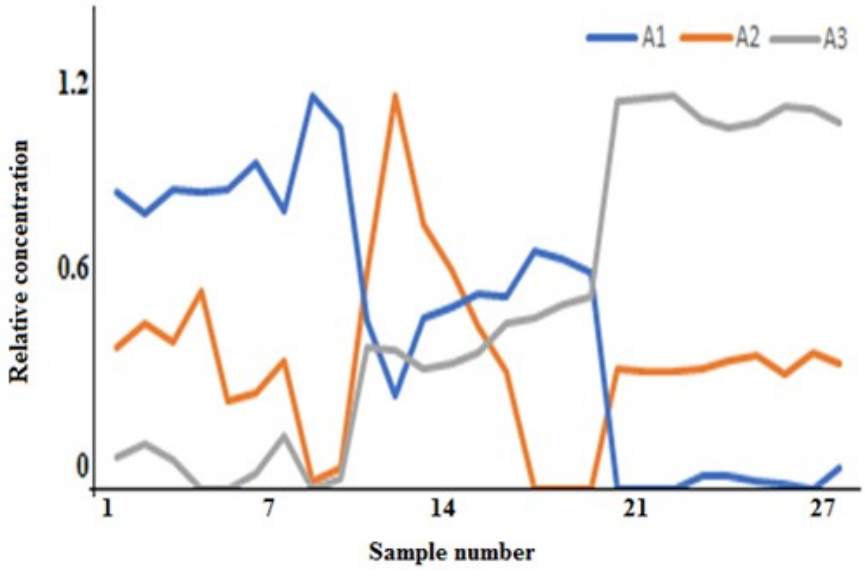


Figure 2-17. Corresponding concentration for pseudo components A_1, A_2, and A_3.

When it comes to comparing the BHC and SMCR to generate nodes among which the Bayesian networks are developed; the latter method provides better results. Applying the BHC technique groups wavenumbers in a cluster without any background constraints on how chemistry changes over different process conditions. On the other hand, SMCR algorithm determines the concentration changes of pseudocomponents across different process conditions along with a full spectrum for each. This makes it more appropriate for the on-line monitoring of change in species during a process. However, in terms of mapping to real chemistry, the results from SMCR are more difficult to interpret.

2.6. CONCLUSIONS

Since biomass can be considered a potential substitute for fossil fuels, there is a need for better understanding of pyrolysis of cellulose as a major component of biomass. In this research, levoglucosan was reacted in the presence of hot water and various catalysts (H_2SO_4 and NaOH) at twenty-seven temperatures and reaction times. In all cases, products were analyzed and identified by FTIR and ^{1}H NMR spectroscopic methods. LG initially underwent depolymerization and dehydration of the larger molecules to produce condensable vapor (liquid) such as formaldehyde or volatiles.

In the first part of this research, BHC was applied for data clustering, and then three different BN approaches (tabu, HC, and MMHC) were used to develop a reaction network. All methods provided the same result, considering three clusters that can be interpreted as either converting levoglucason directly to formaldehyde or producing an intermediate compound such as glucose. The result has a high degree of agreement with the literature. In this paper, the other reaction networks considering four, five, and six clusters were proposed and interpreted based on expert knowledge and literature.

The high-dimensional data set provided by FTIR spectroscopy method with a relatively small number of samples was used to achieve the second goal of this research which was developing a SMCR-ALS algorithm for automatic prediction of pseudo components and then resolving the concentration and spectra provided from FTIR into its chemical ranks. Demanding a little prior knowledge about the system along with requiring a very short time for computational process make this algorithm a useful tool for real-time analysis and quantitative tracking of changes in the process.

2.7. REFERENCES

1. Crocker M. Thermochemical Conversion of Biomass to Liquid Fuels and Chemicals. *Cambridge R Soc Chem*. 2010.
2. IEA. *CO2 Emissions from Fuel Combustion 2016*. IEA; 2016. doi:10.1787/co2_fuel-2016-en
3. AntalJr. MJ. Biomass Pyrolysis: A Review of the Literature Part 1—Carbohydrate Pyrolysis. *Adv Sol Energy*. 1983:61-111.
4. Pan Y, Kong SC. Predicting effects of operating conditions on biomass fast pyrolysis using particle-level simulation. *Energy and Fuels*. 2017;31(1):635-646. doi:10.1021/acs.energyfuels.6b02445
5. Manya JJ, Velo E, Puigjaner L. Kinetics of biomass pyrolysis: a reformulated three-parallel-reactions model. *Ind Eng Chem Res*. 2003;42(3):434-441.
6. Bradbury AGW, Sakai Y, Shafizadeh F. A kinetic model for pyrolysis of cellulose. *J Appl Polym Sci*. 1979;23(11):3271-3280.
7. Minowa T, Zhen F, Ogi T. Cellulose decomposition in hot-compressed water with alkali or nickel catalyst. *J Supercrit Fluids*. 1998;13:253-259.
8. Dote YS, Sawayama SY, Inoue S, Minowa T. Recovery of liquid fuel from hydrocarbon-rich microalgae by thermochemical liquefaction. *Sci dircet*. 1994;73(12):1855-1857.
9. Paine III JBP, Pithawalla YB, Naworal JD. Carbohydrate pyrolysis mechanisms from isotopic labeling Part 2 . The pyrolysis of D -glucose : General disconnective analysis and the formation of C 1 and C 2 carbonyl compounds by electrocyclic fragmentation mechanisms. *J Anal Appl Pyrol*. 2008;82(1):10-41. doi:10.1016/j.jaap.2008.01.002
10. Nimlos MR, Evans RJ. Levoglucosan Pyrolysis. *Fuel Chem Div Prepr*. 2002;47(1):393.
11. Hosoya T, Kawamoto H, Saka S. Different pyrolytic pathways of levoglucosan in vapor- and liquid / solid-phases. *J Anal Appl Pyrolysis*. 2008;83:64-70. doi:10.1016/j.jaap.2008.06.008
12. Rass E-J, McKay B RG. Understanding catalytic biomass conversion through data mining. *Top Catal*. 2010;53:1202-1208.
13. Beyramysoltan S, Abdollahi H, Rajkó R. Newer developments on self-modeling curve resolution implementing equality and unimodality constraints. *Anal Chim Acta*. 2014;827:1-14. doi:10.1016/j.aca.2014.03.019
14. Tefera DT, Agrawal A, Yañez Jaramillo LM, De Klerk A, Prasad V. Self-Modeling Multivariate Curve Resolution Model for Online Monitoring of Bitumen Conversion Using Infrared Spectroscopy. *Ind Eng Chem Res*. 2017;56(38):10756-10769. doi:10.1021/acs.iecr.7b01849
15. Shen, T. T.; Zou, X. B.; Shi, J. Y.; Li, Z. H.; Huang, X. W.; Xu, Y. W.; Chen W. Determination Geographical Origin and Flavonoids Content of Goji Berry Using Near-Infrared Spectroscopy and Chemometrics. *Food Anal Methods*. 2016;9(1):68–79.

16. Faltin F, Kenett R. Bayesian Networks. *Encycl Stat Qual Reliab*. 2007;1(1):4. doi:10.1002/wics.48

17. Lee SM, Abbott PA. Bayesian networks for knowledge discovery in large datasets: Basics for nurse researchers. *J Biomed Inform*. 2003;36(4-5):389-399. doi:10.1016/j.jbi.2003.09.022

18. Sasaki M, Fang Z, Fukushima Y, Adschiri T, Arai K. Dissolution and hydrolysis of cellulose in subcritical and supercritical water. *Ind Eng Chem Res*. 2000;39(8):2883-2890.

19. Demirbas A. Pyrolysis of municipal plastic wastes for recovery of gasoline-range hydrocarbons. *J Anal Appl Pyrolysis*. 2004;72(1):97-102. doi:10.1016/j.jaap.2004.03.001

20. Wu W, Kawamoto K, Kuramochi H. Hydrogen-rich synthesis gas production from waste wood via gasification and reforming technology for fuel cell application. *J Mater Cycles Waste Manag*. 2006;8(1):70-77. doi:10.1007/s10163-005-0138-1

21. Karimi K, Kheradmandinia S, Taherzadeh MJ. Conversion of rice straw to sugars by dilute-acid hydrolysis. *Biomass and Bioenergy*. 2006;30(3):247-253.

22. Wu H, Fu Q, Giles R, Bartle J. Production of mallee biomass in Western Australia: Energy balance analysis. *Energy and Fuels*. 2008;22(1):190-198. doi:10.1021/ef7002969

23. Jin H, Zhao X, Wu Z, Cao C, Guo L. Supercritical water synthesis of nano-particle catalyst on TiO_2 and its application in supercritical water gasification of biomass. *J Exp Nanosci*. 2017;12(1):72-82. doi:10.1080/17458080.2016.1262066

24. Feng W, van der Kooi HJ, de Swaan Arons J. Phase equilibria for biomass conversion processes in subcritical and supercritical water. *Chem Eng J*. 2004;98(1):105-113.

25. Kabir G, Hameed BH. Recent progress on catalytic pyrolysis of lignocellulosic biomass to high- grade bio-oil and bio-chemicals. *Renew Sustain Energy Rev*. 2017;70(January):945-967. doi:10.1016/j.rser.2016.12.001

26. Mistry B. *A Handbook of Spectroscopic Data Chemistry (UV, IR, PMR, 13CNMR and Mass Spectroscopy)*. Oxford Book Company; 2009. doi:10.1063/1.2719251

27. Everitt, Brian, Landau S, Leese M SD. *Cluster Analysis*. Wiley Series in Probablity and Statistics; 2001.

28. Sharma S. *Applied Multivariate Techniques*. John Wiley & Sons, Inc.; 1996.

29. McLachlan, G. J and Peel DA. *Finite Mixture Models*. John Wiley & Sons; 2000.

30. Fraley, C. and Raftery A. Model-Based Clustering, Disccriminant Analysis, and Density Estimation. *Am Statitical Assoc*. 2002;97(458):611-631.

31. Cooke E, Savage R, Kirk P, Darkins R WD. Bayesian hierarchical clustering for microar-ray time series data with replicates and outlier measurements. *BMC Bioinformatics*. 2011;12(399).

32. Sirinukunwattana K, Savage RS, Bari MF, Snead DRJ, Rajpoot NM. Bayesian hierarchical clustering for studying cancer gene dxpression Data with unknown statistics.

PLOS One. 2013;8(10). doi:10.1371/journal.pone.0075748

33. Tadesse MG, Sha N, Vannucci M. Bayesian variable selection in clustering high-dimensional data. *J Am Stat Assoc*. 2005;100(470):602-617. doi:10.1198/016214504000001565

34. Tefera DT, Yañez Jaramillo LM, Ranjan R, Li C, De Klerk A, Prasad V. A Bayesian Learning Approach to Modeling Pseudoreaction Networks for Complex Reacting Systems: Application to the Mild Visbreaking of Bitumen. *Ind Eng Chem Res*. 2017;56(8):1961-1970. doi:10.1021/acs.iecr.6b04437

35. Blanco M, Castillo M, Peinado A, Beneyto R. Application of multivariate curve resolution to chemical process control of an esterification reaction monitored by near-infrared spectroscopy. *Appl Spectrosc*. 2006;60(6):641-647. doi:10.1366/000370206777670710

36. Sanchez, F.C.; van den Bogaert, B.; Rutan, S.C.; Massart DL. Orthogonal projection approach applied to peak purity assessment. *Chem*. 68(1):139-171.

37. Keller, H.R., Massart, D.L., and De Beer JO. Window evoloving factor analysis for assesment of peak homogeneity in liquid chromatography. *Chem*. 1993;65(4):471-474.

38. Jiang J, Ozaki Y. Self -modeling curve resolution (SMCR): principals, techniques, and applications T. *Ind Eng Chem Res*. 2017;4928(October). doi:10.1081/ASR-120014359

39. Awa K, Okumura T, Shinzawa H, Otsuka M, Ozaki Y. Self-modeling curve resolution (SMCR) analysis of near-infrared (NIR) imaging data of pharmaceutical tablets. *Anal Chim Acta*. 2008;619(1):81-86. doi:10.1016/j.aca.2008.02.033

40. Sequaris J-ML, Koglin E. Evolving Factor Analysis for the Resolution of Overlapping Chromatographic Peaks. *Anal Chem*. 1987;59(6):527-530. doi:Doi 10.1021/Ac00130a035

41. Rissanen J. *Information and Complexity in Statistical Models*. Springer Publishing Company; 2007.

42. Koller D FN. Probabilistic graphical models: principles and techniques. *MIT Press Cambridge*. 2009.

43. Yang H, Yan R, Chen H, Lee DH, Zheng C. Characteristics of hemicellulose, cellulose and lignin pyrolysis. *Fuel*. 2007;86(12-13):1781-1788. doi:10.1016/j.fuel.2006.12.013

44. Demirbaş A. Mechanisms of liquefaction and pyrolysis reactions of biomass. *Energy Convers Manag*. 2000;41(6):633-646. doi:10.1016/S0196-8904(99)00130-2

45. Li G, Torraca G, Jing W, Wen Z qing. Applications of FTIR in identification of foreign materials for biopharmaceutical clinical manufacturing. *Vib Spectrosc*. 2009;50(1):152-159. doi:10.1016/j.vibspec.2008.10.016

46. Fränti VHCKK. Improving K-Means by Outlier Removal. In: Scandinavian Conference on Image Analysis; 2005:978-987.

47. Ronaldus R. Bayesian belief networks: from construction to inference. 1995.

48. Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network

structure learning algorithm. *Springer Sci*. 2006:31-78. doi:10.1007/s10994-006-6889-7

49. Wu S, Shen D, Hu J, Xiao R, Zhang H. Pyrolysis TG-FTIR and Py-GC – MS analysis of a model compound of cellulose – glyceraldehyde. *J Anal Appl Pyrolysis*. 2013;101:79-85. doi:10.1016/j.jaap.2013.02.009

50. Shen D, Xiao R, Gu S, Zhang H. The Overview of Thermal Decomposition of Cellulose in Lignocellulosic Biomass. *INTECH*. 2013. doi: 10.5772/51883.

51. Shen D, Xiao R, Luo K. The pyrolytic behavior of cellulose in lignocellulosic biomass : a review. *RSC Adv*. 2011:1641-1660. doi:10.1039/c1ra00534k

52. Glover, F., Laguna M. *Tabu Search.* Kluwer Academic Publishers; 1998.

53. Piskorz J, Radlein D SD. On the mechanism of the rapid pyrolysis of cellulose. *Anal Appl Pyrolysis*. 1986;9:121-137.

54. Lee H, Dharma D, Hsie B, Tang G, Chakrabarti D, De Klerk A, Prasad V. Biomass conversion through hydrous pyrolysis. In: *Innovation, Industry and Internationalization*. London,ON; 2011:348-349.

55. Pouwels, A.D, Eijkel , G.B, and Boon JJ. Pyrolysis. *J Anal Appl*. 1989;14(4):237–280.

56. Shafizadeh F. Introduction to pyrolysis of biomass. *J Anal Appl Pyrolysis*. 1982;3(4):283-305.

57. Shen DK GS. The mechanism for thermal decompostion of cellulose and its main products. *Bioresour Technol*. 2009;100:6496-6504.

58. Yu Y, Bartle J, Mendham D, Wu H. Site Variation in Life Cycle Energy and Carbon Footprints of Mallee Biomass Production in Western Australia. *Energy & Fuels*. 2015;29(6):3748-3752. doi:10.1021/acs.energyfuels.5b00618

59. Heller K a., Ghahramani Z. Bayesian hierarchical clustering. *Proc 22nd Int Conf Mach Learn*. 2005:297-304. doi:10.1145/1102351.1102389

60. Robert M. Silverstein, Francis X. Webster DJK. Spectrometric identification of organic compounds. *J Mol Struct*. 2005:512. doi:10.1016/0022-2860(76)87024-X

61. Wang J, Wei Q, Zheng J, Zhu M. Effect of pyrolysis conditions on levoglucosan yield from cotton straw and optimization of levoglucosan extraction from bio-oil. *J Anal Appl Pyrolysis*. 2016;122:294-303. doi:10.1016/j.jaap.2016.09.013

62. Nybacka L. FTIR spectroscopy of glucose FTIR spectroscopy of glucose. 2016.

63. Zhang X, Yang W, Blasiak W. Thermal decomposition mechanism of levoglucosan during cellulose pyrolysis. *J Anal Appl Pyrolysis*. 2012;96:110-119. doi:10.1016/j.jaap.2012.03.012

64. Cao F, Schwartz TJ, McClelland DJ, Krishna SH, Dumesic JA, Huber GW. Dehydration of cellulose to levoglucosenone using polar aprotic solvents. *Energy Environ Sci*. 2015;8(6):1808-1815. doi:10.1039/c5ee00353a

65. Xu X, Matsumura Y, Stenberg J, Antal MJ. Carbon-Catalyzed Gasification of Organic

Feedstocks in Supercritical Water †. *Ind Eng Chem Res*. 1996;35(8):2522-2530. doi:10.1021/ie950672b

66. DiLeo GJ, Neff ME, Savage PE. Gasification of guaiacol and phenol in supercritical water. *Energy and Fuels*. 2007;21(4):2340-2345. doi:10.1021/ef070056f

3. APPLICATION OF DATA COMBINATION AND DATA MINING TECHNIQUES TO INVESTIGATE THE CHEMISTRY OF CELLULOSE AND LIGNIN DERIVATES IN HYDROUS PYROLYSIS

A version of this chapter will be submitted as; Fereshteh Sattari, Arno de Klerk and Vinay Prasad, “Application of Data Combination and Data Mining Techniques to Investigate the Chemistry of Cellulose and Lignin Derivatives in Hydrous Pyrolysis".

3.1. BACKGROUND

Increasing energy demand as well as concerns regarding climate change related to consuming fossil fuels and the depletion of non-renewable resources have increasingly shifted attention towards liquid biofuels. Currently, biofuel is not as cost-competitive as compared to crude oil refineries. Therefore, one necessary aspect of oil production could be focusing on effective cellulose and lignin utilization. [1]

Cellulose and lignin, which are considered the two most important parts of biomass, have many applications in industrial processes. While cellulose, in principle, is a major ingredient for papers, foods, chemicals, and medicines, [2] lignin is a necessity for the production of many chemicals such as benzene, xylene, and toluene, due to its aromatic structure.[3] Unlike the homogeneity structure of cellulose, which is made of a polysaccharide of glucose, [4] the amorphous structure of lignin mostly consists of phenylpropane monomers with methoxy and hydroxy groups attached to the rings.[5]

Over the past few decades, bio-oil production has been the center of attention for many researchers. Malauan focused on the production of different chemicals by pyrolyzing waste biomass through non-catalyzed reactions in the presence of near-critical or supercritical water.[6] This process, which was later called hydrous pyrolysis, is a possible thermochemical conversion technique for converting the biomass into different products – mostly liquids and gases. This process happens at temperatures below 400°C in the presence of water and the absence of oxygen. Therefore, designing and optimizing biomass pyrolysis processes requires a clear understanding of how cellulose and lignin thermally decompose. In order to investigate the behavior of the compositions of these complex processes and monitor the chemical pathways, developing the reaction network as an abstraction representing the chemical reaction can be a useful technique.[7] Concentrations and/or spectral information can provide the data sets required to develop a reaction network that can ultimately help demonstrate the elementary reactions and their rate coefficients.[8] In other words, reaction networks permit the analysis of reaction dynamics in complex chemistry without using complicated calculation techniques.[9]

There are many different approaches to develop a reaction network, but when it comes to a smaller number of samples, a graphical structure method, called the Bayesian networks (BNs)

approach, representing information or knowledge about an uncertain domain is one of the proper techniques.[10] This method belongs to the probabilistic model family and uses nodes and arcs or edges to represent a random variable and the probabilistic dependencies between these variables. To estimate these conditional dependencies, the BNs use the new available evidence to compute the posterior.[11]

Data required for developing the BNs could be provided from different sources. For the past few decades, usage of FTIR spectroscopy has been dramatically increased in the areas of oil and fuel refining, online and real-time analysis, and modelling because it can provide fast and non-destructive vibrational data.[12] Since crude oil and its products are rich in carbon and hydrogen, another useful data analysis technique is ^{1}H NMR spectroscopy.[13]

Applying different analytical technology processes or multiple spectroscopic methods can increase the information and variation found in the data set by bringing forth more reliable information. At the same time, it comes with a big challenge related to how to best and effectively collect, capture, mine, and extract the available information within these large amounts of data sets.

To this end, algorithms and mathematical procedures such as data fusion or data combination, with different levels of complexity can offer a method of combining multiple sources of data to have the optimal chance of mapping and extracting the information of a data space. In the 1960s, data fusion was introduced into the mathematical field for the first time as a way for data manipulation, and later, in the 1970s, it was executed in the defense and robotics fields.[14] Currently, data fusion has many applications in different areas, such as mine detection, maintenance engineering, and weather prediction.[15]

Since there is no one-size-fits-all model for all subjects required for the applications of data fusion, the first step before starting a data fusion task is establishing a strategy in a robotic and organized way, which can help obtain a solution to the problem. Researchers have introduced different algorithms for data fusion. In 1988, Luo and Kay[16] developed multi-sensor integration as a generic structure for data fusion. Based on their proposal, a fusion center in a hierarchical manner is required to combine the data provided from different sources, and there is now a distinction between data integration and data fusion. Based on their definition, while the former

is applied to completing a particular task by using multiple sources of data, the latter can happen when there is actual fusion of the data, and it could also be applied during any stage of the integration process. In this study, data integration method has been applied to develop the final data set for causality detection in a complex reacting system, such as those in petrochemical products or the thermochemical conversion of biomass.

3.2. MATERIALS

In this study, levoglucosan (LG) and 2-Phenoxyethyl benzene (PEB) were used as model compounds to represent cellulose and lignin. While LG is not a model compound for cellulose, it is known that when cellulose undergoes thermal decomposition or pyrolysis, the most abundant product in the primary reaction is LG. [17,18] Although LG is insoluble in non-hydroxylic solvents, it is soluble in water; and its molecular structure includes a pyranose ring connected to three hydroxyl and two ether groups. PEB can be considered a model compound for the dominant ether linkage in lignin, [19] with a molecular formula of $C_{14}H_{14}O$ and an average mass of 198.26 Da. In this study, LG and PEB underwent the hydrous pyrolysis reactions in non-catalyzed conditions as well as in the presence of sulfuric acid (0.05 M) and sodium hydroxide (1 M), separately.

3.3. METHODS

3.3.1. Hydrous Pyrolysis

Among different biomass thermochemical conversion techniques, this research targets hydrous pyrolysis in regard to its advantages over other methods. One benefit could be the kind of products (mainly pyrolysis oil containing various chemicals) that do not require high temperatures to be formed but are easy to store and transport. These high-quality products are capable of many added-value applications.[20] Hydrous pyrolysis is a thermal decomposition in the absence of an oxidizing agent, such as air or oxygen, but in the presence of water. In these kinds of reactions, many factors such as time, temperature, and the heating rate have strong effects on the proportion of the products. Pyrolysis happens at lower temperatures (less than 400°C) compared to other reactions such as gasification; low temperature and long residence times promote char formation while high temperature and less residence time are mostly required for

gas production. However, the optimal conditions for oil production which is the target of this study are medium temperature and short residence time.[21]

Many studies have targeted the pyrolysis of biomass pyrolysis in the presence of supercritical water (SCW). While Boblete and Adchiri focused on the hydrothermal degradation of cellulose with hot water [22, 23] and Mok and his colleagues used an acid catalyst for the hydrolysis reaction of cellulose in hot compressed water; [24] we manipulated the temperature and residence time to obtain the maximum liquid with the desired characteristics (Table 3-1).[25] In all conditions, the initial pressure was considered constant (0.1 MPa), and the volume ratio of individual LG and PEB to the medium was 1:10.

Table 3.1. Variation in ranges of temperature and residence time for LG and PEB in a stainless-steel micro batch-reactor.

Medium	T (°C)	t1(min)	t2(min)	t3(min)
Subcritical water	150	15	35	55
Subcritical water	200	15	35	55
Subcritical water	250	15	35	55
Sulfuric acid	150	10	20	30
Sulfuric acid	180	10	20	30
Sulfuric acid	240	10	20	30
Sodium Hydroxide	100	10	60	110
Sodium Hydroxide	150	10	60	110
Sodium Hydroxide	200	10	60	110

T = temperature, t = time, min = minutes

The experimental setup for these reactions has been presented in our previous work.[26] The stainless-steel batch microreactors (24 cm long and 2.1 cm in diameter) were installed with a thermocouple, heated to the temperature and time planned and left to cool down to 100°C before quenching to room temperature with water. In all reactions, the initial pressure was kept constant at 0.1 MPa by closing the pressure relieve valve.

3.3.2. Spectroscopy Methods

After hydrous pyrolysis, the samples were collected for characterization. FTIR and ^{1}H NMR spectroscopy techniques were used for characterization. In this study, the absorbance signal was acquired in the 4000–600 cm^{-1} range, with a 2 cm^{-1} acquisition step, using an ABB MB 3000

FTIR spectrometer. In total, fifty-four processed samples were considered for both LG and PEB (Figure 3-1) and a handbook of spectroscopic data was used to identify the functional groups. [27]

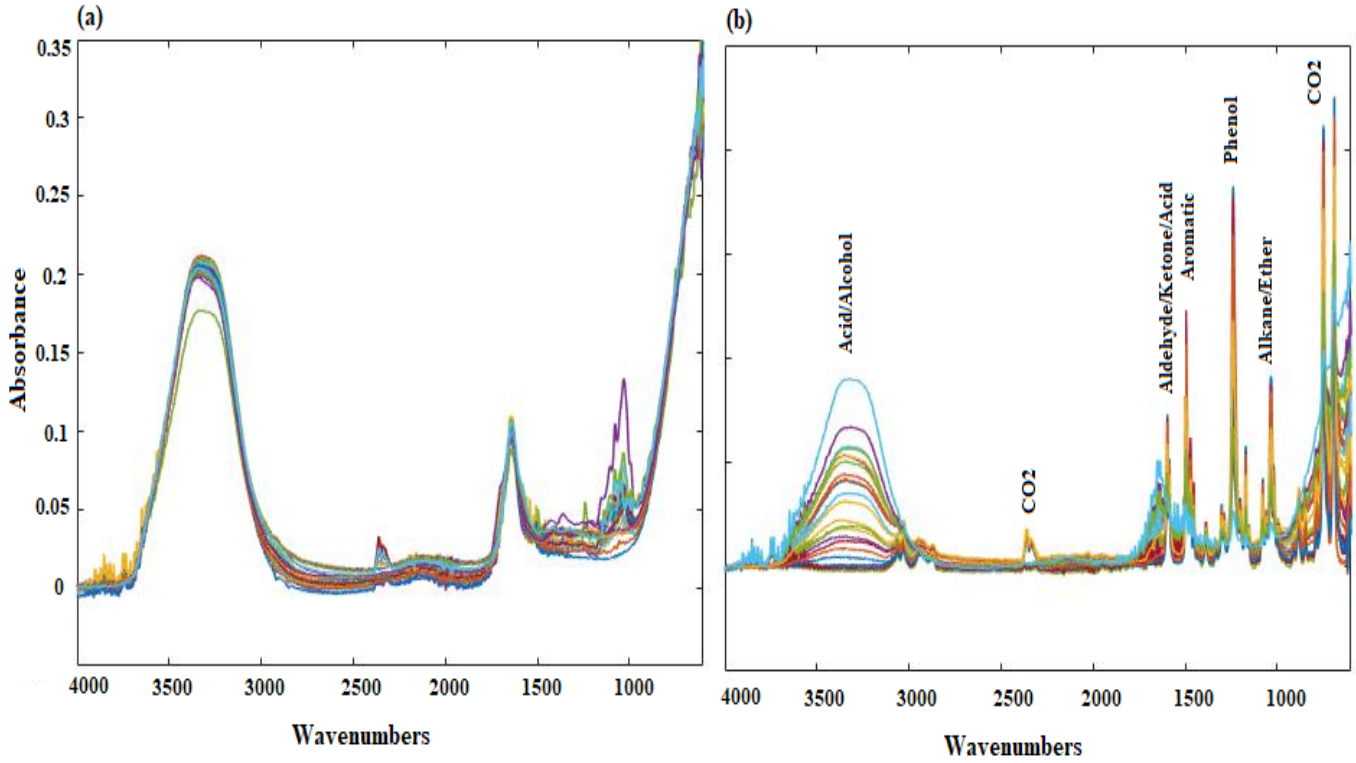


Figure 3-1. Fifty-four FTIR spectra of HTL of (a) LG, and (b) PEB in various conditions.

To have a better and more complete understanding about the data acquired from FTIR, another spectroscopy technique (^{1}H NMR) was selected and used. For all ^{1}H NMR measurements and data acquisition, spectra were acquired using an offline NMReady at the frequency of 60 MHz with a resolution of FWHM < 1.0 Hz (20 ppb). The following figure illustrates the spectrum for the LG and PEB conversion in the presence of SCW and NaOH across eighteen different conditions.

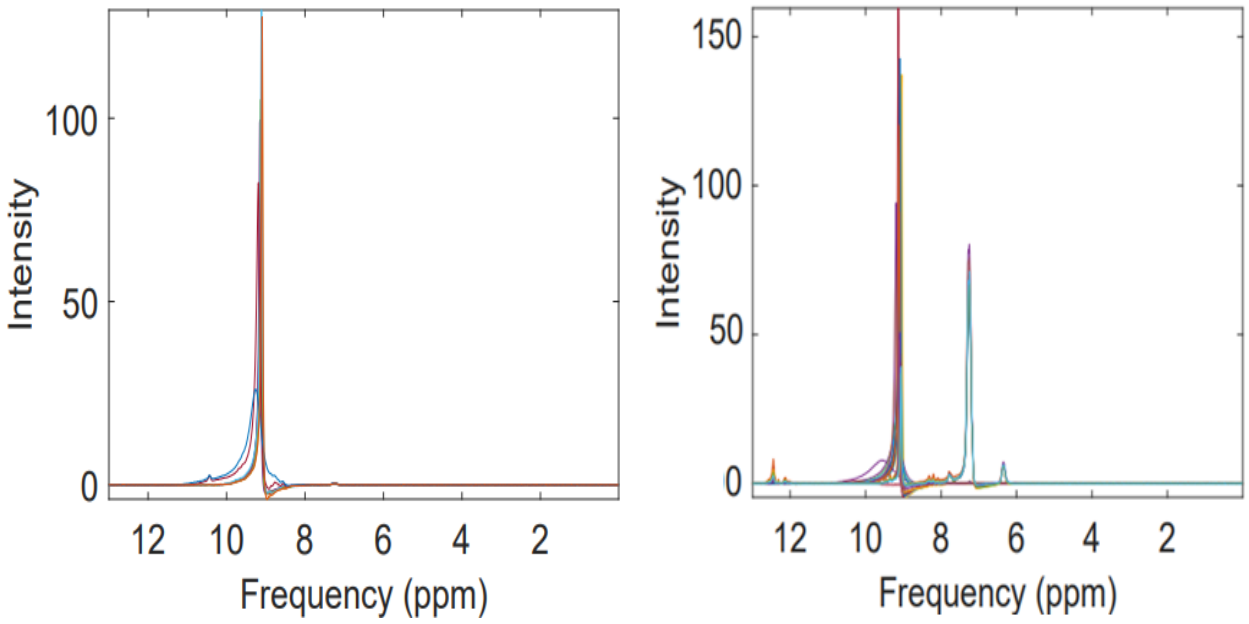


Figure 3-2. Eighteen ^{1}H NMR spectra for (left) LG and SCW, and (right) PEB and NaOH.

3.3.3. Data Combination Technique

The large amounts of data points obtained from FTIR and ^{1}H NMR carry some useful information regarding different features of samples. These data points can provide more accurate and reliable information if they are synergistic combined and analyzed. One goal of this research was to find a final model presenting how best fuse or combine these data pints to have the advantages of increasing the accuracy while reducing the overall uncertainty compared to models produced using the individual spectroscopic entities. By using the final model, a user can extract all useful and meaningful information from a large number of different data sets. Many research studies have confirmed that the potential benefits of applying data fusion or data combination could be complementary, redundancy, and cost of information.[28, 29]

In this study, data combination method was used, since there is no logical outcome corresponding with the input data points. Another aspect considered for developing the final algorithm was incorporating data-mining techniques to facilitate this process. In a systematic manner, data mining can be applied to search for general relationships within data points contained in large amounts of raw data. In other words, data-mining techniques, such as

clustering, can help in the positional integration and object-identity processes. Figure 3-3 illustrates the steps required for this method.

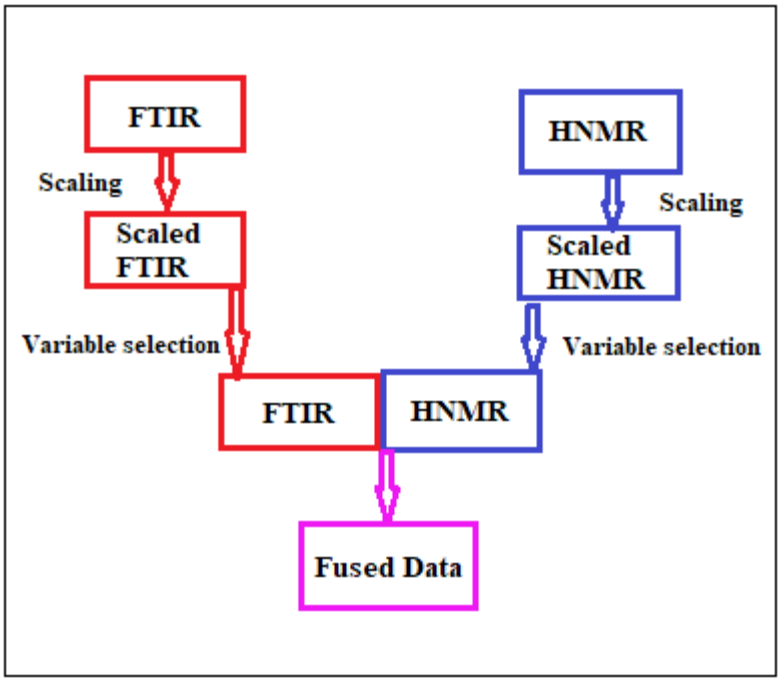


Figure 3-3. Proposed steps for data combination.

3.3.4. Cluster Analysis by Applying BHC SMCR-ALS Algorithms

Cluster analysis is a necessary and useful step to organize the spectroscopic data. In this area, researchers have proposed many different clustering methods based on their definitions of what a cluster actually is. For example, some density methods, such as the density-based algorithm for discovering clusters in large spatial databases with application of noise (DBSCAN) [30] and ordering points to identify the clustering structure (OPTICS), [31] propose to form a single cluster by merging the densest area in the space. In contrast to these methods, there is another method called hierarchical clustering. The goal of this method is to determine the various clusters within a data set by arranging the data points in a hierarchical manner.[32]

Since this research needs to deal with a high dimensional data set provided by FTIR and ^{1}H NMR, another cluster analysis method based on the Dirichlet process mixture model (DPM) seems to be a useful technique. This method, which is called the Bayesian hierarchical clustering

algorithm (BHC), uses a Bayesian algorithm to improve the quality of the resulting clusters and the structure.[33] In this framework, clusters merge based on the probability of their similarity.[34] For a detailed description of the method, we refer readers to our previous work. [35]

Since FTIR and ^{1}H NMR provide high dimensional data sets that are difficult to analyze and interpret,[37] this research also applies a chemometric method called SMCR, which is based on a matrix of mixture spectra. the algorithm uses a minimal constraint of non-negativity for absorbances and concentrations to bilinearly decompose spectroscopic data into concentration profiles and absorbance spectra. The goal of this algorithm was to find the number of species involved in the process and to be sensitive to the detector. To do this, the experimental data matrices D were decomposed into two new matrices according to Eq. (3-1), and their spectral or concentration profiles were estimated: [36]

$$D = CS^T + E \qquad (3\text{-}1)$$

Where C contains information about the concentration and S^T represents the absorbance spectral profiles of the pseudocomponents involved in the process. E is the noise, which can be described as the matrix of the residuals that contains the variance unexplained by CS^T and can be minimized by solving Eq. (3-1) iteratively. Figure 3-4 presents the structure and SMCR -ALS treatment for the individual data matrices (D_{IR} is obtained from FTIR spectra recorded from the samples).

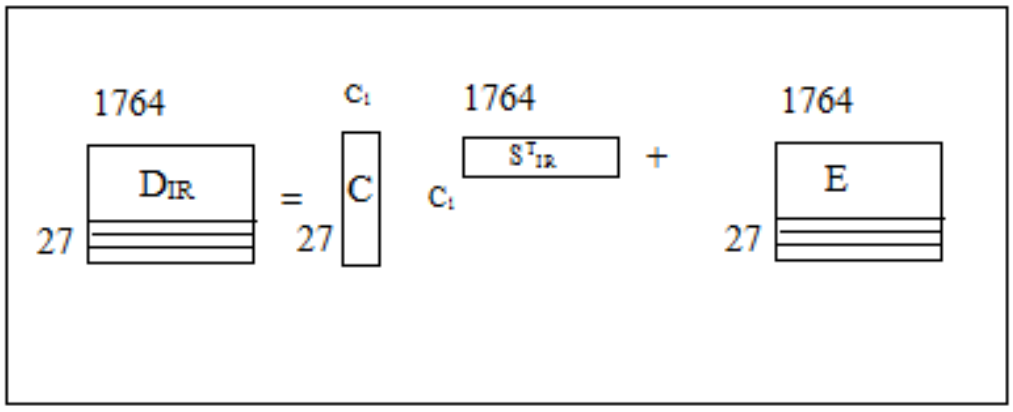


Figure 3-4. The structure and SMCR-ALS treatment for the experimental data matrices.

To compute the number of species in the process, we used an empirical measure of ROD in terms of the errors of reconstruction of the data using the k orthogonal vectors obtained from the singular value decomposition of the data matrix: [37]

$$D = USV^T \quad (3\text{-}2)$$

Where U and V^T represent the left and right singular vectors, respectively, and the diagonal values of S are the singular values of D. In this study, singular value decomposition projected k sample rows in the orthogonal uncorrelated space of m vectors, creating a 1764 x 27 matrix and a 27 x 27 matrix of S (1764 refers to the number of wavenumbers resulted from FTIR, and 27 is the total number of samples).

After determining the chemical rank, to compute the initial estimation of the spectral profile, the interactive self-modelling mixture of the individual data matrices D_{IR} was used, and then it was optimized by using the constrained alternating least squares regression procedure.[26] Non-negativity was applied for the constraints of both the concentration and spectral profile, and the lack of fit (LOF) and the percentage of explained variance (R^2) were applied for the evaluation of the quality of the SMCR-ALS:

$$\text{LOF}\ (\%) = \ 100 * \sqrt{((\sum e_{ij}^2\ )/(\sum d_{ij}^2\ ))} \quad (3\text{-}3)$$

$$R^2(\%) = [1 - ((\sum d_{ij}^2\ - \sum e_{ij}^2\ )/(\sum d_{ij}^2\ )] \quad (3\text{-}4)$$

Where e_{ij} is the residual (the difference between the experimental and produced data from this algorithm), and d_{ij} is the data point of the experimental data matrix. Two different data matrixes (same number of elements) from the raw experimental data and principal component analysis reproduced were considered to compute the R^2 and LOF. The data set that produced the values closer to zero and hundred, respectively, were better suited for the SMCR-ALS algorithm. In another step, the initial concentration was calculated by applying the Evolving Factor Analysis (EFA) approach, which was used as the input to SMCR-ALS. The resolved concentration matrix C from SMCR-ALS was used to develop the BN (Figures 3-7 and 3-9). All the steps and calculations were explained in detail in the previous paper.[38] Figures 3-(6 &8) illustrate the results of resolving the data (D) by applying SMCR-ALS to the chemical ranks while trying to minimize the residual.

3.3.5. The Bayesian Networks

In the previous step – the cluster analysis – we obtained enough variables to develop the most probable reaction network (BNs) representing the investigated chemical reactions. By developing the BNs, we wanted to create a graphical comprehensive framework to remodel the probabilistic dependencies between these random variables $\mathbf{X} = \{X_1, X_2, \ldots, X_p\}$ as a directed acyclic graph (DAG), where $G = (\mathbf{V}, A)$. Each node $v_i \in \mathbf{V}$ corresponds to a random variable X_i.[39]

Two steps are required to complete the task of fitting a BN. The first step, structure learning, is completed by identifying the graphical structure of the BN, and the next step, called parameter learning, includes the estimation of the parameters of the global distribution obtained from the previous step.[40] To have an optimal BN, most of the time, more than one learning approach is applied.

3.4. RESULTS AND DISCUSSION

3.4.1. FTIR and ^{1}HNMR Results

After performing the hydrous pyrolysis experiments in the micro-batch reactor, samples were analyzed using two different spectroscopy techniques: FTIR and ^{1}H NMR. Figure 3-1 illustrates the FTIR spectra representing a functional group compositions analysis of the products of LG and PEB pyrolysis, and Table 3-2 presents the main functional groups according to the data reported in the literature.[41, 42]

By comparing the results, it can be seen that the major differences between samples provided from the pyrolysis of LG and PEB are those peaks located between 1500-1310 cm^{-1}, which represent the presence of aromatics and compounds containing phenolic functional groups in the PEB samples. Furthermore, mono, polycyclic, and substituted aromatic groups are presented by the absorption peaks between 1500-1400 cm^{-1}. However, in both cases, the other major oxygen-containing organics could be alcohols, carboxylic acids, ketones, and aldehydes.

Table 3-2. The main functional groups from the pyrolysis of LG and PEB.

Wave number(cm-1)	Functional groups	Compounds
3500-3200(s,b)	O-H stretch, H-bonded	Alcohols, Phenols
3300-3250(m)	O-H stretch	Carboxylic acids
3100-3000(s)	C-H stretch	Aromatics
3100-3000(m)	=C-H stretch	Alkenes
3000-2850 (m)	C-H stretch	Alkanes
1760-1665(s)	C=O stretch	Carbonyls(general)
1760-1690(s)	C=O stretch	Carboxylic acids
1740-1720(s)	C=O stretch	Aldehydes
1715((s)	C=O stretch	Ketones, Saturated aliphatic
1680-1640(m)	-C=C- stretch	Alkenes
1500-1400(m)	C-C stretch(in-ring)	Aromatics
1410-1310	C-O stretch	Phenols
1370-1350(m)	C-H rock	Alkanes
1320-1000(s)	C-O stretch	Alcohols, carboxylic acids, esters, ethers
1300-1100(w)	C=O stretch	Ketones
1000-650(s)	=C-H bend	Alkenes
725-720(m)	C-H rock	Alkanes

s=strong, m=medium, w=weak, n=narrow, b=broad, sh=sharp

Figure 3-2 represents the ^{1}H NMR spectrum, and Table 3-3 shows the hydrogen distribution of ^{1}H NMR of the products from the hydrous pyrolysis of LG and PEB. Resonances between 9 and 9.6 ppm were assigned to aldehyde structures. Therefore, the results from the hydrous pyrolysis of LG indicated the presence of a singlet peak around 9.60 ppm presenting an aldehyde with no neighboring carbon. This description is consistent with formaldehyde.[27] Results from the pyrolysis of PEB also presented resonances between 7.0 and 8.0 ppm, which were attributed to aromatic structures, and resonances between 5 and 7 ppm, which were mainly considered phenolic or non-conjugated olefins.

Table 3-3.^{1}H NMR results of the products of the pyrolysis of LG and PEB.

Type of hydrogen	Chemical shift[1]
Aldehyde	9.0-10.0
Aromatic	7.0-8.0
Phenolic or olefinic proton	5.0-7.0

1=ppm

3.4.2. Data Processing (Data Combination)

After signals were produced, the next step was completed by combining the data points (spectra and curves) obtained from FTIR and ^{1}H NMR spectrometers to produce a more reliable and

comprehensive data set than produced by each individual spectrometer. This step was completed using a low-level fusion procedure. Since the ^{1}H NMR spectra were acquired using an offline NMReady at the frequency of 60 MHz, there was no need to reduce the magnitude. Signals obtained from the two different spectrometer devices had different intensities. Therefore, they needed to be normalized. Normalizing to unit length and maximum peak intensity were considered, and the latter approach was chosen. Normalization was followed by concatenation of the FTIR and ^{1}H NMR signals to form the final fused signal. In the final spectrum, the first portion of the spectrum is the FTIR spectrum, and the next portion of the spectrum is the ^{1}H NMR spectrum. Figure 3-5 represents the final fused spectrum, which contained 3528 and 3665 data points from LG and PEB, respectively. All of these steps were implemented in MATLAB version 2018b and R version 3.5.1. The fused data were then processed and analyzed using standard multivariate processing methods, such as principal component analysis. The application of PCA was to assign data points in an order based on their contribution to the total variance of the data set. The final fused data sets were applied to develop the SMCR-ALS algorithms and the BNs. Finally, these results were compared with the ones provided by using data from only FTIR.

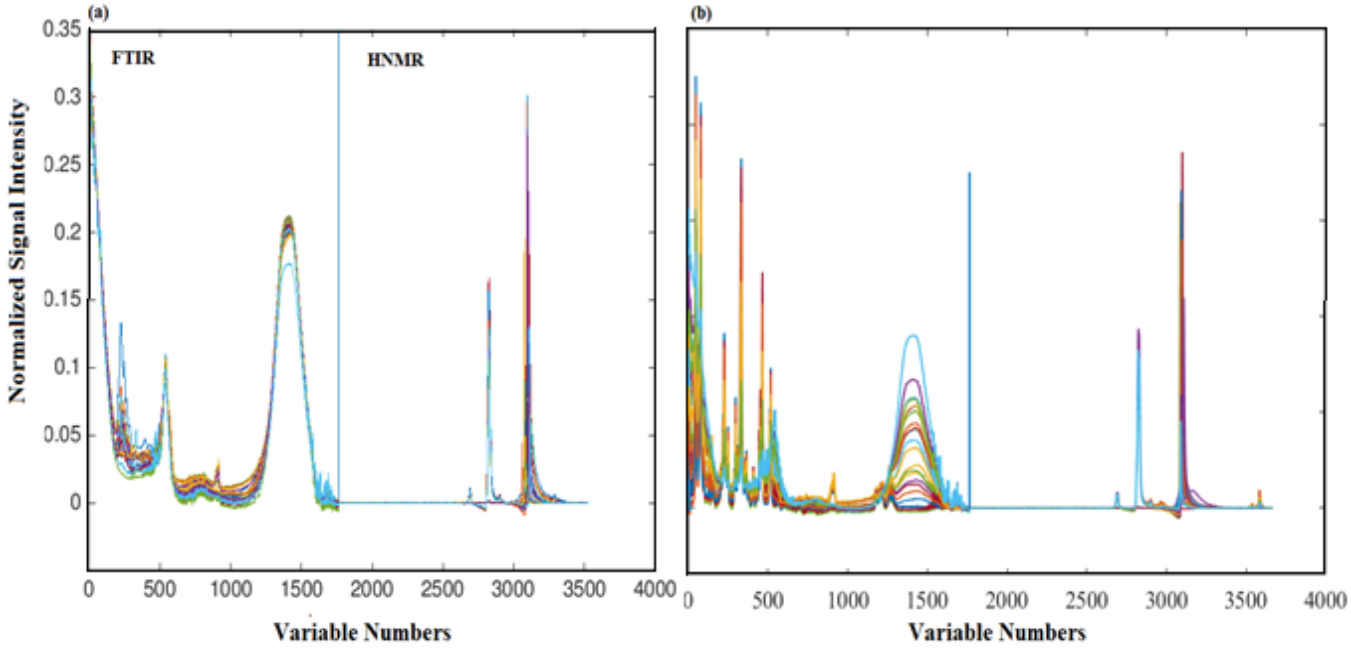


Figure 3-5. The final fused spectra for (a)LG, and (b) PEB.

3.4.3. Developing SMCR-ALS algorithm for PEB Decomposition

3.4.3.1. Data provided by FTIR Spectroscopy

The SMCR-ALS results for LG decomposition based on FTIR data were described in the previous paper; [26] therefore, the next part will present the SMCR-ALS algorithm just for PEB. Applying Eq. (3-1) helped us achieve the goal of this stage, which was deconvolving the spectra obtained from the investigated reactions into the spectral profiles and concentrations of those active chemical species (pseudocomponents) with distinct spectra. The novelty of this stage was including an optimization method (the alternating least squares regression procedure) to enable SMCR to incorporate any data-specific constraints. Figure 3-6 shows the result of developing this algorithm by using the data points provided by the FTIR technique, and Figure 3-7 presents the reaction network between these pseudo components.

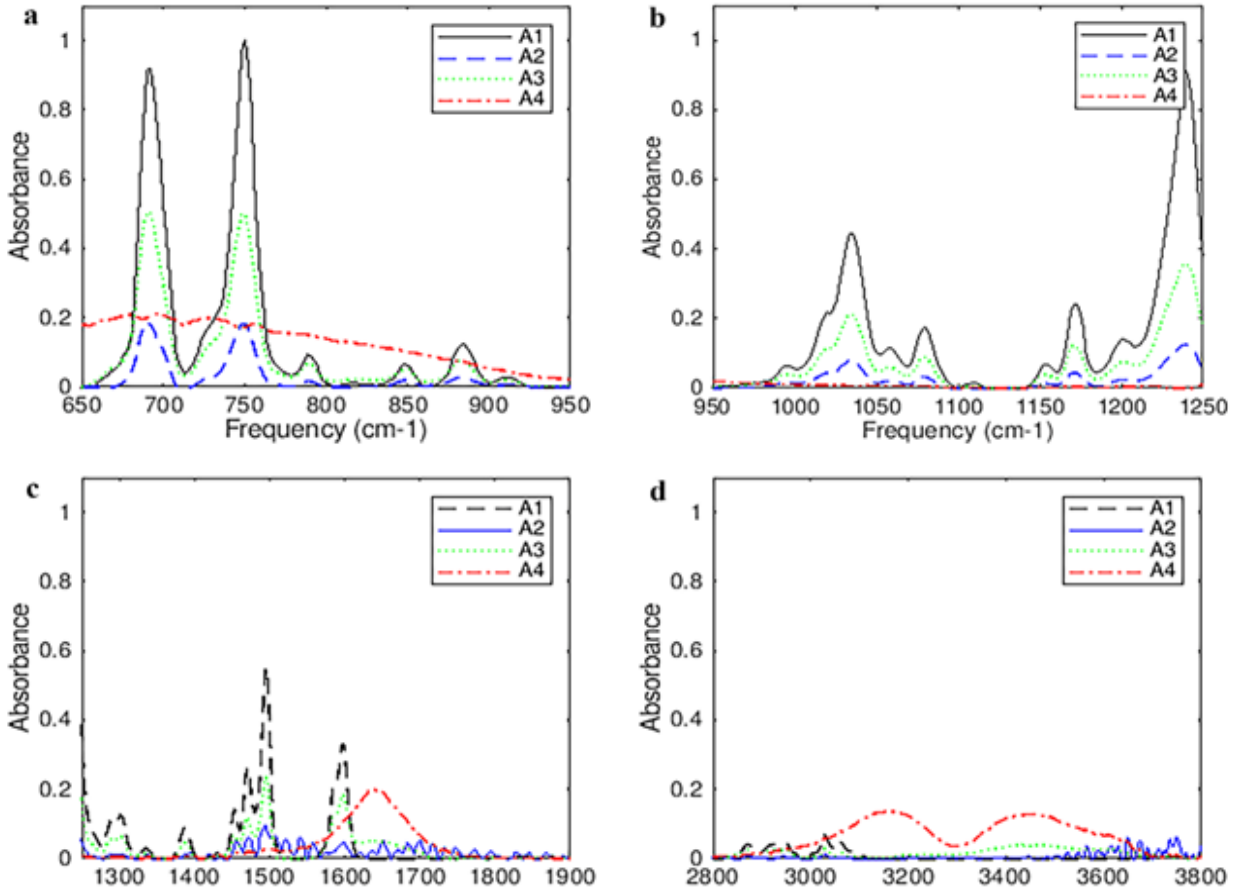


Figure 3-6. SMCR-ALS results for PEB conversion (data from FTIR) representing the resolved spectra for A_1, A_2, A_3, and A_4 by focusing on the major peaks.

Bear in mind that to have an appropriate interpretation regarding the spectra collected from the above algorithm; A_1, A_2, A_3, and A_4 (representing the first, second, third, and forth pseducomponents) should be considered to represent the actual molecules. To develop the reaction network, both greedy research methods (tabu and hill climbing) were used for the concentrations of A_1, A_2, A_3, and A_4, and both methods provided the same results (Figure 3-7). The one explanation for this BN could be that the first cluster (A_1) represents pseudocomponents mostly containing phenolic groups (the absorbance of peaks between 1410 and 1310 cm^{-1}). Cluster 2 (A_2) can been identified as aromatic compounds by C-C in-ring stretching vibrations with absorbance peaks from 1500 to1400 cm^{-1}, or C-H stretching vibrations at 900-675 cm^{-1}. Cluster 3 or A_3 (indicating compounds with carbonyl groups such as acids, aldehydes, and ketones with absorbance peaks around 1320-1000 cm^{-1}) can be produced by the cleavage of C-O-H and C=C bonds from clusters 1 and 2. Finally, cluster 4 (A_4) can represent pseudocomponents holding alkene groups (1650-1585 cm^{-1}, and 3100-3000 cm^{-1}) which can be produced by ring opening from cluster 2 or bond cleavage from cluster 3.

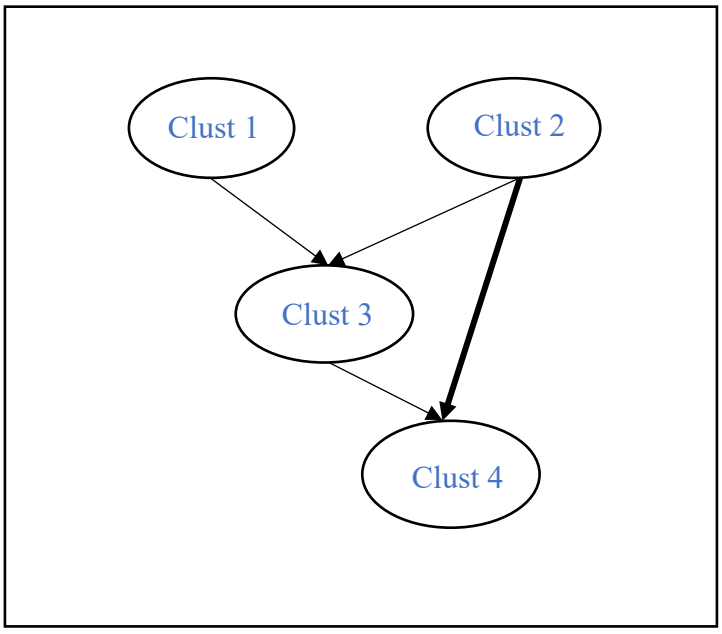


Figure 3-7. The BN obtained from SMCR-ALS for PEB decomposition.

The next Table (3-4) illustrates the strength values of each arc in the tabu and HC search-based BN learning approach. These values are valuable for two reasons. First, they show the strength of each arc (the highest negative number presents the most probable pathway or dependency).

Second, they have the benefit of being automatically updated as soon as the process receives the new information. According to the data provided by this table, the arcs between clusters 2 and 4(-24.62), and clusters 3 and 4 (-15.19) have the strongest dependencies or the highest probabilities for updating themselves. In other words, by removing these arcs, the overall score of the network will be decreased automatically by 24.26 and 15.19, respectively.

Table 3-4. The strength values of each arc in the tabu and HC search-based BN.

From	To	Arc strength
Cluster 2	Cluster 3	-8.01
Cluster1	Cluster 3	-7.81
Cluster 2	Cluster 4	-24.62
Cluster 3	Cluster 4	-15.19

As mentioned earlier, the second part of developing the BN is parameter learning aiming to quantify the uncertainty regarding the model which had been already developed. This step was completed based on the Markovian property which explains in a single path, the conditional probability of a parent (the ancestor) is the only one who is responsible for its child (the descendent). Hence, the conditional probability distribution of each group is used to develop a model for the mean value of the intensity. This value is determined by calculating the root mean square (RMS) value of the absorbance intensities of the group. To describe Θ which represents the set of parameters of the network, two constraints have been defined: X_i and μ_i. The former one has been considered as the intensity value of the i_{th} variable (i= 1,2,3), and the latter one as the mean value of X_i. Considering these conditions, the conditional probability distribution of each group has been shown in the pseudokinetic equations 3-5 to 3-8, and the model describing mean intensity can be found in the pseudokinetic equations 3-9 to 3-12.

$$P(X_1) \sim N(\mu_1, 0.36^2) \quad (3\text{-}5)$$
$$P(X_2) \sim N(\mu_2, 0.41^2) \quad (3\text{-}6)$$
$$P(X_3 | X_1, X_2) \sim N(\mu_3, 0.23^2) \quad (3\text{-}7)$$
$$P(X_4 | X_1, X_3) \sim N(\mu_4, 0.10^2) \quad (3\text{-}8)$$
$$\mu_1 = 0.36 \quad (3\text{-}9)$$
$$\mu_2 = 0.29 \quad (3\text{-}10)$$
$$\mu_3 = 0.90 - 0.64\mu_1 - 0.54\mu_2 \quad (3\text{-}11)$$
$$\mu_4 = 0.80 - 0.85\mu_1 - 0.50\mu_3 \quad (3\text{-}12)$$

The last four equations (3-9 to 3-12) not only indicate how much the mean value of the probability distribution of each cluster is related to others, but also make it possible to monitor an

online process in real time analysis. The latter task can be done by manipulating and controlling the process variables.

3.4.3.2. Data Acquired by the Data Combination Method

In the next step, the data set obtained by combination of the FTIR and ^{1}H NMR spectra from the PEB process was used to develop the SMCR-ALS algorithm. Figure 3- 8 presents the results. At first glance, there are still 4 pseudocomponents. Along with weak C=O stretching vibrations with absorbance peaks from 1300 to 1100 cm^{-1}, representing ketones or pseudocomponent one, there are three absorbance peaks in different regions presenting A_2, A_3, and A_4. The absorbance peaks between 3100 and 3000 cm^{-1} (=CH str) can represent the presence of alkenes (A_2), and the confirmation of the presence of compounds containing mostly phenolic groups (A_3) can be found by the absorbance peaks located between 1410 and 1310 cm^{-1}. Finally, the strong absorbance peak located at 2850 cm^{-1} confirms the species with aldehyde groups (C-H str) or A_4.

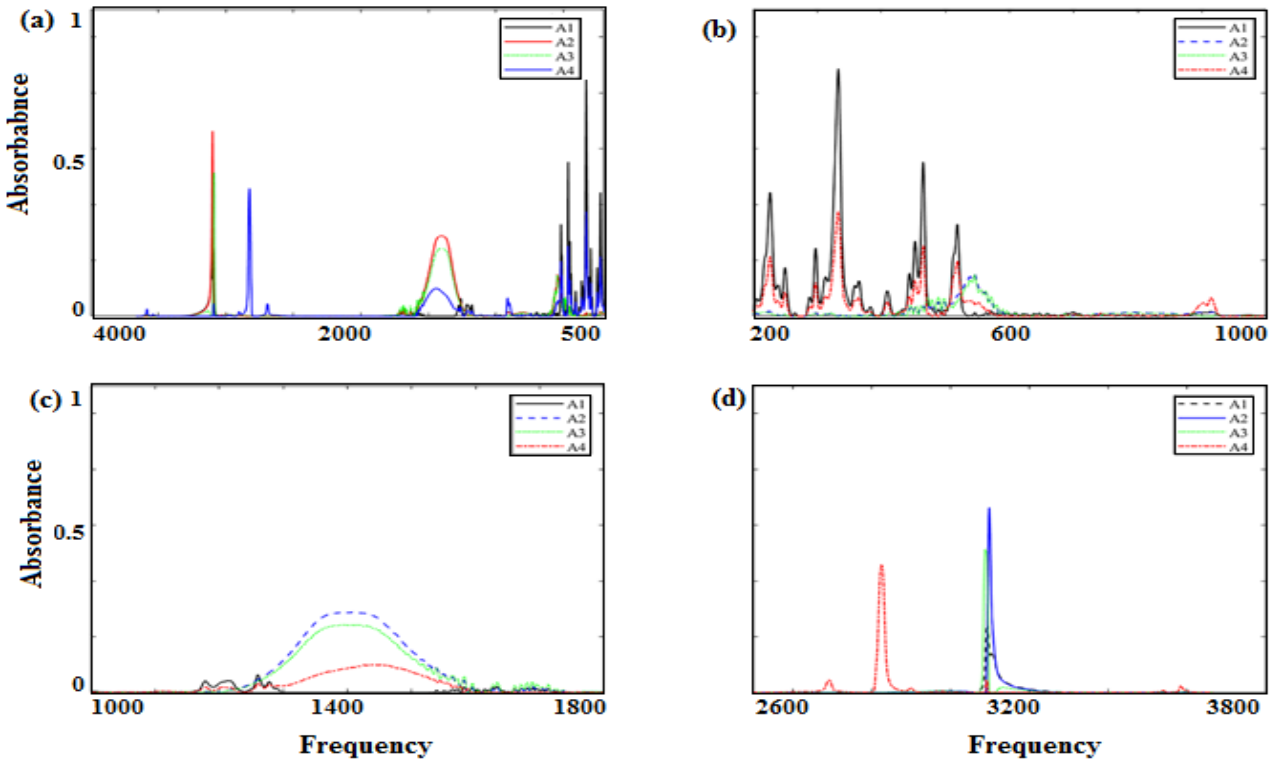


Figure 3-8. SMCR-ALS results for PEB conversion (data from data combination technique) representing the resolved spectra for A_1, A_2, A_3, and A_4, focusing on the major peaks.

Applying the BN developed the reaction network shown in Figure 3-9 between A_1, A_2, A_3, and A_4. This reaction network showed that the compounds mostly containing phenolic groups can be converted to alkenes and the major oxygen-containing organics, including ketones and aldehydes (a detailed explanation is provided with Figure 3-12).

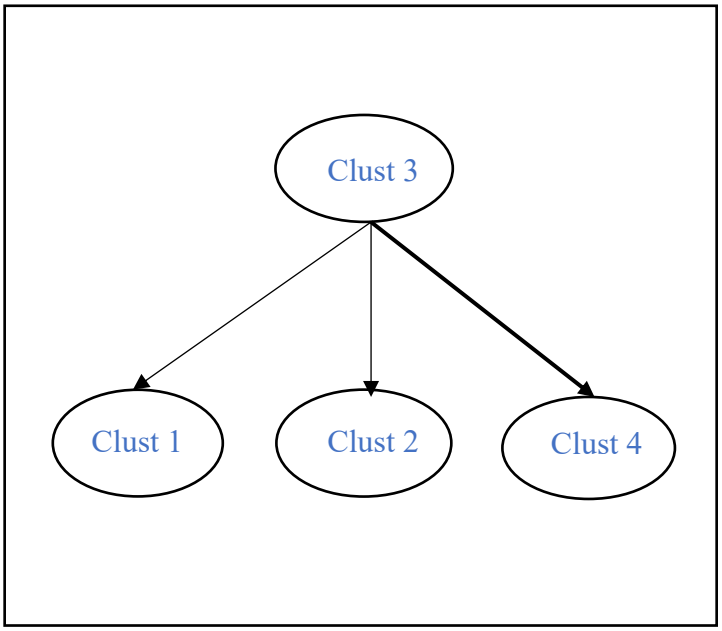


Figure 3-9. BN obtained for pseudocomponents from SMCR-ALS results for PEB decomposition by using fused data points.

According to the Table 3-5, the highest dependency or the most probable pathway (in this BN), belongs to the arc showing the conversion of compounds with phenolic functional groups to aldehyde groups (-7.67).

Table 3-5. The strength values of each arc in the tabu and HC search-based BN.

From	To	Arc strength
Cluster 3	Cluster 2	-1.19
Cluster 3	Cluster 1	-0.86
Cluster 3	Cluster 4	-7.67

To better understand the second part of BN, which is parameter learning, equations 3-13 to 3-16 present each group's conditional probability, and the models for these groups' mean intensities are shown in equations 3-17 to 3-20.

$P(X_1) \sim N(\mu_1, 0.26^2)$ (3-13)
$P(X_2 | X_1) \sim N(\mu_2, 0.57^2)$ (3-14)
$P(X_3 | X_1) \sim N(\mu_3, 0.28^2)$ (3-15)
$P(X_4 | X_1) \sim N(\mu_4, 0.19^2)$ (3-16)
$\mu_1 = 0.97$ (3-17)
$\mu_2 = 0.41 - 0.44\mu_1$ (3-18)
$\mu_3 = 0.75 - 0.56\mu_1$ (3-19)
$\mu_4 = 0.89 + 0.35\mu_1$ (3-20)

3.4.4. Three- cluster Bayesian Network for LG and PEB Decomposition

As mentioned in the previous paper, to develop an ultimate BN, three different methods – including two score-based (tabu and hill climbing) and one hybrid structure learning method (max–min hill climbing) – were applied.[26] All three learning structure methods (using data from FTIR) developed the same reaction networks with three clusters for both LG and PEB decomposition. Since the interpretation for LG decomposition using FTIR data was explained in the previous paper,[26] for now, we will mostly focus on PEB decomposition.

The chemical reactions and the physical processes in the thermochemical processes of lignin, including pyrolysis, gasification, and hydrothermal carbonization, are very complex and based on many factors.[21] Among different kinds of links in lignin, which need different energy to crack, aromatic units are usually bonded by interunit of C-C and C-O aryl ether bonds, including β-*O*-4 and α-O-4 linkages.[43] Between these two linkages, the C-O linkage is easier to break down due to its lower bond energy.[44]

The main products obtained in the pyrolysis of lignin are poly-substituted phenols, mono-phenols (phenol, syringol, and catechol), aromatic compounds, gases (H_2, CH_4, C_2H_4, C_2H_6, CO, and CO_2), and volatile compounds (acetone, acetaldehyde, and methanol).[45] Based on several research studies, the dissociation of the weakest linkages in lignin (the α-O-4 and β-O-4; BDE = 50-65 and 60-70 kcal. mol^{-1}, respectively) seemed to be responsible for forming these compounds. [46]

In this study, PEB, with the bond energy of the β-O4 bond of about 69 $kcal.mol^{-1}$, was chosen as the model compound for the lignin. According to Figure 3-10, at first glance, the phenoxy and 2-phenethyl radicals are expected to form from PEB decomposition. However, new pathways are possible. Britt et al.[47] proposed a Maccoll-like pathway involving intramolecular hydrogen

transfers via a six- and four-centred transition state, which results in closed-shell species (styrene and phenol). On the other hand, a Hoffman or retro-ene mechanism was proposed by Klein and Virk,[48] which can result in forming styrene and 2,4-cyclohexadienone. However, distinctions between these pathways are difficult to determine for two reasons: phenol and styrene can also be produced through the elimination or abstraction of phenoxy and phenylethyl radicals (the products of radical channel) in a reactor, and 2,4-cyclohexadienone and phenol can undergo rapid isomerization.

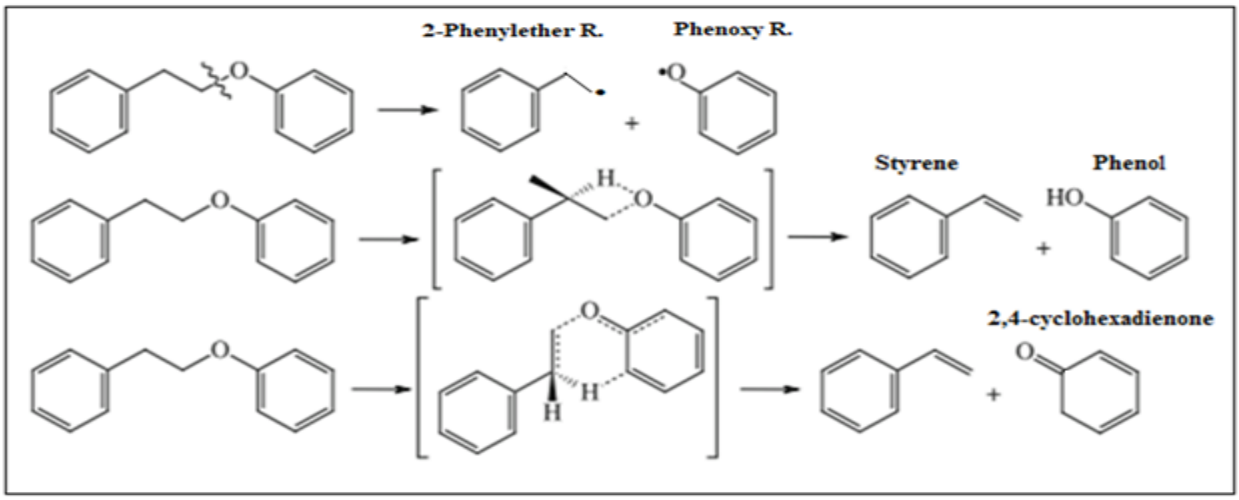


Figure 3-10. Pathways considered for the unimolecular decomposition of PEB.

According to Figure 3-11, the reaction networks for both LG and PEB are the same. However, it needs to be reminded that wavenumbers grouped in each network represent different functional groups.

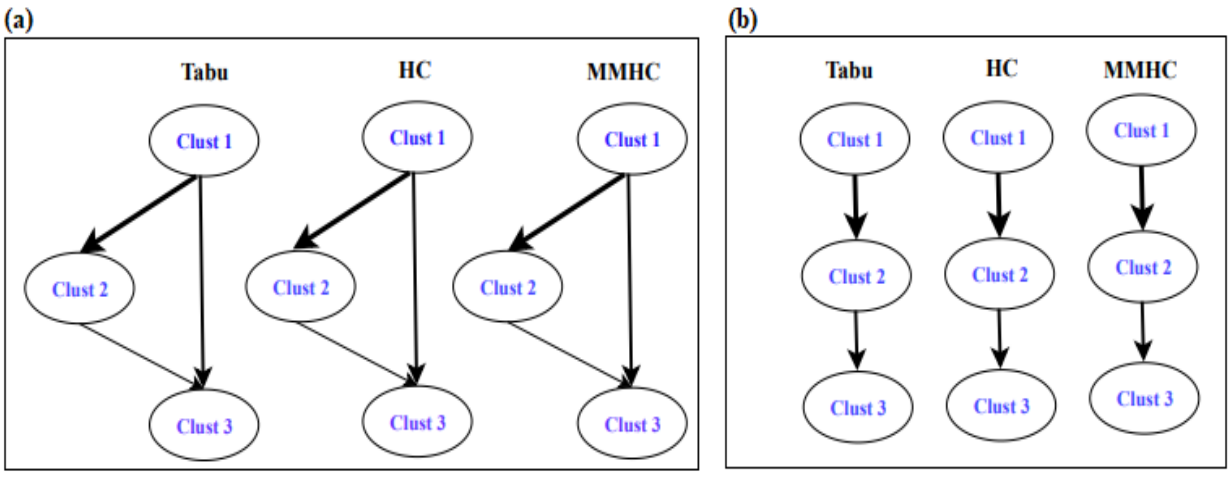


Figure 3-11. The BNs for LG and PEB conversion, data from (a) FTIR and (b) data combination.

According to these reaction networks, apparently by breaking down the β-O4 linkages, phenolic dimers such as bibenzyl and 2-benzylphenol can be formed through the hydrous pyrolysis of pseudocomponent 1 (PBE). To successfully describe these experimental observations, one explanation could be the initiation of a free-radical reaction pathway by the homolysis of the β-O_4 bond.[3] Per the proposed reaction network shown in Figure 3-12, cluster 2 (containing mostly phenolic groups) can be the result of hydrogen abstraction by the phenoxy radicals from homolysis.[49] It has to be kept in mind that due to the relatively stable formation of the phenoxy and benzyl radicals, PEB homolysis needs to be energetically favored. An unpaired electron could be responsible for resonance stabilization because the π bond is adjacent to single electrons in these radicals. Finally, lots of organic compounds, such as aromatics, alcohols, aldehydes, ketones, and alkenes – which were identified by absorbance peaks between 1900 and 1000 cm^{-1} – could be produced by the cleavage of C-O-C and C-O bonds and classified in cluster 3.

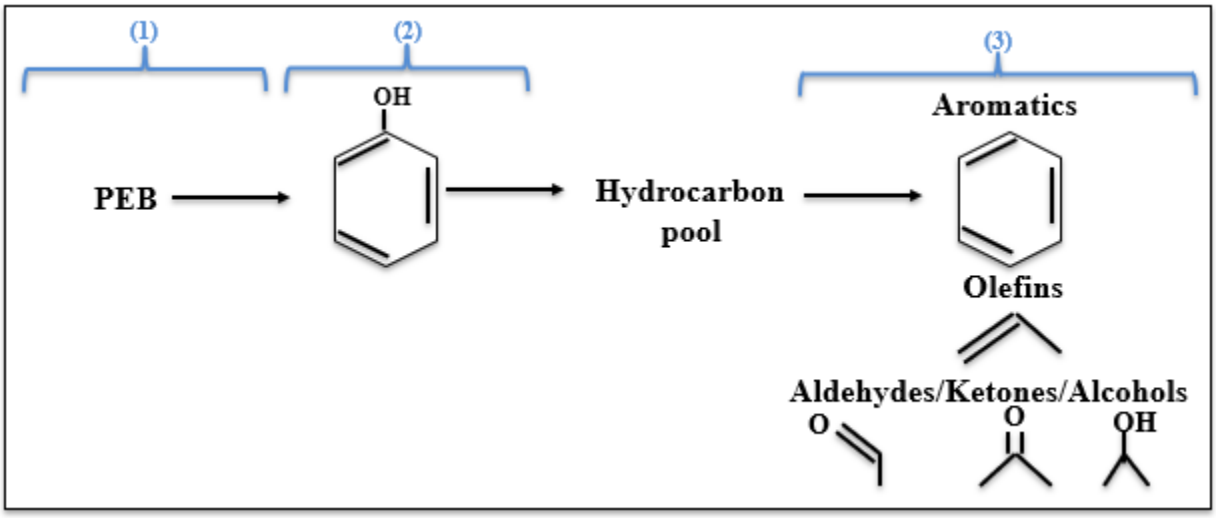


Figure 3-12. Proposed reaction network for PEB decomposition.

From the pyrolysis of the cellulosic biomass, a potential product could be LG (1, 6-anhydro-b-D-glucopyranose), which was used in this study. Transglycosilation is a common process in this reaction, even though there is a disagreement between its homolytic vs. heterolytic mechanisms. Through the thermal decomposition of LG, the glycosidic 1, 6-acetal bond can undergo ring opening, which can be accelerated in the presence of acidic catalysts in the heterolytic mechanism. Under the investigated conditions for the pyrolysis of LG, regardless of which data were used, one of the main final products was formaldehyde (cluster 3), which has a consistent agreement with the literature. Li and his colleagues used IR spectroscopy to measure the

formation of formaldehyde (CH_2O) and some gases, such as carbon monoxide (CO) and carbon dioxide (CO_2), as the major products of the thermal degradation of biomass, and they confirmed that the major precursor of forming formaldehyde was LG.[50] Based on Figure 3-11, it also can be concluded that glucose can be considered an intermediate product for this reaction.

Table 3-6. The strength values of each arc in the tabu, HC, and MMHC search-based BN (data provide by the combination method).

From	To	Arc strength
Cluster 1	Cluster 2	-7.30
Cluster 2	Cluster 3	-7.15

To complete the second part of the BNs, which is parameter learning, equations 3 (21-23) and 3(24-26) represent each group's conditional probabilities and mean intensities, respectively.

$$P(X_1) \sim N(\mu_1, 0.0028^2) \quad (3\text{-}21)$$
$$P(X_2 \mid X_1) \sim N(\mu_2, 0.003^2) \quad (3\text{-}22)$$
$$P(X_3 \mid X_2) \sim N(\mu_3, 0.003^2) \quad (3\text{-}23)$$
$$\mu_1 = 0.023 \quad (3\text{-}24)$$
$$\mu_2 = 0.064 + 1.06\,\mu_1 \quad (3\text{-}25)$$
$$\mu_3 = 0.14 + 0.74\mu_2 \quad (3\text{-}26)$$

3.4.5. Six-cluster Bayesian Network for PEB Decomposition

In addition to a reaction network that included three clusters (Figure 3-11), all three BN approaches provided the same results while considering six clusters (Figure 3-13). Generally, gas hydrocarbons, such as such as CH_4, CO_2, and CO (cluster 6), are the final products from lignin or PEB pyrolysis. Some researchers such as Huang et al.[51] have attempted to theoretically investigate the mechanism of the formation of these gases during the pyrolysis of lignin by applying density functional theory. Based on their proposal, these gases could be the results of concerted reactions, decarboxylation of phenyl (p-hydroxyphenyl, guaiacyl, and syringyl) formic acid, and decarbonylation of phenyl (p-hydroxyphenyl, guaiacyl, and syringyl) acetaldehyde, respectively. Other researchers such as Ferdosian et al. [52] tried to identify the different kinds of products in the pyrolysis of PEB. They confirmed that because of the hydrogen-donor/radical balance, at low temperature (same as was used in this study), volatile liquids, such as aldehydes

(acetaldehyde) and ketones (acetone), could be considered the major products (cluster 4), as could alcohols like methanol (cluster 3). Furthermore, cluster two can be represented by monophenols (phenol, guaiacol [*o*-methoxyphenol, **G**–H], and catechol [*o*-dihydroxybenzene]) and other polysubstituted phenolic compounds. Finally, cluster 5 could be classified as carboxylic acids, such as formic acid.

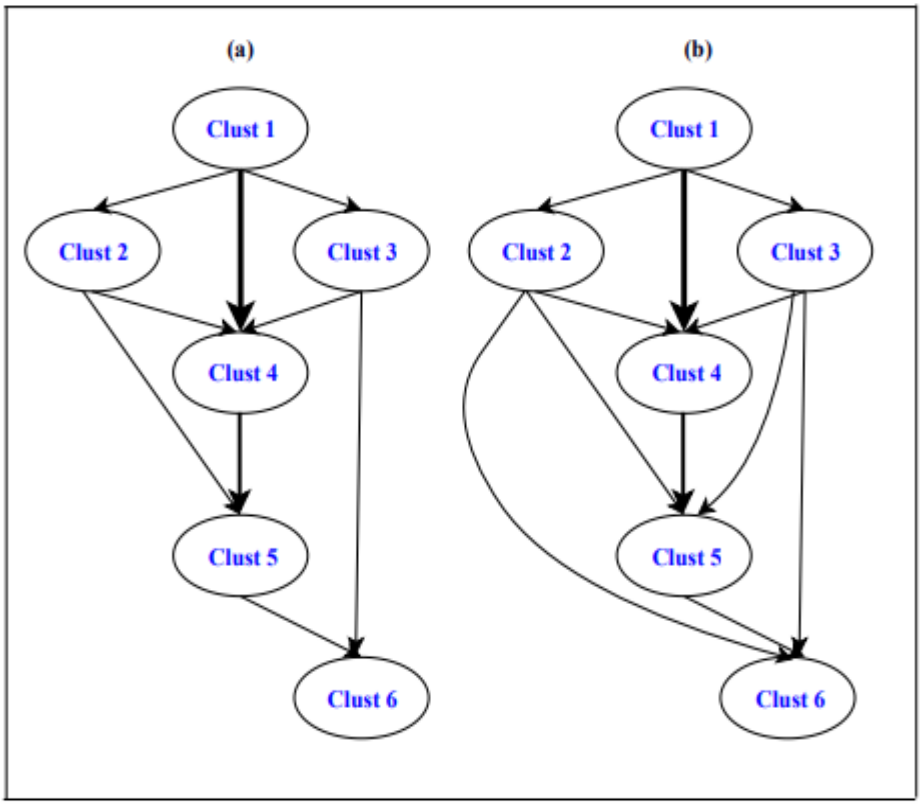


Figure 3-13. Six-cluster BNs for PEB conversion, data from (a) fusion method, and (b) FTIR.

Table 3-7. The strength values of each arc in the HC search-based BN (PBE, fused data, and 6 clusters).

From	To	Arc strength
Cluster 1	Cluster 2	-1.96
Cluster 1	Cluster 3	-11.57
Cluster 1	Cluster 4	-59.41
Cluster 2	Cluster 4	-3.94
Cluster 2	Cluster 5	-35.62
Cluster 3	Cluster 4	-2.01
Cluster 3	Cluster 6	-0.29
Cluster 4	Cluster 5	-43.37
Cluster 5	Cluster 6	-4.45

The conditional probability distribution of each group is described in equations 3(27-32), and the model for the groups' mean intensities is presented in equations 3(33-38).

$P(X_1) \sim N(\mu_1, 0.05^2)$ (3-27)
$P(X_2 | X_1) \sim N(\mu_2, 0.01^2)$ (3-28)
$P(X_3 | X_1) \sim N(\mu_3, 0.003^2)$ (3-29)
$P(X_4 | X_1, X_2, X_3) \sim N(\mu_4, 0.002^2)$ (3-30)
$P(X_5 | X_2, X4) \sim N(\mu_5, 0.009^2)$ (3-31)
$P(X_6 | X_3, X_5) \sim N(\mu_5, 0.009^2)$ (3-32)
$\mu_1 = 0.087$ (3-33)
$\mu_2 = 0.14 + 0.15\mu_1$ (3-34)
$\mu_3 = -0.009 + 0.061\ \mu_1$ (3-35)
$\mu_4 = 0.004 + 0.75\ \mu_1 - 0.12\ \mu_2 - 0.52\ \mu_3$ (3-36)
$\mu_5 = -0.014 + 2.32\ \mu_2 - 1.34\ \mu_4$ (3-37)
$\mu_6 = 0.038 - 0.99\ \mu_3 + 0.16\ \mu_5$ (3-38)

3.5. DISCUSSION

One of the major goals of this study was to investigate whether data fusion or combination can provide better data points as the input for developing the reaction network. After comparing the BNs developed from the FTIR and fused method, it can be concluded that the latter presented two major improvements. According to Figure 3-13(b), there are two arcs pointing out from clusters 2 and 3 to clusters 6 and 5, respectively. These arcs represent the direct conversion of compounds mostly containing phenolic groups to liquefied gases and producing carboxylic acids from alcohols. While several studies have confirmed the formation of gases, such as CO/CO_2, from phenolic compounds and acids from compounds with carboxyl groups through high-temperature and catalysts, such as $Rh/Ce_{0.13}Zr_{0.83}La_{0.04}O_2$ or the ruthenium complex $[RuCl_2(IPr)(p\text{-cymene})]$, [3,53] there is no direct evidence for these reactions under the investigated conditions (low temperatures and in the absence of those catalysts) in this study. Fortunately, the BNs provided from data fusion (Figure 3-13(a)) have outweighed these weaknesses and agree with the existing literature.[54]

By comparing the BNs (3 clusters) developed for LG decomposition and applying data from FTIR and fused data (Figure 3-11), the latter method provides the result that is more consistent with the results provided by the density functional theory (DFT). According to the results provided by this technique, since the activation energy required for dehydration is less than ring

opening for breaking C-O and C-C, converting LG to glucose through hydration is more favorable than converting it directly to the final products. [55] Many studies also confirmed the formation of glucose from the degradation of LG.[4, 56] These reasons confirm that data combination or data fusion can improve the results of reaction networks implemented by the BN.

In terms of generating nodes for the development of the BNs during online monitoring, the SMCR technique proves better in tracking chemistry changes across process conditions as compared to the BHC; as it resolves the concentration changes of pseudocomponent classes alongside returning its spectral signatures. However, BHC which groups the wavenumbers based on prior knowledge performs better in offline mapping to real chemistry without any background constraints on how chemistry changes over different process conditions. On the other hand, when it comes to develop a useful algorithm for monitoring online and controlling of multi-mixture reactions such as thermochemical conversion of biomass, SMCR algorithms can be the answer. Figure 3-14 shows the concentrations of A_1, A_2, A_3, and A_4 over the time of the reactions.

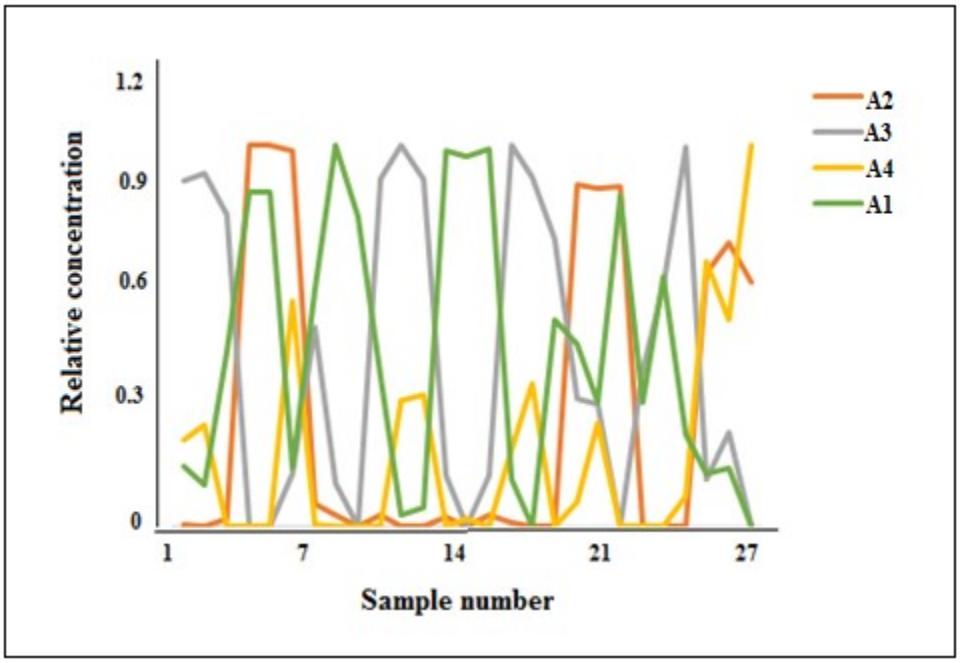


Figure 3-14. Corresponding concentration for pseudo components A_1, A_2, A_3, and A_4.

3.6. CONCLUSION

For the past few decades, there has been an increased interest in understanding and developing the pyrolysis pathways needed to obtain bio-oil. As a result, it seems necessary to have a better understanding of the reaction mechanism of biomass pyrolysis. To do this, there is still a challenge to elucidating the mechanism due to the high complexity of biomass structure and the process involved. This is why this research focused on the decomposition of LG and PEB (due to their simple structures and similar pyrolysis products) to better understand the pyrolysis process of cellulose and lignin. Hence, LG and PEB were investigated under fifty-four different conditions of hydrous pyrolysis reactions, and then the products were analyzed using FTIR and ^{1}HNMR spectroscopy techniques. Moreover, data mining and data fusion were applied. The former method was applied as a tool for data clustering and developing the BNs, and the latter was used as a combination of multiple sources to obtain improved or more relevant information that could contain the necessary aspects for developing the reaction networks for the investigated process. Applying three different Bayesian approaches (data provided by the combination method from FTIR and ^{1}HNMR) for both LG and PEB provided the same results considering three clusters, which agree strongly with the literature. Under the investigated conditions, LG can convert to glucose, which is the precedent for another cluster involving aldehydes/acids (e.g., formaldehyde). Regarding PEB decomposition, the final products (aldehydes/acids) can be produced from intermediate compounds containing mostly phenolic groups. In addition, an SMCR-ALS algorithm was developed for automatic resolution of spectra provided by online spectroscopic techniques which could be used for predicting the effect of intervention qualitatively and quantitatively. This algorithm resulted in deconvolving the whole spectrum into four pseudocomponents for this reaction (ROD=4). Then, the causality map between these pseudocomponents were traced and mapped. While interesting, the real value of this work is in the perception of the chemistry by applying data mining and data fusion techniques.

3.7. REFERENCES

1. Davis GA. The minerals sector, sectoral analysis, and economic development. *Resour Policy*. 1998;24(4):217-228.
2. Sasaki M, Fang Z, Fukushima Y, Adschiri T, Arai K. Dissolution and hydrolysis of cellulose in subcritical and supercritical water. *Ind Eng Chem Res*. 2000;39(8):2883-2890.
3. Choi YS, Singh R, Zhang J, et al. Pyrolysis reaction networks for lignin model compounds: unravelling thermal deconstruction of β-O-4 and α-O-4 compounds. *Green Chem*. 2016;18(6):1762-1773. doi:10.1039/C5GC02268A.
4. Yuan X, Cheng G. From cellulose fibrils to single chains: understanding cellulose dissolution in ionic liquids. *Phys Chem Chem Phys*. 2015;17(47):31592-31607. doi:10.1039/C5CP05744B.
5. Binder JB, Gray MJ, White JF, Zhang ZC, Holladay JE. Reactions of lignin model compounds in ionic liquids. *Biomass and Bioenergy*. 2009;33(9):1122-1130. doi:10.1016/j. biombioe.2009.03.006.
6. Malaluan RM. A Study on Cellulose Decomposition in Subcritical and Supercritical Water. Ph.D. Dissertation, Tohoku University, Sendai, Japan, 1995.
7. Alon U. Network motifs: Theory and experimental approaches. *Nat Rev Genet*. 2007;8(6):450-461. doi:10.1038/nrg2102.
8. Soranzo N, Altafini C. ERNEST: A toolbox for chemical reaction network theory. *Bioinformatics*. 2009;25(21):2853-2854. doi:10.1093/bioinformatics/btp513.
9. Zhang S, Ding HUA, Wang X. Research and application of structure learning algorithm. 2003; *International Journal of Engineering and Advanced Technology*. (November):2-5.
10. Pourret O. *Bayesian Networks: A Practical Guide to Applications*. Chichester, West Sussex, Eng: John Wiley; 2008.
11. Neapolitan RE. *Learning Bayesian Networks*. Upper Saddle River, NJ: Pearson Prentice Hall; 2004.
12. Abidi N, Manike M. X-ray diffraction and FTIR investigations of cellulose deposition during cotton fiber development. *SAGE.* 2018;(3). doi:10.1177/0040517516688634.
13. Darbeau RW. Nuclear Magnetic Resonance (NMR) spectroscopy: A review and a look at its use as a probative tool in deamination chemistry. *Applied Spectroscopic Reviews*. 2006;4928. doi:10.1080/05704920600726175.
14. Ayari I HJ. A framework for multi-sensor data fusion. *Proc IEEE Symp Emerg Technol Fact Autom*. 1995;2:51-59.
15. Cremer F, den Breejes E KS. Sensor fusion for anti-personnel land mines detection. *Proc 3rd eurofusion Conf*. 1998;4:63-70.
16. Luo RC, Kay MG. Multisensor integration and fusion: Issues and approaches. *Proc SPIE*. 1988;0931(August 1988):42-49. doi:10.1117/12.946646.
17. Bennett NM, Helle SS, Duff SJB. Extraction and hydrolysis of levoglucosan from pyrolysis oil. *Bioresour Technol*. 2009;100(23):6059-6063. doi:10.1016/j.biortech.2009.06.067
18. Minowa T, Zhen F, Ogi T. Cellulose decomposition in hot-compressed water with alkali or nickel catalyst. *Journal of Super Critical Fluids*. 1998; 13:253-259.
19. Kidder MK, Britt PF, Chaffee L, Buchanan AC. Confinement effects on product selectivity in the pyrolysis of phenethyl phenyl ether in mesoporous silica. *Chem Soc*. 2007:52-54. doi:10.1039/b613757a.

20. Nygård HS, Olsen E. Review of thermal processing of biomass and waste in molten salts for production of renewable fuels and chemicals. *Int J Low-Carbon Technol*. 2012;7(4):318-324. doi:10.1093/ijlct/ctr045.
21. Smolders K, Van de Velden M BJ. Operating parameters for the bubbling fluidized (BFB) and circulating fluidized bed (CFB) processing of biomass, In: Proceedings of ‘‘use of renewables.’ *Achema: Frankfurt*. 2006;(Paper 1030).
22. Boblete O. Hydrothermal degradation of polymers derived from plants, progress in polymer science. *Polym Sci*. 1994;19(799).
23. T. Adschiri, S. Hirose, R. Malaluan KA. Noncatalytic conversion of cellulose in supercritical and subcritical water, J. *Chem Eng Jpn*. 1993;26(676).
24. W.S. Mok, M.J. Antal Jr., GV. Productive and parasitic pathways in dilute acid-catalyzed hydrolysis of cellulose,Ind. *Eng Chem*. 1992;32(94).
25. Kabir G, Hameed BH. Recent progress on catalytic pyrolysis of lignocellulosic biomass to high-grade bio-oil and bio-chemicals. *Renew Sustain Energy Rev*. 2017.
26. Sattari F, Tefera A, Sivaramakrishnan K, De Klerk A, Prasad V. Postulating Pseudo-reaction Networks for the Conversion of Levoglucosan in Hydrous Pyrolysis Using Spectroscopic Data & Self-Modeling Multivariate Curve Resolution. *Submitted to I & EC*. 2019.
27. Mistry B. *A Handbook of Spectroscopic Data Chemistry (UV, IR, PMR, 13CNMR and Mass Spectroscopy)*. Oxford Book Company. 2009. doi:10.1063/1.2719251.
28. Roussel S, Bellon-Maurel VE, Roger JM, Grenier P. Authenticating white grape must variety with classification models based on aroma sensors, FT-IR and UV spectrometry. *J Food Eng*. 2003;60(4):407-419. doi:10.1016/S0260-8774(03)00064-5.
29. Edwards I, Gross XE, Lowden DW SP. Fusion of NDT data. *Br J NDT*. 1993;35(12):710-713.
30. Ester, M., Kriegel, H.P., Sander, J., Xu X. A Density-based algorithm for discovering clusters in large spatial databases with noise. *AAAI Press*. 1996:226-231.
31. Ankerst M, Breunig MM, Kriegel H-P, Sander J. Optics: Ordering points to identify the clustering structure. *ACM Sigmod Rec*. 1999:49-60. doi:10.1145/304182.304187.
32. F. Murtagh P. Contreras, Methods of hierarchical clustering. Department of Computer Science, Royal Holloway, University of London. 2011:CoRR abs/1105.0121.
33. Heller K a., Ghahramani Z. Bayesian hierarchical clustering. *Proc 22nd Int Conf Mach Learn*. 2005:297-304. doi:10.1145/1102351.1102389.
34. Sirinukunwattana K, Savage RS, Bari MF, Snead DRJ, Rajpoot NM. Bayesian hierarchical clustering for studying cancer gene expression data with unknown statistics. *PLOS One*. 2013;8(10). doi:10.1371/journal.pone.0075748.
35. Tefera DT, Agrawal A, Yañez Jaramillo LM, De Klerk A, Prasad V. Self-modeling multivariate curve resolution model for online monitoring of bitumen conversion using Infrared Spectroscopy. *Ind Eng Chem Res*. 2017;56(38):10756-10769. doi:10.1021/acs.iecr.7b01849.
36. Zhang X, Tauler R. Application of multivariate curve resolution alternating least squares (MCR-ALS) to remote sensing hyperspectral imaging. *Anal Chim Acta*. 2013;762:25-38. doi:10.1016/j.aca.2012.11.043.
37. D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi and JS-V. Handbook of chemometrics and qualimetrics: Part A. *Chem Inf Comput Sci*. 1998;38(6):1234–1254.

38- Tefera D, Tamiru Y J, Maria L, Ranjan R, Li C, De Klerk A, Prasad V. A Bayesian learning approach to modeling pseudoreaction networks for complex reacting systems: Application to the mild visbreaking of Bitumen. *I & EC*. 2017;56(8): 1961-1970.
39. Nagarajan R, Scutari M, Lèbre S. *Bayesian Networks in R with Applications in Systems Biology*. Springer science.; 2013. doi:10.1007/978-1-4614-6446-4.
40. Koller D FN. Probabilistic graphical models: principles and techniques. *MIT Press Cambridge*. 2009.
41. Li G, Torraca G, Jing W, Wen Z qing. Applications of FTIR in identification of foreign materials for biopharmaceutical clinical manufacturing. *Vib Spectrosc*. 2009;50(1):152-159. doi:10.1016/j.vibspec.2008.10.016.
42. Bassilakis R, Carangelo RM, Wojtowicz MA. TG-FTIR analysis of biomass pyrolysis. *Fuel*. 2001;80(12):1765-1786.
43. Dorrestijn E, Laarhoven LJJ, Arends IWCE, Mulder P. Occurrence and reactivity of phenoxyl linkages in lignin and low rank coal. *J Anal Appl Pyrolysis*. 2000;54(1):153-192. doi:10.1016/S0165-2370(99)00082-0.
44. Guadix S, Meenakshisundaram M. Review on catalytic cleavage of C – C inter-unit linkages in lignin model compounds: Towards lignin depolymerisation. *Top Catal*. 2018;61(3):183-198. doi:10.1007/s11244-018-0909-2.
45. Liu WJ, Jiang H, Yu HQ. Thermochemical conversion of lignin to functional materials: a review and future directions. *Green Chem*. 2015;17(11):4888-4907. doi:10.1039/c5gc01054c.
46. Kim S, Chmely SC, Nimos MR, Bomble YJ, Foust TD, Paton RS BG. Computational study of bond dissociation enthalpies for a large range of native and modifi ed lignins. *J Phys Chem Lett*. 2011;2(22):2846–2852.
47. Britt PF, Buchanan AC, Cooney MJ MD. Flash vacuum pyrolysis of methoxy-substituted lignin model compounds. *J Org Chem*. 2000;65(5):1376-1389.
48. Klein MT VP. Model pathways in lignin thermolysis 1. Phenethyl phenyl ether. *Ind Eng Chem Fund*. 1983;22(1):35-45.
49. Jarvis MW, Daily JW, Carstensen H, et al. Direct detection of products from the pyrolysis of 2-Phenethyl Phenyl Ether. *J. Phys. Chem.* 2011:428-438.
50. Li S, Lyons-Hart J, Banyasz J, Shafer K. Real-time evolved gas analysis by FTIR method: An experimental study of cellulose pyrolysis. *Fuel*. 2001;80(12):1809-1817. doi:10.1016/S0016-2361(01)00064-3.
51. Huang J, Liu C, Tong H, Li W, Wu D. A density functional theory study on formation mechanism of CO, $CO_2$and CH_4 in pyrolysis of lignin. *Comput Theor Chem*. 2014;1045:1-9. doi:10.1016/j.comptc.2014.06.009.
52. Ferdosian CX and F. Conversion of lignin into bio-based chemicals and materials, green chemistry and sustainable technology. *Springer-Verlag GmbH Ger*. 2017;DOI 10.100.
53. Constantinou DA, Álvarez-galván MC, Luis J, Fierro G, Efstathiou AM. Applied catalysis B: environmental low-temperature conversion of phenol into CO, CO_2 and H_2 by steam reforming over La-containing supported Rh catalysts. *Applied Catal B, Environ*. 2012;117-118:81-95. doi:10.1016/j.apcatb.2012.01.005.
54. Santilli C, Makarov IS, Fristrup P, Madsen R. Dehydrogenative synthesis of Carboxylic acids from primary alcohols and hydroxide catalyzed by a Ruthenium N - heterocyclic carbene complex. *JOC*. 2016;. doi:10.1021/acs.joc.6b02105.
55. Nimlos MR, Evans RJ. Levoglucosan pyrolysis. *Fuel Chem Div Prepr*. 2002;47(1):393.

56. Abella L, Nanbu S, Fukuda K, May R. A theoretical study on Levoglucosan pyrolysis reactions yielding aldehydes and a ketone in biomass. *Kyushu University*. 2007;67(2):67-74.

4. DEVELOPING THE MOST PROBABLE REACTION NETWORKS FOR THERMOCHEMICAL CONVERSION OF A PHYSICAL MIXTURE OF LEVEOGLUCOSAN AND 2-PHENOXYETHYL BENZENE BY USING DATA FUSION AND CHEMOMTERIC TECHNIQUES (SMCR-ALS)

A version of this chapter will be submitted as; Fereshteh Sattari, Arno de Klerk and Vinay Prasad," Developing the Most Probable Reaction Networks for Thermochemical Conversion of a Physical Mixture of Levoglucosan and 2-Phenoxyethyl benzene by Using Data Fusion and Chemometric Techniques (SMCR-ALS)".

4.1. BACKGROUND

Different forms of fossils fuels such as coal, natural gas, and petroleum, currently, are the main suppliers of energy.[1] However, researchers have increasingly been investigating alternative renewable energy sources such as hydroelectricity, solar, or biomass. The most important reasons behind this are: (a) reducing environmental impacts related to fossil fuels (b) increasing and securing the energy sources and (c) preserving and increasing agricultural activities.[2] Among various kinds of renewable sources, biomass has received lot of attention because it is the world's most abundant such resource.[3]

Biomass as an organic compound mostly contains extensive chains of carbon atoms connected to macromolecules. The major components of biomass which are cellulose, hemicellulose, and lignin, are the results of the chemical linkages between these carbons and other elements such as oxygen or sometimes nitrogen or sulfur. Cellulose (40-50%) is a homogeneous polymer that consists of dehydrated glucose molecules linked by β-1, 4-glycoside linkages $(C_6H_{10}O_5)_n$. In contrast, hemicellulose (15-30%) with an amorphous and heterogeneous structure consists of a group of monomers including glucose, glucuronic acid, arabinose, xylose, mannose and galactose. Hemicellulose also acts as a bond between cellulose and lignin. Lastly, lignin with different phenylpropane units has a very complex three-dimensional structure. Interestingly, these units are linked by (C-C) or (C-O) bonds linked to some side chains like methoxy.[4]

Among different methods for biomass conversion, thermo-chemical and bio-chemical have been the two most important processes. Hydrous pyrolysis or hydrothermal liquefaction (HTL) is a method of the former technique in which air is absent and hot water acts as a medium. Amongst different advantages linked to this process, safety, high quality of the products and the maximum liquid yield are the most significant ones.[5]

In a complex system such as HTL, one of the ultimate goals is to understand how the behavior of the system or reaction networks could be affected by the interactions at the molecular level or by network topology.[6] To this end, an accurate model is required to develop the reaction network and reveal essential characteristics of the investigated system.[7] However, encoding the presence of multispecies in a complex system along with developing the causality between these groups is considered a major challenge for researchers.[8] This is why for the past few decades, analytical

science has focused and made great progress in developing mathematical approaches to trace and map the complexity of these systems.

Currently, Granger and the Bayesian (BNs) causality detections are the two main approaches to interpret the causal relationships between elements (genes or chemical species) of a group of multispecies in a data set.[9, 10] The aim of the former method's (initially developed by Wiener and Granger) is to determine the causal impact of one time series on another by predicting how the knowledge of a time series can be improved with the information of the second one. This approach has the benefit of finding the frequencies responsible for interactions of those elements by having a corresponding frequency domain decomposition.[11] In contrast, the Bayesian network (often applied to statistical data) represents the causal relationship among the nodes by using the joint probability distribution which is factorized in terms of a Markovian blanket. There are some rationalizations why the BN is a more appropriate technique for this study. First, this learning method is more useful when it comes to dealing with high dimensional data sets with fewer samples (applicable to the conditions of this study, with 1764 wavenumbers or variables and 27 samples). [12] Second, BNs approach is computationally tractable and general enough to be applicable to real processes, such as combustion, petroleum refining and HTL process.[13]

In this study, the required data for developing the BNs was collected by either a single spectroscopic technique such as Fourier transform infrared spectrometry (FTIR) or multiple techniques such as Proton nuclear magnetic resonance (^{1}H NMR) and FTIR combined. Then, a low-level data fusion technique was applied to combine data sets collected from FTIR and ^{1}H NMR as if they were a single signal. The goal of this step was to utilize, analyze and collect more comprehensive measure resulting from different types of sources in order to ultimately capture more information and characteristics of the research object.[14] In general, multi-sensor data fusion offers substantial statistical advantages (via redundant observations) compared to a single data source. This can happen by increasing the accuracy while observing and characterizing a quantity (or an objective) from different points of view. [15]

In data analyzing field, spectroscopic techniques are considered useful. However, there are some concerns about their interpretations due to their high dimensional structure that results in overlapping between their data points.[16, 17] The second goal of this study was to resolve this problem by applying a family of chemometric methods such as multivariate curve resolution

techniques called self-modeling multivariate curve resolution (SMCR).[18] This approach is a mathematical technique to obtain the required necessary information related to the objects of interest by resolving the whole spectrum into spectral profile and concentration of pseudo-components changing during the reaction [19] and it was initially introduced by Lawton and Sylvester for multivariate imaging data analysis.[20] Resolved spectral profiles and concentrations of pseudo-components can be considered as a tool for on-line monitoring in real time analysis and developing the most likely reaction network for biomass conversion, respectively.

4.2. MATERIALS

More than 80% of biomass is made of cellulose and lignin. Therefore, in this study, a physical mixture (PM) of levoglucosan (LG) and 2-Phenoxyethyl benzene (PEB) were utilized to represent cellulose and lignin in biomass. LG can be considered as an anhydosugar which is the result of the thermal decomposition of cellulose (150- 350 °C)[21] and its molecular structure consists of three hydroxy and two ether groups connected to a pyranose ring. [22] PEB ($C_{14}H_{14}O$) can be considered as a model compound for lignin by representing the dominant ether linkage in lignin.[23]As shown in Figure 4-1, the methodology of this research mostly includes data acquisition and data processing in which each step will be explained in detail in the following sections.

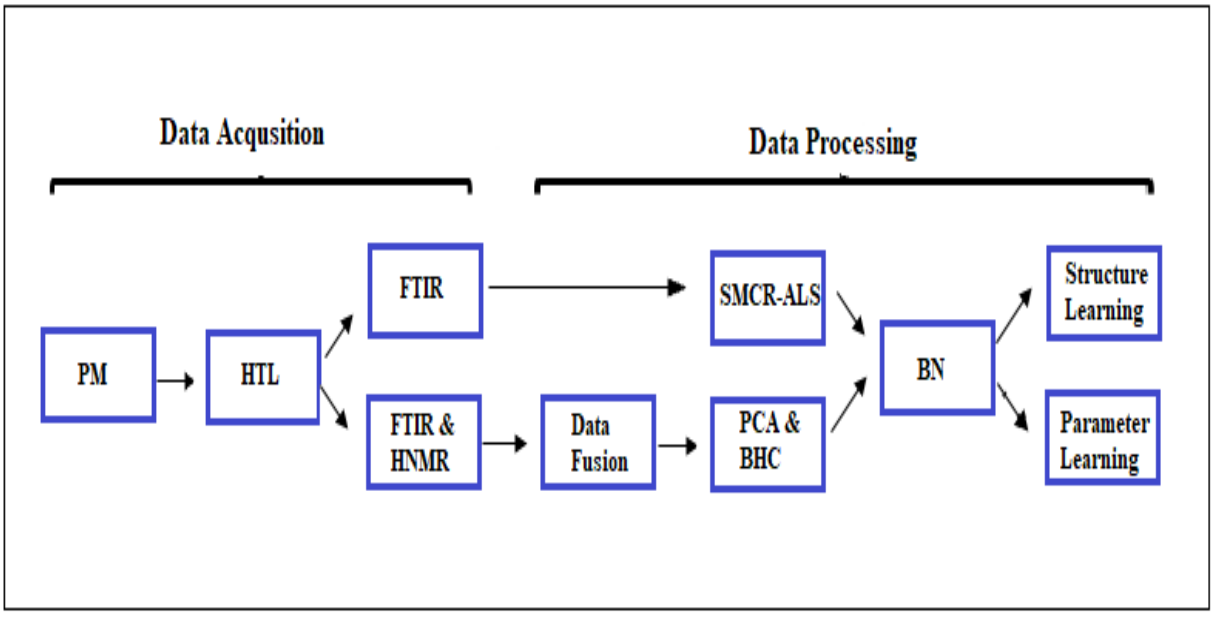


Figure 4-1. Proposed research methodology.

4.3. DATA ACQUISITION

For data collection, a physical mixture of LG and PEB (1:1 weight ratio) went through the twenty-seven different conditions of hydrous pyrolysis reactions in non-catalyzed conditions, and in the presence of catalysts- sulfuric acid (0.05 M) and sodium hydroxide (1 M). In all conditions, the volume ratio of the physical mixture to the medium was 1:10. Supplementary Table 4-A$_{-1}$ and Figure 2-1 present the variation in ranges of temperature and residence time in a stainless-steel micro batch-reactor and the experimental setup for the hydrous pyrolysis of the physical mixture respectively.

After hydrous pyrolysis, samples were ready for characterization. In this stage, two different spectroscopic techniques (FTIR and ^{1}H NMR) were applied. FTIR was chosen since it has the ability of fast measurement besides providing information related to asymmetric stretching of bonds. Therefore, the spectra were collected using an ABB MB 3000 FTIR spectrometer, at 2 cm^{-1} resolutions over the spectral range 4000 to 600 cm^{-1} and the results are shown in Figure 4-2.

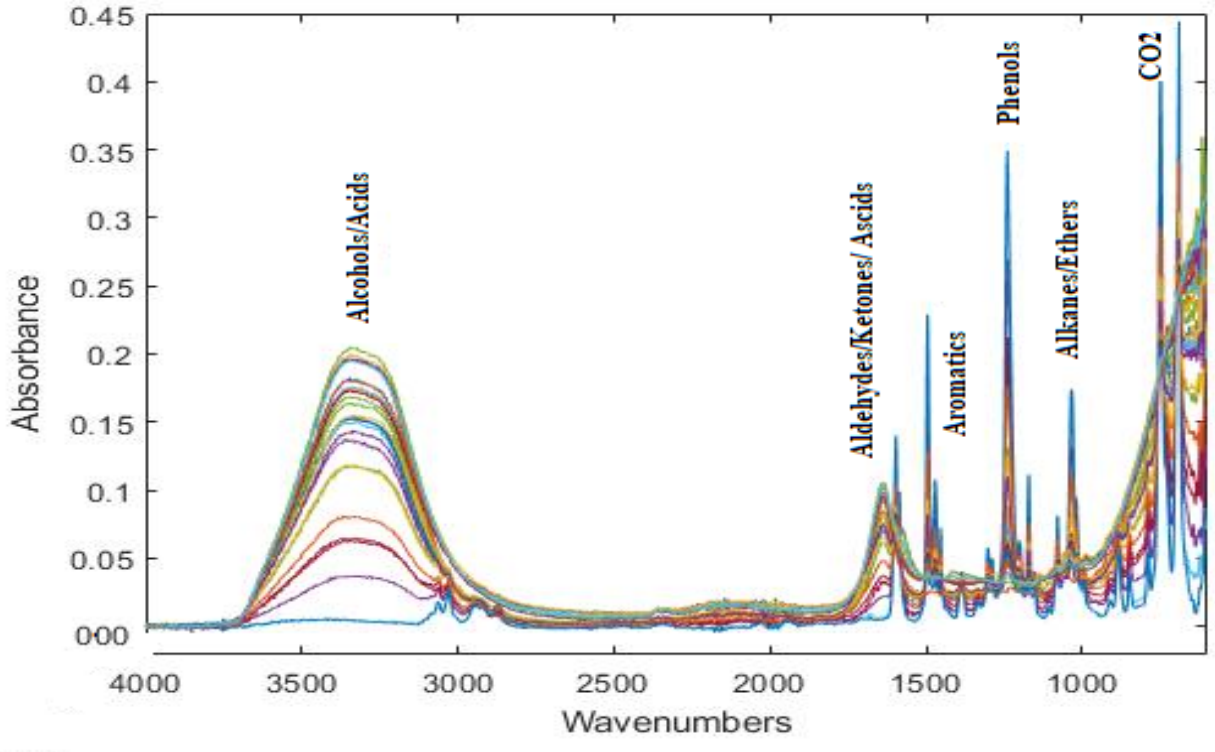


Figure 4-2. Twenty-seven FTIR spectra of HTL of the physical mixture (PM) of LG and PEB.

To determine the entire chemical structure of samples, a magnetic resonance imaging technique called ^{1}H NMR was used. This one-dimensional technique takes advantage of distinguishing between aromatic and saturated chains of alkanes by generating spectra from unique proton environments. NMReady at the frequency of 60 MHz with a resolution of FWHM<1.0 Hz (20 ppb) was used for all ^{1}H NMR measurements and data acquisition (Figure4-3). To provide an appropriate interpretation, these steps were followed by reviewing some handbooks of spectroscopic data.[24, 25]

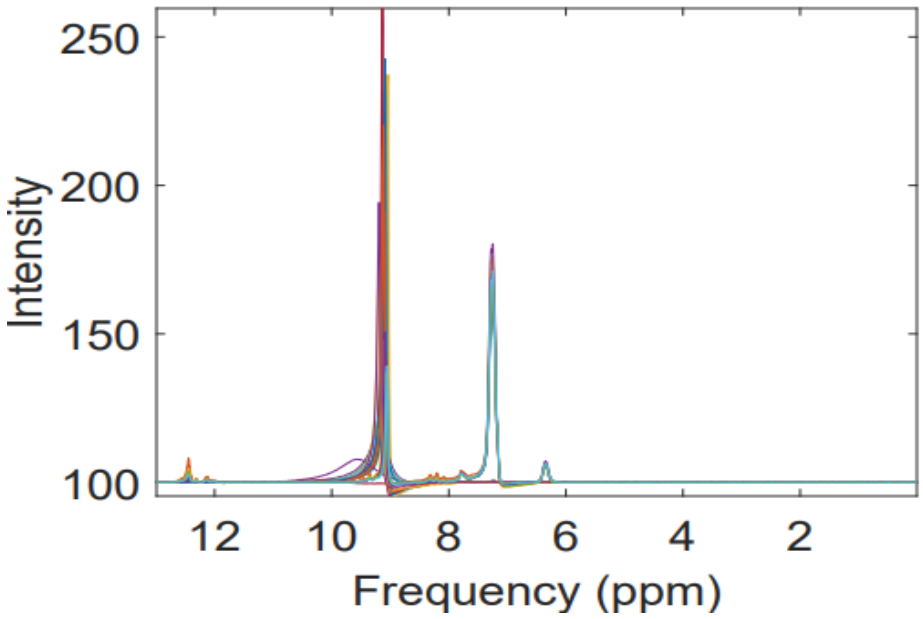


Figure 4-3.^{1}H NMR spectra for PM conversion in the presence of SCW and NaOH.

4.4. DATA PROCESSING

Over the past few decades, data fusion and data mining have been applied extensively for obtaining meaningful information from raw data. Data fusion was employed focusing on combining data from multiple sources to obtain improved and more reliable information compared to the result from a single source. Data mining has been used to extract implicit and useful information from investigated datasets by clustering their data points (by finding a common control unit) and then predicting the outcomes.

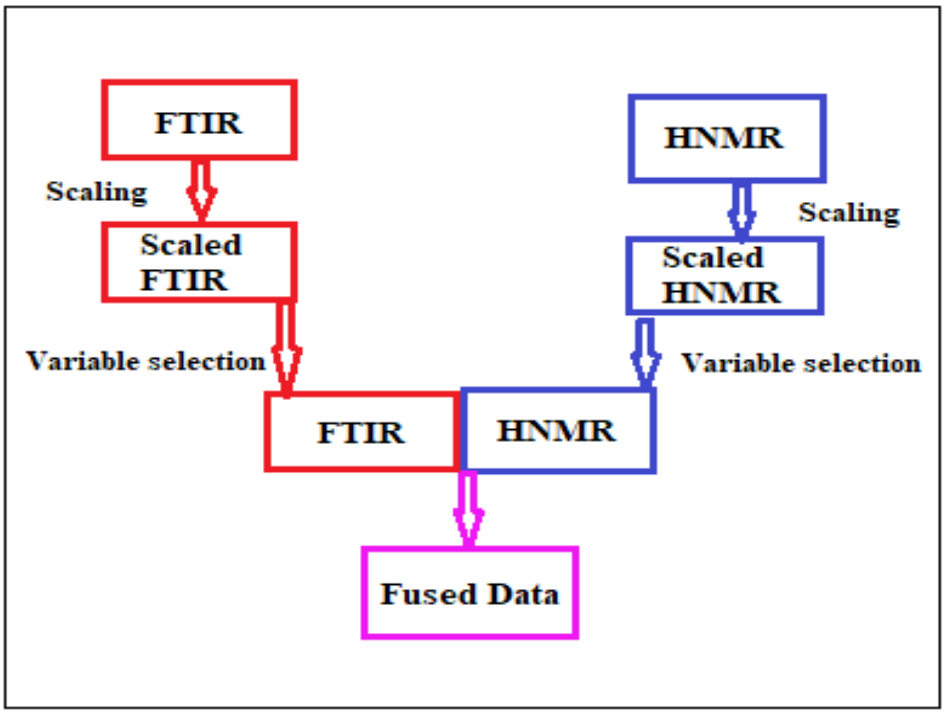
4.5. DATA COMBINATION

The usage of integrating different sources to explore a situation is very old, using sound and sight to cross the road. However, its application in digital information sources is somewhat new. In the 1980s, for the first time in signal processing, the idea of data fusion was introduced in defense applications (target identification or land to air defense). Later, its usage in different fields (medical diagnosis or equipment or process monitoring) has rapidly increased.[26]

The US Department of Defense Joint Directors of Laboratories Data Fusion Subpanel defines data fusion as a multi-level technique, which combines information and data from different sources in order to get more accurate results, and it has many applications in dealing with estimation, correlation, fault detection, monitoring and controlling the system. [27] It has also been investigated and confirmed by many researchers that combined models provide fewer errors and better outcomes compared to the individual models. [28,29] To this end, data fusion (integrating IR with NMR) has been applied and confirmed improved results in petroleum properties. [15]

In a common data fusion classification, there are two kinds of fusion: a fusion of data types and fusion of data relations. The former method, which is more practical for dissertations or journal articles, takes the advantages of merging different data types into the same analysis entity. Contrary, in the latter method, the focus is on identifying and characterizing the relationships between individual objects by merging these multiple data relations into a new unit. [14] Morris et al. mentioned that to fully understand a research field or process, there is a need to observe and analyze the relationships from different angles since each individual relationship provides only a fraction of the whole picture. [30] Based on this classification, this study used the integration of multiple relationships to contribute to a comprehensive understanding of the PM decomposition.

Based on the abstraction levels in the data, this research applied a low-level fusion method [31] by importing the raw data sets which had been already provided by FTIR and ^{1}H NMR into the fusion algorithm. At the lowest level, preprocessing was accomplished by reducing the amount of data by scaling and normalizing them to the maximum peak intensity while retaining the useful information. Then, newly developed data points were connected to form the final data set, which had a lower signal to noise ratio implying that data is more accurate compared to individual data sets.



ss

Figure 4-4. Proposed steps for data combination.[32]

4.6. BAYESIAN HIERARCHICAL CLUSTERING (BHC)

Clustering is one of the most important and applied techniques in data mining. In many clustering algorithms, Euclidean distance is considered as the basic determination for similarity between objects (samples) in a group; hence, the probability of an object belonging to a specific group is either one or zero. This approach brings up issues while dealing with objects made near the border of a group. On the other hand, BHC, based on Bayesian probabilistic principles, offers significant advantages, such as higher accuracy in clustering samples that are close to the borders of clusters. [33] The Bayesian clustering approach was applied in this study for two reasons. First, in pattern classification, most of the popular classifiers are based on maximizing the posterior probability which is the core principle of the Bayesian approach.[34] Second, BHC is an appropriate method for dealing with a high dimensional data set (the IR and data fusion samples had 1764 and 3665 variables, respectively).[12]

For these reasons, this study proposed a BHC algorithm that can maximize hidden variables by marginalizing the random parameter irregularity in large data sets to improve the precision of the

results by accurately depicting the information. All steps were explained in more details in previous paper.[35]

4.7. SELF- MODELING MUTIVARIATE CURVE RESOLUTION- ALTERNANTING LEAST SQUARE (SMCR-ALS)

The second goal of this study was to resolve the issues related to interpreting the high dimensional set collected by FTIR and ^{1}H NMR, as well as develop an algorithm for online monitoring and controlling of a multi-mixture process such as biomass conversion in the real-time analysis. To achieve these goals, SMCR-ALS technique was applied which is one of the most popular techniques which allows mathematical resolution of the mixture into spectral profiles and concentrations of pseudo-components in the mixture.[36, 37] Differently put, the aim of this step was to develop individual factors for single species in a multi-component mixture by developing a mathematical data decomposition procedure which requires only some general information related to the variables (non-negativity of spectral profiles and concentrations).[20] This method takes the advantages of requiring less costs, chemical efforts and time, [38] and as its terminology (self-modeling) shows, it does not need prior information about samples or connecting the data to resolved pseudo-components.[18] Moreover, this study develops an algorithm for online monitoring and controlling of a multi-mixture process such as biomass conversion within a decomposition framework consistent with the Beer's law that adds physical significance to the interpretation of pseudo-component spectra and concentrations, all without any prior knowledge.

All the steps, equations and procedures were explained in detail in the previous papers.[35] The core idea behind SMCR bilinear model is that the data matrix D can be obtained from the following multivariate measurements:

$$D = D_{SMCR} + E_{SMCR} = CS^{T} + E_{SMCR} \quad (4\text{-}1)$$

Where the reconstructed data matrices, row vector, column vector and the residual matrix are presented by D_{SMCR}, C, S^{T}, and E_{SMCR}, respectively.

4.8. THE BAYESIAN NETWORK (BN)

Recently, data sets with large variables and only a few samples (same as data sets obtained in this study) have become very common, consequently, many researchers believe that the Bayesian approach is the best choice to develop a reaction network in these cases. [39, 13] BNs are graphical models including nodes or vertices connected by directed edges or arcs in an acyclic manner. Hence, BNs are based on a set of random variables $X = \{X_1, \ldots, X_N\}$ which each X presents a node of a directed acyclic graph (DAG) and arcs present direct dependency between these variables (X_i).

To develop a BN, two steps are required: structure and parameter learning. To learn these steps from a data set, the space(s) of all possible networks are searched by algorithms that make locally optimal choices based on maximizing the likelihood or scoring function (Equation 4-2). The graph presenting the maximum score will be chosen as the most likely network to present the data set. In other words, structure learning includes determining the DAG that presents the dependency relationship structure of the data and parameter learning involves in estimating the parameters of the global distribution obtained from structure learning.[40] To have an optimal BN, this study applied three different learning approaches- tabu, Hill climbing (HC) and Maximum-minimum hill climbing (MMHC).

The following Bayesian information criterion (BIC) was applied for scoring function in which the dimensions of the graph (d), number of samples by (n) and finally, last term was added to reduce the effect of overfitting:

$$BIC = \sum_{i=1}^{n} \text{logPxi (Xi } |\Pi \text{ xi)} - \frac{d}{2} \log(n) \qquad (4\text{-}2)$$

4.9. RESULTS AND DICSUSSION

4.9.1. FTIR, ^{1}H NMR and Final Fused Spectra

Bio-oil obtained from HTL of the physical mixture of LG and PEB was characterized for various functional groups using a simple, fast and reliable method such as FTIR.[41] In total, twenty-seven processed samples were analysed for the physical mixture, and the handbook of spectroscopic data was used to identify the functional groups.[42]

The functional groups and related classification of compounds are listed in Table 4-1. The oxygenated compounds (O−H group) are presented by absorption peaks at 3200–3600 cm^{-1}. Furthermore, methyl (alkanes) and methylene (alkenes) groups are indicated by peaks at 2800 and 3000 cm^{-1} (C−H stretching vibrations); and 1350 and 1475 cm^{-1} (C−H deformation vibrations), respectively. [43] Alkene also is presented by the peak 1656 cm^{-1} (C=C stretching vibrations). The presence of ketones, carboxylic acid or aldehydes groups is confirmed by absorbance at 1700 cm^{-1} (C= O deformation vibrations). Moreover, the acid compounds are represented in the range of 1210-1320 cm^{-1} (C−O stretching vibrations). In addition, aromatic compounds in bio-oil are observed by the peaks appearing in the range of 1475–1525 cm^{-1} (C=C ring stretching) and 700 and 900 cm^{-1}(C-H deformation). Finally alcohols, phenols, ethers and esters are indicated by the peaks between 1000 and 1200 cm^{-1} (C−O stretching vibrations).[44]

Table 4-1: FTIR bands and functional groups of bio-oil obtained from HTL.

Frequency range, cm-1	**Group**	**Class of compound**
3600-3200	O-H stretching	Polymeric O-H
3050-2800	C-H stretching	Alkanes
1750-1690	C=O stretching	Ketones, aldehydes, Carboxylic acids
1680-1570	C=C stretching	Alkenes
1525-1475	C=C ring stretching	Aromatics
1475-1330	C-H deformation	Alkanes
1280-1200	C-H stretching	Aromatics
1200-1000	C-O stretching	Alcohols, esters, ethers
885	C-H deformation	Aromatics
750	Adjacent C-H deformation	Aromatics
696	Out of plan =CH deformation	Alkenes
615	Out of plan O-H deformation	Polymeric O-H

Another analysis was carried out using ^{1}H NMR to have a clearer understanding of the distribution of species in the whole bio-oil. Consequently, the obtained ^{1}H NMR spectra provided complementary functional group information to the FTIR spectrum. Figure 4-4 illustrates the spectra confirming the presence of aliphatic protons, aromatic, olefinic, alcohols and aldehyde groups for the bio-oil samples in the presence of SCW and NaOH across nine different conditions. Aliphatic protons attached to at least two carbon atoms ($-CH_3$, $-CH_2-$) are presented in the first region (0.5-1.5 ppm). The presence of compounds containing aromatic or

olefinic functional groups is confirmed by the most up-field region from 1.5 to 3.0 ppm (aliphatic carbon attached to a C=C double bond). In addition, the peaks in the next region (3.0-4.5 ppm) correspond to protons on carbon atoms next to an aliphatic alcohol (-CH_2O-).Furthermore, the presence of phenolic and aromatic groups can be confirmed by peaks in the next regions (5.0-7.0 ppm) and (7.0 8.0ppm), respectively Lastly, the spectrum in the last area (9.0-10.0 ppm) resembles protons associated with aldehydes (-CHO). Overall, the presence of these typical functional groups in the whole region (0.5–9.5 ppm) had been already confirmed by FTIR results.

After producing and analyzing individual data sets obtained from FTIR and ^{1}H NMR, the next step was accomplished by using a low-level fusion procedure (integrating data points, spectra and curves) to produce a more reliable and comprehensive data set than produced by each individual spectrometer. For data fusion procedures, usually there is a need to down-sample the NMR spectra from 600 MHz to 60 MHz to provide a better representative of online process NMR data, allowing a fair comparison with the FTIR data. However, this study did not need this step since an NMReady at the frequency of 60 MHz was used.

There is another requirement for data fusion that is normalization of signals with different intensities produced by FTIR and ^{1}H NMR. To do this and develop a better optimal model, different techniques such as normalizing to unit length and maximum peak intensity exist; however, the latter was chosen in this study. After normalization to the maximum peak intensity, the final fused signal (Figure 4-5) was formed by concatenating the normalized FTIR and ^{1}H NMR spectra. In the final spectrum, the first portion of the spectrum (variables from 1 to 1764) is the IR spectrum, and the next portion of the spectrum (variables from 1765 to 3665) is the NMR spectrum. Since the final spectrum includes a large number of variables (3665 variables), a standard multivariate processing method called principal component analysis (PCA) was used in order to systematically reduce to a smaller but more meaningful set of variables, [45] and also to allocate variables in an order based on their contribution to the total variance of the data set. After applying PCA, 93 % of all information was captured and explained just by the first two principal components. All these steps were performed in MATLAB version 2018b and R version 3.5.1 and the final fused data sets were clustered using BHC (Figure 4-6).

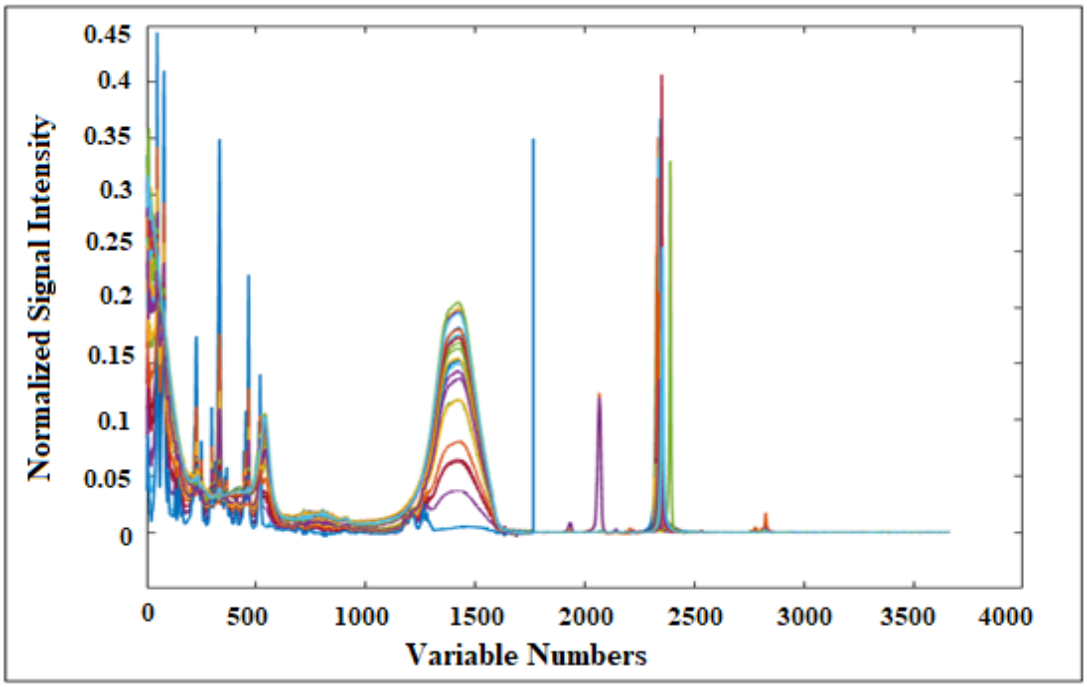


Figure 4-5. The final fused spectra for the products of PM conversion.

4.9.2. BHC and the BNs

In the BHC approach, there is a variety of distance criteria (based on types of variables used in clustering) for the determination of similarity between entities of a group. In this study, the functional groups and their related wavenumbers have been considered as a factor for conducting the hierarchical clustering. In addition, to achieve the aim of merging the similar clusters without overfitting, the probability of all the data in a potential merger was calculated and the results were compared to the lower levels of the dendrogram. This step was accomplished by using R version 3.5.1.

The literature indicates that the main products in the pyrolysis of cellulose and lignin are mostly: acids, aldehydes, ketones, gases, phenols and alcohols. [4,46] Therefore, in this study the number of clusters were chosen between 3 and 6. After developing the three different Bayesian learning approaches, networks with 3 and 4 clusters were the only ones, which had optimal solutions (Figures 4-(7 & 8)). This is why, in this paper, the focus will be on these BNs. In the following dendrograms, the total number of variables or wavenumbers have been grouped into four clusters.

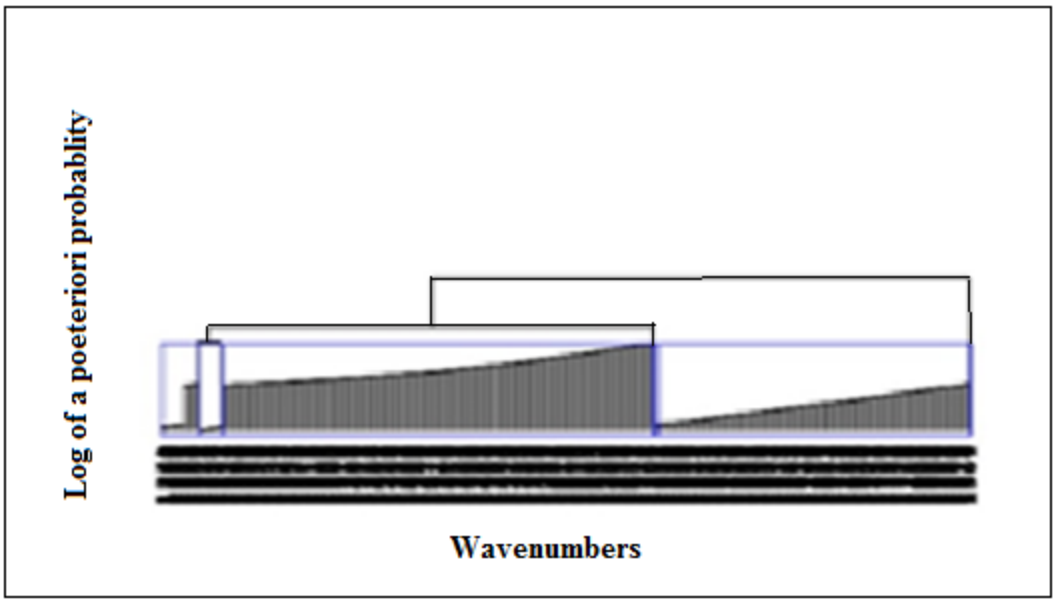


Figure 4-6. Dendrogram obtained by applying BHC cluster analysis results (4-cluster).

Before interpreting the developed BNs, three important points need to be highlighted. First, in thermal decomposition of the binary mixture of LG and PEB, even though each object goes under a separate mechanism, their interactions promote each other's reaction.[47] Next, it is already confirmed by other researchers that in a mixture of cellulose and lignin, the pyrolytic products of cellulose (based on similarity in the functional group and molecular weight) can be divided into three classes (except water, gases and char): (i) dehydrated sugars (5-6 carbons), (ii) furan derivatives (4-6 carbon numbers) and (iii) low molecular weight products (1-3 carbons). [4] These results are consistent with the pyrolysis mechanism which will be proposed later in the discussion section in which a competitive glycosidic bond and C−C bond breaking are postulated to be the primary reactions for the thermal conversion of LG. Finally, most of pyrolytic phenols products of lignin are formed through the cleavage of hydroxyl and methoxyl groups on aromatic rings and fractionation of different linkages (ß-O-4, a-O-4, ß-5) of basic lignin units. Figure 4-7 illustrates the reaction networks for the physical mixture of LG and PEB using data provided by data combination applying the tabu, HC and MMHC BN approaches.

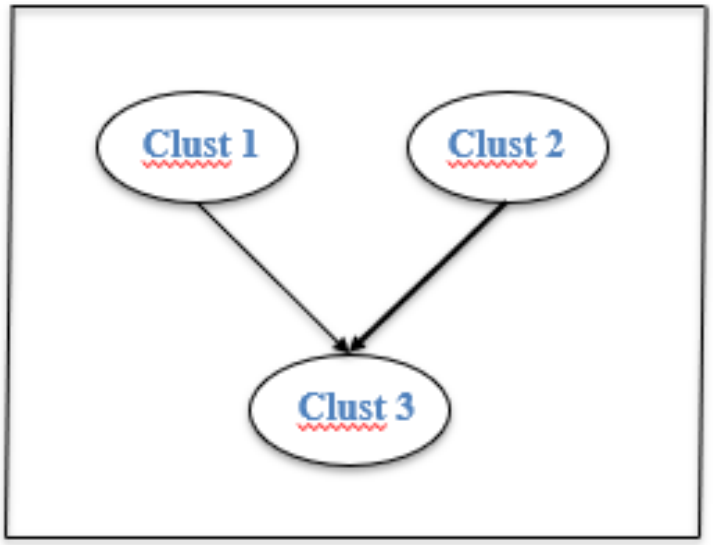


Figure 4-7. Three- cluster Bayesian network (data from data fusion method) for HTL of PM.

It can be seen from these networks that considering three clusters, all three different search algorithms (tabu, hill climbing and max-min hill climbing) provided the same results strengthening the confidence in the validity of the reaction network. While cluster 1 (compounds containing mostly phenolic groups) can be the result of hydrogen abstraction by the phenoxy radicals from hemolysis of PEB, cluster 2 consists mostly of compounds with alcohol and ether functional groups. Cluster 3 includes many organic compounds such as ketones, aldehydes, aromatics, alkynes, alkanes and alkenes. These compounds are the result of the cleavage of C-O-C and C-O bonds and can be identified by absorbance peaks between 1900 and 1000 cm^{-1}. Therefore, we can argue that by moving from cluster 1 to 3, the number of compounds with carbonyl and aliphatic groups have increased.

To present the dependency and probable pathways between the clusters in a network, the strength value of each arc is calculated and reported in Table 4-2 (the strongest dependency belongs to the highest negative number). These values indicate the strength of the causality between the cause and effect confirming in which extent a product can be influenced by the change in the reactant. These numbers take the advantages of being automatically updated when new information comes in and presents the change in the overall score of the network. Therefore, according to data provided by this table, the arc pointing out from cluster 2 to the cluster 3 shows

the highest probability for updating itself (43.99) which means removing it, results in the overall score of the network to decrease automatically by 43.99.

Table 4-2. Arc strength calculated by tabu, HC, and MMHC search methods.

From	To	Arc strength
Cluster 1	Cluster 3	-7.47
Cluster 2	Cluster 3	-43.99

After constructing the directed causal map between clusters (DAG structure) and calculating the strength value for each arc, the next step was parameter learning which involved quantifying the uncertainty about the model (Θ). This was done by using the Markovian property in which the conditional probability of each child is only on its parents considering that in a single path or arc, a child is a descendant of its parent in the sequence of the ordered nodes.[48] The conditional probability distribution of a group was used to compute a model for the mean value of the intensity (root mean square value of the absorbance intensities, RMS) of that group.

$$P(X_1) \sim N(\mu_1, 0.0023^2) \quad (4\text{-}3)$$
$$P(X_2) \sim N(\mu_2, 0.0315^2) \quad (4\text{-}4)$$
$$P(X_3 | X_1, X_2) \sim N(\mu_3, 0.0108^2) \quad (4\text{-}5)$$
$$\mu_1 = 0.025 \quad (4\text{-}6)$$
$$\mu_2 = 0.215 \quad (4\text{-}7)$$
$$\mu_3 = -0.087 - 3.29\,\mu_1 + 1.28\mu_2 \quad (4\text{-}8)$$

Equations 4(3-5) depict the conditional probability distribution of each group, and equations 4(6-8) show the mean intensity (X_i is the intensity value of the i_{th} variable (i= 1,2,3) and μ_i is the mean value of X_i). The application of the last three equations 4(7-9) reveals to what extent the mean value of the probability distribution of each cluster is related to others in the network. This can be viewed as a pseudokinetic model and can be valuable and beneficial to monitor an online process in real time analysis by applying appropriate controls over process variables.

In addition to a reaction network that included three clusters, all three different search algorithms provided the same results while considering four clusters (Figure 4-8). As mentioned earlier, the interactions of cellulose and lignin are very intensive and at low temperature, they promote the conversion each other's pyrolytic products.[49] By grouping the wavenumbers into four clusters, and comparing it with three-cluster BN, the initial and final clusters include the same functional groups, so the only difference is in cluster 3. This cluster mostly includes wavenumbers related

to the vibrations of the benzene and aromatic skeleton bearing methoxy or ethoxy substituents observed in the FTIR spectra following the pyrolysis of PM as multiple peaks at 2760–3150 cm^{-1} and 690–1630 cm^{-1}. These are due to such compounds as 1,2,4-trimethoxybenzene, eugenol, and 1,2-diethoxybenzene, which are considered as phenol alkylation[47] or rearrangement or dehydration of LG under suitable experimental conditions.[50]

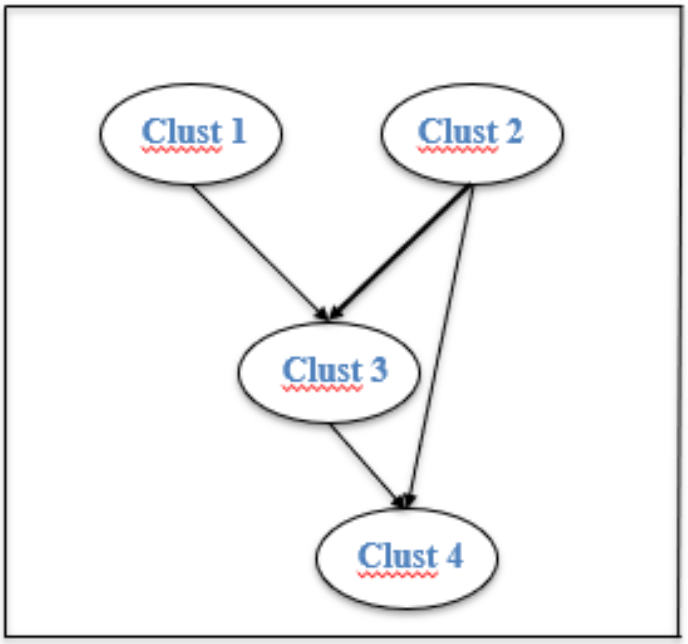


Figure 4-8. Four- cluster Bayesian network for PM conversion (data provided by data fusion).

According to data provided by Table4- 3 (in 4-cluster BN), the arc from cluster 2 to 3 elucidates the highest dependency or the most probable pathway showing the conversion of compounds containing ether and alcohol functional groups (such as LG) into aromatic groups (-35.36). This finding is consistent with the literature. [50]

Table 4-3. Arc strength calculated by tabu, HC, and MMHC search methods

From	To	Arc strength
Cluster 1	Cluster 3	-4.08
Cluster2	Cluster 3	-35.66
Cluster 2	Cluster 4	-32.79
Cluster 3	Cluster 4	-10.61

Equations 4(9-12) demonstrate the conditional probability distribution of each group and equations 4(13-16), describe mean intensity for those groups.

$P(X_1) \sim N(\mu_1, 0.002^2)$ (4-9)
$P(X_2) \sim N(\mu_2, 0.053^2)$ (4-10)
$P(X_3 | X_1, X_2) \sim N(\mu_3, 0.009^2)$ (4-11)
$P(X_4 | X_2, X_3) \sim N(\mu_4, 0.003^2)$ (4-12)
$\mu_1 = 0.025$ (4-13)
$\mu_2 = 0.202$ (4-14)
$\mu_3 = 0.309 + 2.78\,\mu_1 - 0.62\,\mu_2$ (4-15)
$\mu_4 = 0.066 + 0.587\mu_2 - 0.308\,\mu_3$ (4-16)

4.9.3. SMCR-ALS & the BN

The purpose of the SMCR method is to determine the region, concentration and spectral profile of each component in a multi-mixture compound by resolving the data while having the minimum information of the system (non-negative constraints). Additionally, including the alternating least squares regression procedure (ALS) as an optimization method enables SMCR to incorporate any data-specific constraints.

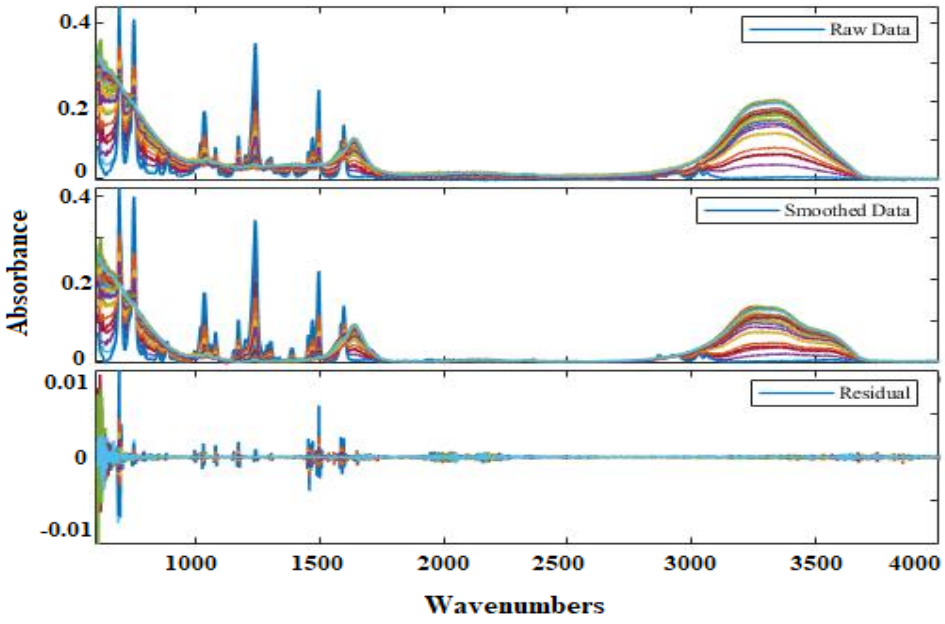


Figure 4-9. FTIR raw and smoothed spectra and removed residual for PM conversion.

To develop the SMCR-ALS algorithm, all steps and calculation methods were explained in detail in the previous papers.[35] While Figure 4-9(a) shows the FTIR results from twenty-seven samples of hydrous pyrolysis of PM in various conditions, Figure 4-9(b) illustrates the denoised and background-corrected signals for these spectra by using the Savitsky-Golay filtering algorithm (the residual noise removed is also shown in the figure). Equation (4-1) and the ratio of the

second and third derivatives (ROD) were computed in order to dissolve the whole spectra into the spectral profiles and concentrations of those active chemical species or chemical ranks involved in the samples to compute the chemical ranks, respectively.[51] Figure 4-10 illustrates that the number of species changing during the reaction reaches a maximum value of 4, meaning that the chemical rank is 4.

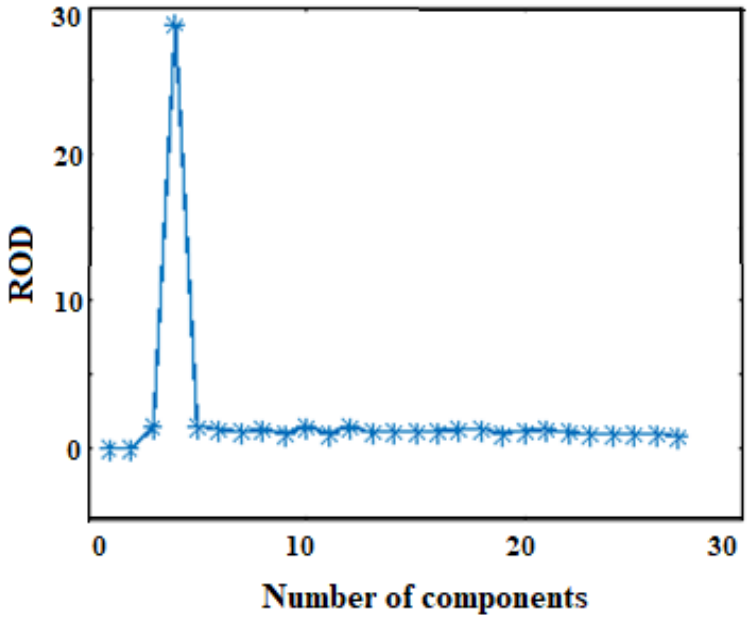


Figure 4-10. Calculating chemical rank using ROD.

Figure 4-11 shows the result of using this algorithm with the FTIR data. Figure 4-11 (a) demonstrates the overall resolved spectra for the pseudo-components (A_1, A_2, A_3, and A_4) for the whole region and Figures 4-11 (b-d) illustrate the regions containing the major peaks of these pseudocomponents. It can be seen that these pseudocomponents have distinct spectra in the desired wavelength regions and they contain spectral structures that can represent the actual molecules.

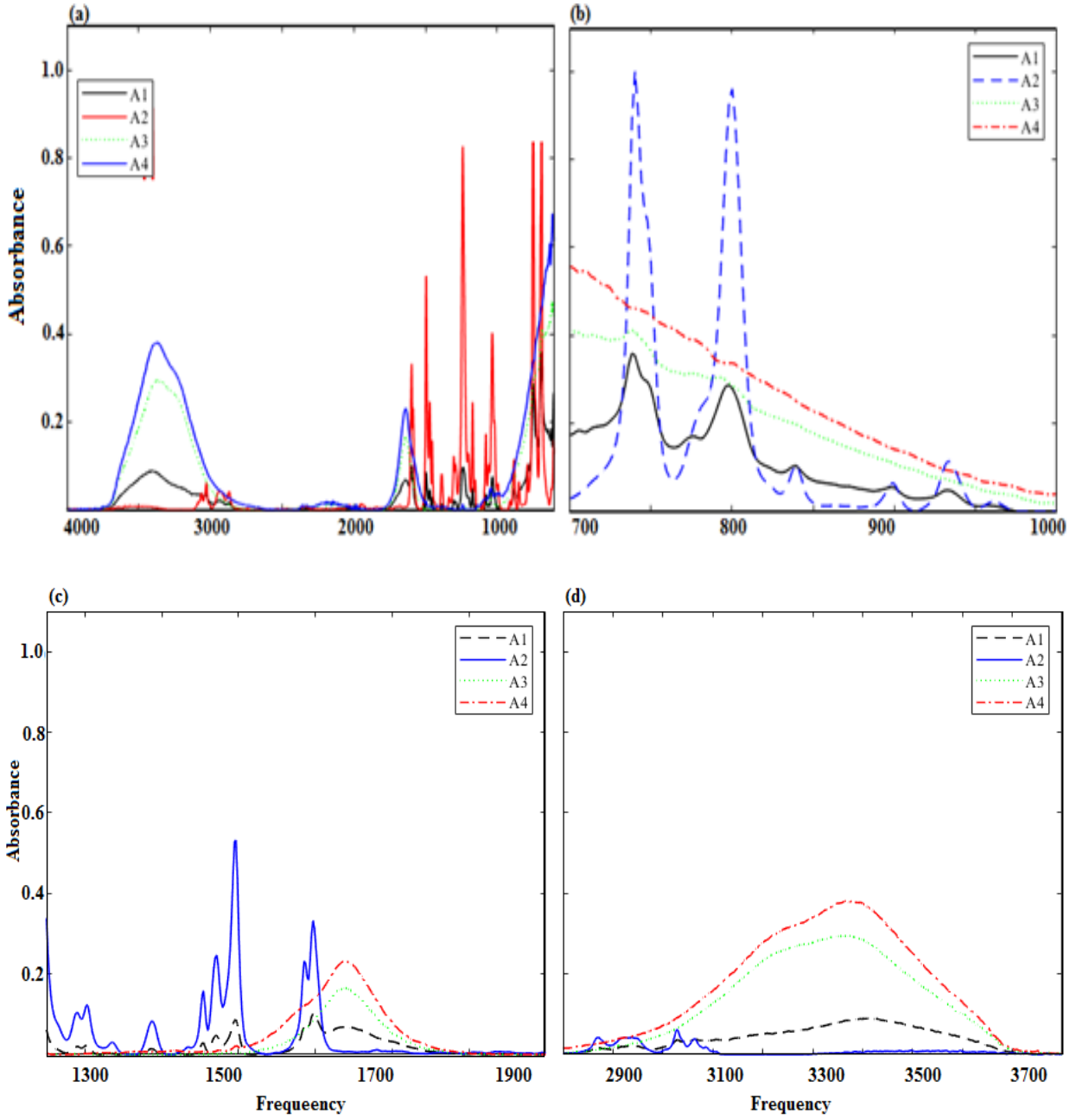


Figure 4-11. (a) Resolved spectra for the pseudocomponents over the whole region; and (b, c, and d) the resolved spectra for the pseudocomponents by focusing on the major peaks.

To develop the reaction network (Figure 4-12), the concentrations of A_1, A_2, A_3, and A_4 were estimated by applying EFA (evolving factor analysis) and later used as input data for the BNs approach. The one explanation for this BN could be that the first cluster (A_1) represents pseudo-components mostly containing phenolic groups (the absorbance of peaks between 3676 and 3584 cm^{-1} or peak at 1200 cm^{-1}). Cluster 2 (A_2) can be identified as aromatic compounds by C=C

vibrations with absorbance peaks from 1625 to1575 cm^{-1} or C-C in-ring stretching at 1500-1400 cm^{-1}. The third cluster (A_3) indicates pseudocomponents with alcohol functional groups (3676-3584 cm^{-1}) or ethers (3150-3050 cm^{-1}). Cluster 4 (A_4) represents compounds with carbonyl groups which can be produced by the cleavage of C-O-H and C=C bonds from clusters 1 and 3. The fourth group of pseudo components (A_4) indicates the presence of carboxylic acids (3550-3500 cm^{-1}), ketones (3550-3205 cm^{-1}), and aryl aldehydes (1715-1695 cm^{-1}) as well as pseudocomponents with alkene groups (1650-1585 cm^{-1}, and 3100-3000 cm^{-1}) which can be mostly produced by a ring- opening from cluster 2 or a bond cleavage from cluster 3. After calculating the strength value for each arc, the strongest dependency belonged to the arc pointing out from cluster 2 to the cluster 4 shows the highest probability for updating itself (-73.69) implying that its removal causes the overall score of the network to decrease automatically by 73.69.

Table 4-4. Arc strength calculated by tabu, HC, and MMHC search methods

From	To	Arc strength
Cluster 1	Cluster 2	-35,21
Cluster 1	Cluster 4	-15.46
Cluster 2	Cluster 4	-73.69
Cluster 3	Cluster 4	-20.61

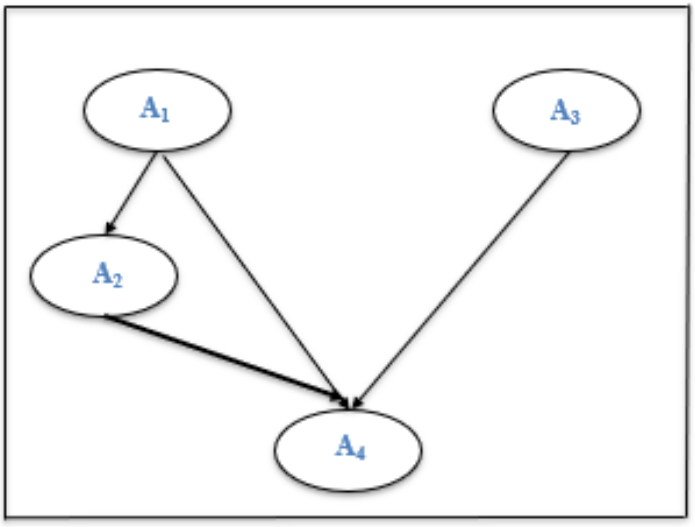


Figure 4-12. BN obtained from SMCR-ALS for the pseudocomponents.

4.9.4. Discussion

This study employed a treated physical mixture of LG and PEB to present the real cellulose and lignin in biomass considering that pyrolytic products of un-treated and treated cellulose and lignin mixture are almost the same.[52] It has been already confirmed by the previous study[32] that using data fusion provides the same BNs for both LG and PEB conversion (Figure 4-13(a & b)). According to Figure 4-13(a), wave numbers collected in the first and second clusters representing LG and glucose can be mostly identified by alcohol and ether functional groups. Additionally, in the pyrolysis of PEB, clusters 1 and 2 represent PEB itself and compounds containing mainly phenolic groups resulting from hydrogen abstraction by the phenoxy radicals in the hemolysis of PEB. However, in both cases, the final clusters include many organic compounds, such as aromatics, alcohols, aldehydes, ketones and acids produced by the cleavage of C-O-C and C-O bonds.

The 3-cluster BN for the physical mixture of LG and PEB presents some very interesting findings (Figure 4-13 (c)). First, the pathway from cluster 1 to 3 presents the conversion of PEB and phenolic groups into the final products with the almost same arc value (-7.47). Second, the arc pointing out from cluster 2 to 3 shows the conversion of LG and glucose to the final products, again, presenting the same dependency strength (-43.99). More interestingly, for both LG and PEB conversion, many wavenumbers grouped in clusters 3 can be traced in the third cluster for the conversion of the physical mixture which were identified by absorbance peaks between 1900 and 1000 cm^{-1}. These findings are consistent with the other researchers claiming that lignin would promote the decomposition of cellulose to furans and low molecular weight products and the cellulose would also promote the decomposition of lignin to increase some phenolic compounds.[49, 53, 54]

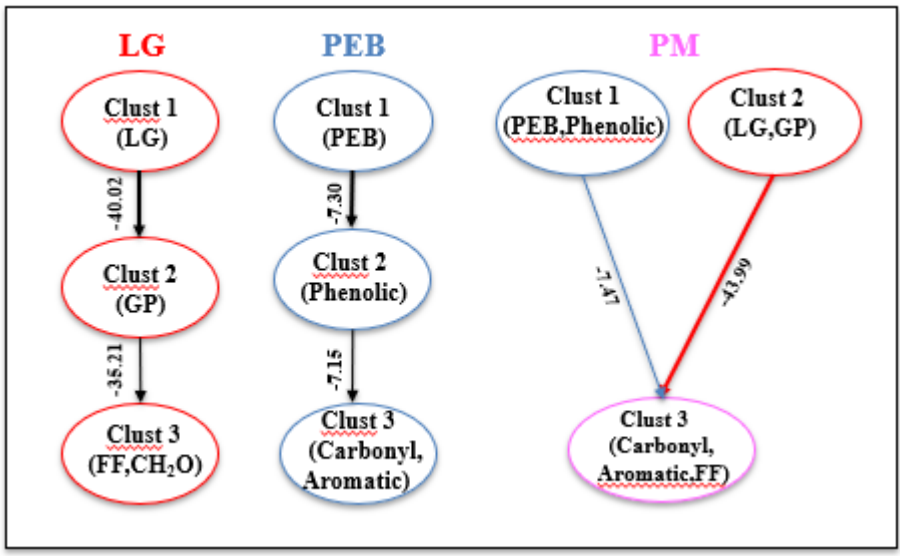


Figure 4-13. Three-cluster BNs for LG, PEB, and PM (physical mixture) based on data fusion.

Another interesting aspect was observed when the BNs for the physical mixture (PM) resulting from SMCR-ALS analysis was compared with the individual results from LG and PEB (Figure 4-14). It can be seen that most of the wave numbers found in clusters 1 and 2 in LG conversion (Figure 4-14(a)) showing the pseudocomponents including hydroxyl and ether groups can be found in the third pseudocomponent (A_3) in the conversion of PM. At the same time, A_1 and A_2 from both PEB and PM consist of pseudocomponents containing the same functional groups (phenolic and aromatic groups). It is also interesting that cluster 4 (A_4) in PM conversion (Figure4- 13(c)) includes all final products of LG and PEB conversion. As a result, we concluded that BNs resulted from BHC and SMCR-ALS of PM conversion could not only be a proper network for their conversion but also an appropriate representation for each individual element of the mixture.

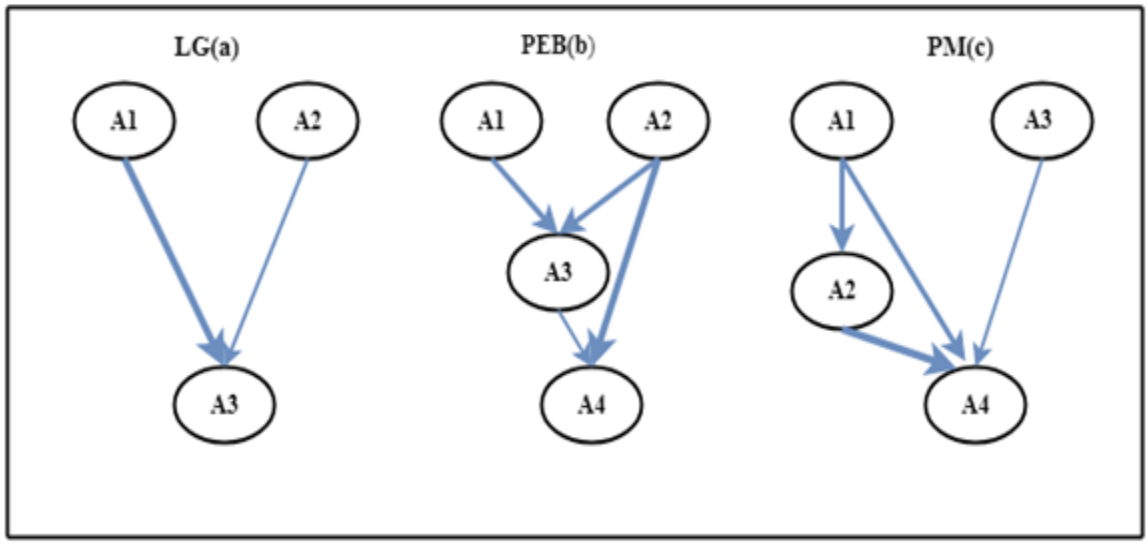


Figure 4-14. BNs from SMCR-ALS results for (a) LG, (b) PEB, and (c) PM conversion.

In addition, in order to develop an algorithm for the online monitoring of species conversion in a complex reaction, we traced the time evolution of concentrations of A_1, A_2, A_3 and A_4 (Figure 4-15). As the reaction progressed over time, we observed an increase in the concentration of A_4, with a simultaneous decrease in the concentrations of A_1, A_2, and A_3. By integrating this information with a suitable control strategy to adjust process conditions for yield maximization of the desired product, we obtained the algorithm for online monitoring of a complex reaction, such as PM conversion.

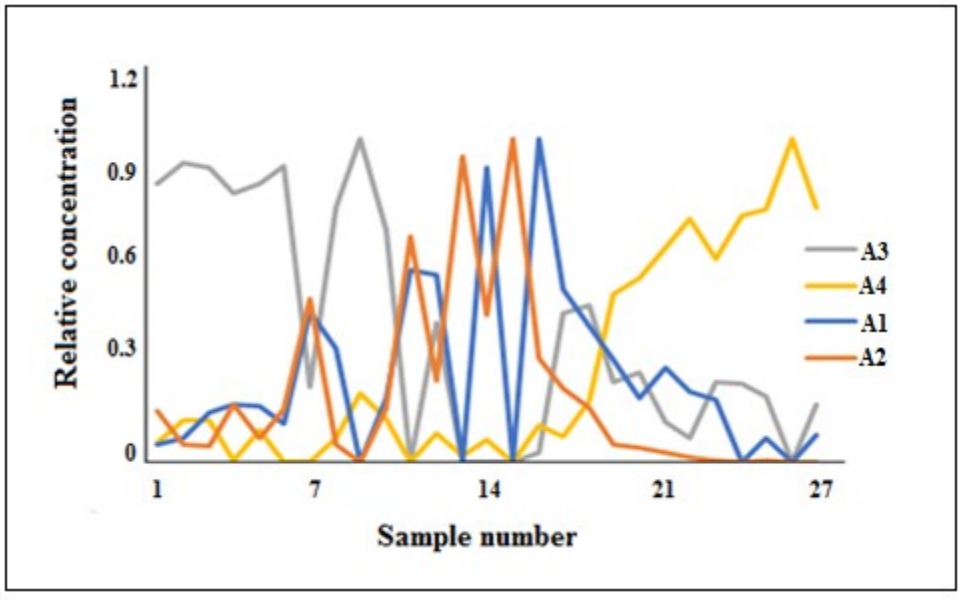


Figure 4-15. Corresponding concentration for A_1, A_2, A_3, and A_4.

After developing the reaction networks, the following mechanism was proposed for hydrous pyrolysis of PM (Figure 4-16). According to this pathway, the pyrolytic products of D-glucose (LG dehydration's result) can be categorized into furans, anhydrosugars and linear carbonyls by C–C bond cleavage, isomerization reaction, ring opening reaction, dehydration reaction, keto-enol tautomerization reaction and cyclization (ring-formation) reaction.[55] As a result, furfural (FF) and formaldehyde (CH_2O) could be the result of the dehydroxymethylation reaction of the side chain of the furan-ring and 5-hydroxymethylfurfural (5-HMF) is generated through the dehydration of the hydroxyl groups (C-2).

On the other side, C_β-O homolysis is responsible for initial decomposition during the pyrolysis of PEB which resulted in forming radicals. Different phenolic products such as 4-hydroxymethyl-2-methoxyphenol, 2-hydroxy-benzaldehyde and 2-methoxyphenol could be the results of conversion of these radicals. [56] Additionally, volatile liquids such as aldehydes (acetaldehyde) and ketones (acetone), could be considered as the other major products. They are produced because of the hydrogen-donor/radical balance at low temperature (which was used in this study).

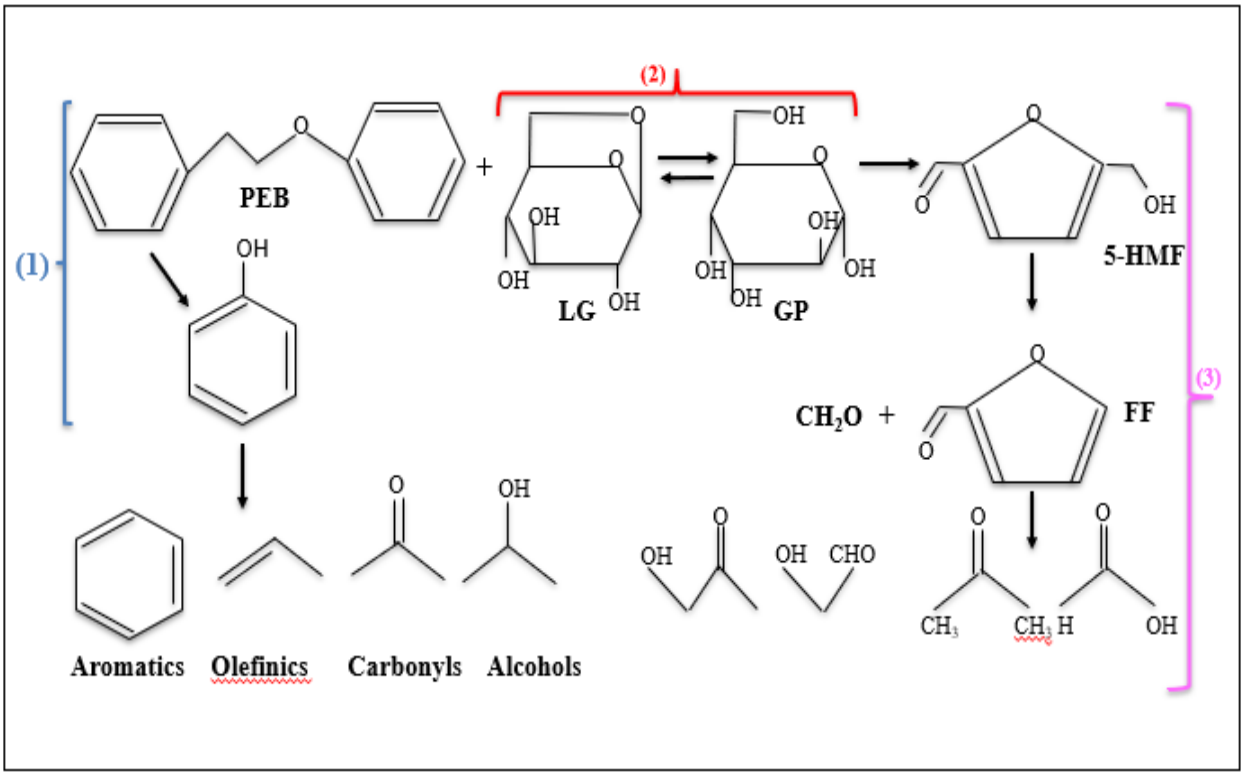


Figure 4-16. Proposed mechanism for PM conversion.

To have a complete argument considering the PM pyrolysis interaction mechanism, we propose that this mechanism can be recognized as the reactions of H-donors (LG derived pyrolytic products) and H-acceptors (lignin-derived pyrolytic products). However, based on other theoretical studies, there is a possibility that the H-donation and H-abstraction reactions are not the only responsible interactions during the LG and PEB pyrolysis process. This is due to the fact that in cellulose pyrolysis process, the decomposition mechanism might happen without the formation of free radicals. [53]

4.10. CONCLUSION

To overcome the complexity of the mechanism of a multi-mixture system such as hydrous pyrolysis of biomass, the first part of this research made use of data fusion and data mining methods to develop the most probable reaction network. The principal motivation for data fusion was to improve the quality of information and deliver a more complete image of the process under observation. The BNs, which are intuitively understandable and mathematically rigorous, provide a simple and effective representation of conditional dependence between any two distinct nodes. Hence, this study manipulated a set of random variables (wavenumbers) to compute their joint probability distribution (JPD) and applied three different score-based greedy search algorithms (tabu, hill climbing, and max-min hill climbing) to develop the most probable reaction networks for hydrous pyrolysis of a mixture of LG and PEB. All algorithms mapped the same BNs with three and four clusters confirming the optimal solutions. In these BNs, while initial clusters mostly contain compounds with alcohols, ethers and phenolic groups, the final cluster is a group of wavenumbers related to aromatics, aldehydes, ketones, alkynes, alkanes and alkenes. In other words, by moving from clusters representing feed to final products, the number of compounds with carbonyl groups and aliphatic groups increases.

In addition, in the second part of this study, an SMCR-ALS algorithm was developed in order to predict the reaction chemistry of converting a physical mixture of LG and PEB into products along with the ability for real-time analysis and quantitative tracking of changes in the process. In this part, it was confirmed that only four pseudocomponents took part the reaction (based on changes in their pseudoconcentrations). Therefore, the whole spectra were resolved into spectral profiles of these pseudocomponents as well as computing their concentrations that were used as

an input for causality detection. The strength of this algorithm is that it requires a smaller amount of quantitative data to be effective and does not require vast prior knowledge of the system. In other words, we believe that mathematical modeling and principles rooted in chemistry and chemical engineering could be useful and practical techniques to understand a complex reaction as well as how to influence them.

4.11. REFERENCES

1. Cherubini F, Strømman AH. Principles of biorefining. *Biofuels*. 2011:3-24. doi:10.1016/B978-0-12-385099-7.00001-2

2. Lange J. Lignocellulose conversion: an introduction to chemistry, process and economics. 2007;1(1):39–48.

3. Liu X, Zu X, Liu Y, Sun L. Conversion of waste water hyacinth into high-value chemicals by iron (III) chloride under mild conditions. *BioResources*. 2018;13(2):2293-2303. doi:10.15376/biores.13.2.2293-2303

4. Yang H, Yan R, Chen H, Lee DH, Zheng C. Characteristics of hemicellulose, cellulose and lignin pyrolysis. *Fuel*. 2007;86(12-13):1781-1788. doi:10.1016/j.fuel.2006.12.013

5. Collard FX, Blin J. A review on pyrolysis of biomass constituents: Mechanisms and composition of the products obtained from the conversion of cellulose, hemicelluloses and lignin. *Renew Sustain Energy Rev*. 2014;38:594-608. doi:10.1016/j.rser.2014.06.013

6. Wong ASY, Huck WTS. Grip on complexity in chemical reaction networks. *Beilstein J Org Chem*. 2017;13:1486-1497. doi:10.3762/bjoc.13.147

7. Ross, J.; Schreiber, I.; Vlad MO. *Determination of Complex Reaction Mechanisms: Analysis of Chemical, Biological, and Genetic Networks*. Oxford University Press: New York; 2006.

8. Faulon JL, Sault AG. Stochastic Generator of Chemical Structure. 3. Reaction Network Generation. *J Chem Inf Comput Sci*. 2001;41(4):894-908. doi:10.1021/ci000029m

9. J P. Causality: Models, Reasoning, and Inference. *Cambridge Univ Press*. 2000.

10. Ugi, I.; Bauer, J.; Bley, K.; Dengler, A.; Dietz A.; Fontain, E.; Gruber, B.; Herges R., Knauer, M.; Reitsam, K.; Stein N. Computer-Assisted Solution of Chemical Problems-The Historical Development and the Present State of the Art of a New Discipline of Chemistry. *Chem Int Ed Engl*. 1993;32:201-227.

11. Zou C, Denby KJ, Feng J. Granger causality vs. dynamic Bayesian network inference: A comparative study. *BMC Bioinformatics*. 2009;10. doi:10.1186/1471-2105-10-401

12. Tefera DT, Yañez Jaramillo LM, Ranjan R, Li C, De Klerk A, Prasad V. A Bayesian Learning Approach to Modeling Pseudoreaction Networks for Complex Reacting Systems: Application to the Mild Visbreaking of Bitumen. *Ind Eng Chem Res*. 2017;56(8):1961-1970. doi:10.1021/acs.iecr.6b04437

13. Faltin F, Kenett R. Bayesian Networks. *Encycl Stat Qual Reliab*. 2007;1(1):4. doi:10.1002/wics.48

14. Xu HY, Yue ZH, Wang C, Dong K, Pang HS, Han Z. Multi-source data fusion study in scientometrics. *Scientometrics*. 2017;111(2):773-792. doi:10.1007/s11192-017-2290-5

15. Castanedo F. A review of data fusion techniques. *Sci World J*. 2013;2013. doi:10.1155/2013/704504

16. Awa K, Okumura T, Shinzawa H, Otsuka M, Ozaki Y. Self-modeling curve resolution (SMCR) analysis of near-infrared (NIR) imaging data of pharmaceutical tablets. *Anal Chim Acta*. 2008;619(1):81-86. doi:10.1016/j.aca.2008.02.033

17. RóbertRajkó. Additional knowledge for determining and interpreting feasible band boundaries in self-modeling/multivariate curve resolution of two-component systems. *Anal Chim Acta*. 2010;661(2):129-132.

18. Shinzawa H, Jiang JH, Iwahashi M, Noda I, Ozaki Y. Self-modeling curve resolution (SMCR) by particle swarm optimization (PSO). *Anal Chim Acta*. 2007;595(1-2 SPEC. ISS.):275-281. doi:10.1016/j.aca.2006.12.004

19. Rahimdoust Mojdehi N, Sawall M, Neymeyr K, Abdollahi H. Investigating the effect of flexible constraints on the accuracy of self-modeling curve resolution methods in the presence of perturbations. *J Chemom*. 2016;30(5):252-267. doi:10.1002/cem.2787

20. W.H. Lawton EAS. Self modeling curve resolution. *Technometrics*. 1971;13(3):617-633.

21. Simoneit BRT, Schauer JJ, Nolte CG, et al. Levoglucosan, a tracer for cellulose in biomass burning and atmospheric particles. *Atmos Environ*. 1999;33(2):173-182. doi:10.1016/S1352-2310(98)00145-9

22. Schreuder LT, Hopmans EC, Stuut JBW, Sinninghe Damsté JS, Schouten S. Transport and deposition of the fire biomarker levoglucosan across the tropical North Atlantic Ocean. *Geochim Cosmochim Acta*. 2018;227:171-185. doi:10.1016/j.gca.2018.02.020

23. Kidder MK, Britt PF, Chaffee L, Buchanan AC. Confinement effects on product selectivity in the pyrolysis of phenethyl phenyl ether in mesoporous silica. 2007:52-54. doi:10.1039/b613757a

24. Nybacka L. FTIR spectroscopy of glucose FTIR spectroscopy of glucose. 2016.

25. Bonnefont C, Guerra A, Théron L, Molette C, Canlet C, Fernandez X. Metabolomic study of fatty livers in ducks : Identification by 1 H-NMR of metabolic markers associated with technological quality. 2011:1542-1552.

26. E. Waltz JL. Multisensor Data Fusion. *Artech House, Norwood MA*. 1990.

27.] L. A. Klein. Sensor and Data Fusion Concepts and Applications. *SPIE Opt Eng Press Bellingham WA,*. 1999.

28. Blackman SS. Association and fusion of multiple sensor data," in Multitarget-Multisensor: Tracking Advanced Applications. *Artech House*. 1990:187–217.

29. Luo RC, Kay MG. Multisensor Integration And Fusion: Issues And Approaches. *Proc SPIE*. 1988;0931(August 1988):42-49. doi:10.1117/12.946646

30. Morris, S.A., Van der Veer Martens B. Mapping research specialties. *Annu Rev Inf Sci Technol*. 2008;42(1):213-295.

31. Llinas DLH and J. An introduction to multisensor data fusion. *Proc IEEE*. 1997;85(1):6-23.

32. Sattari F, De Klerk A, Prasad V. Application of data combination and data mining techniques to investigate the chemistry of cellulose and lignin derivatives in hydrous pyrolysis. *Submiited to I&EC*. 2019.

33. Baghbani G, Eskandari F. Calculating the required cash in bank branches: a Bayesian-data mining approach. *Neural Comput Appl*. 2018;30(9):2831-2841. doi:10.1007/s00521-017-2888-9

34. Hammer HL, Yazidi A, Oommen BJ. "Anti-Bayesian" flat and hierarchical clustering using symmetric quantiloids. *Inf Sci (Ny)*. 2017;418-419(August 2016):495-512. doi:10.1016/j.ins.2017.08.017

35. Sattari F, Tefera D, Sivaramakrishnan K, De Klerk A, Prasad V. Postulating Pseudo-reaction Networks for the Conversion of Levoglucosan in Hydrous Pyrolysis Using Spectroscopic Data & Self-Modeling Multivariate Curve Resolution. *Submiited to I&EC*. 2019.

36. Mehdi Jalali-Heravi , Hadi Parastar HE-N. Self-modeling Curve Resolution Techniques Applied to Comparative Analysis of Volatile Components of Iranian Saffron from Different Regions. *Anal Chim Acta*. 2010;(662):143-154.

37. J Jaumot, A de Juan RT. MCR-ALS GUI 2.0: New Features and Applications. *Chemom Intell Lab*. (140):1-12.

38. Jiang J, Ozaki Y. Self -modeling curve resolution (SMCR): principals, techniques, and applications T. *Ind Eng Chem Res*. 2017;4928(October). doi:10.1081/ASR-120014359

39. Scutari M, Vitolo C, Tucker A. Learning Bayesian Networks from Big Data with Greedy Search: Computational Complexity and Efficient Implementation. 2018. http://arxiv.org/abs/1804.08137.

40. Nagarajan R, Scutari M, Lèbre S. *Bayesian Networks in R*.; 2013. doi:10.1007/978-1-4614-6446-4

41. Kadiroğlu P. FTIR spectroscopy for prediction of quality parameters and antimicrobial activity of commercial vinegars with chemometrics. *J Sci Food Agric*. 2018;98(11):4121-4127. doi:10.1002/jsfa.8929

42. Mistry B. *A Handbook of Spectroscopic Data Chemistry (UV, IR, PMR, 13CNMR and Mass Spectroscopy)*. Oxford Book Company; 2009. doi:10.1063/1.2719251

43. Uncu O and Ozen B. Prediction of various chemical parameters of olive oils with Fourier transform infrared spectroscopy. *LWT – Food Sci Technol*. 2015;63:978–984.

44. Francavilla M, Manara P, Kamaterou P, Monteleone M, Zabaniotou A. Cascade approach of red macroalgae Gracilaria gracilis sustainable valorization by extraction of phycobiliproteins and pyrolysis of residue. *Bioresour Technol*. 2014;184:305-313. doi:10.1016/j.biortech.2014.10.147

45. Dunteman BGH. *Principal Components Analysis*. SAGD publications, Inc; 1976.

46. Demirbaş A. Mechanisms of liquefaction and pyrolysis reactions of biomass. *Energy*

Convers Manag. 2000;41(6):633-646. doi:10.1016/S0196-8904(99)00130-2

47. Wang S, Guo X, Wang K, Luo Z. Influence of the interaction of components on the pyrolysis behavior of biomass. *J Anal Appl Pyrolysis*. 2011;91(1):183-189. doi:10.1016/j.jaap.2011.02.006

48. Lee SM, Abbott PA. Bayesian networks for knowledge discovery in large datasets: Basics for nurse researchers. *J Biomed Inform*. 2003;36(4-5):389-399. doi:10.1016/j.jbi.2003.09.022

49. Long Y, Zhou H, Meng A, Li Q, Zhang Y. Interactions among biomass components during co-pyrolysis in (macro)thermogravimetric analyzers. *Korean J Chem Eng*. 2016;33(9):2638-2643. doi:10.1007/s11814-016-0102-x

50. Wang Y, Li X, Mourant D, Gunawan R, Zhang S, Li C-Z. Formation of Aromatic Structures during the Pyrolysis of Bio-oil. *Energy & Fuels*. 2012;26(1):241-247. doi:10.1021/ef201155e

51. Tefera DT, Agrawal A, Yañez Jaramillo LM, De Klerk A, Prasad V. Self-Modeling Multivariate Curve Resolution Model for Online Monitoring of Bitumen Conversion Using Infrared Spectroscopy. *Ind Eng Chem Res*. 2017;56(38):10756-10769. doi:10.1021/acs.iecr.7b01849

52. Zhang J, Choi YS, Yoo CG, Kim TH, Brown RC, Shanks BH. Cellulose–Hemicellulose and Cellulose–Lignin Interactions during Fast Pyrolysis. *ACS Sustain Chem Eng*. 2015;3(2):293-301. doi:10.1021/sc500664h

53. Ye X ning, Lu Q, Jiang X yan, et al. Interaction characteristics and mechanism in the fast co-pyrolysis of cellulose and lignin model compounds: A joint experimental and theoretical study. *J Therm Anal Calorim*. 2017;130(2):975-984. doi:10.1007/s10973-017-6465-3

54. Hosoya T, Kawamoto H SS. Cellulose-hemicellulose and Cellulose-lignin Interactions in Wood Pyrolysis at Gasification Temperature. *J Anal Appl Pyrolysis*. 2007;80:118–25.

55. Paine JB, Pithawalla YB NJ. Carbohydrate pyrolysis mechanisms from isotopic labeling Part 4. The pyrolysis of Dglucose: the formation of furans. *J Anal Appl Pyrolysis*. 2008;83:37-63.

56. Beste A BA. Kinetic Analysis of the phenyl-shift reaction in b-O-4 lignin model compounds: a computational study. *J Org Chem*. 2011;76:2195–203.

5. THE APPLICATION OF DATA FUSION AND CHEMOMETRIC TECHNIQUES FOR UNDERSTANDING THE CONVERSION OF BIOMASS BY HYDROUS PYROLYSIS BASED ON SPECTROSCOPIC DATA

A version of this chapter will be submitted as; Fereshteh Sattari, Dereje Tefera, Arno de Klerk and Vinay Prasad, “The Application of Data Fusion and Chemometric Techniques for Understanding the Conversion of Biomass by Hydrous Pyrolysis Based on Spectroscopic Data”.

5.1. INTRODUCTION

Biomass produces numerous, cheap and sustainable resources when used as a chemical feedstock for generating bio-oil, and it serves as an intriguing substitute for conventional fuels. Generally, biomass is CO_2-neutral, and it is the only renewable energy source that contains carbon.[1] Biomass alteration takes place via biochemical or thermochemical processes. Pyrolysis is one of the essential processes of biomass thermal conversion; it produces bio-oil as a primary product. Bio-oil is an important source for chemicals in biorefineries, [2] and it is also used in manufacturing biological pesticides.[3] Gases produced via pyrolysis can be used to generate heat and power.[4]

HTL, also known as hydrous pyrolysis, is an adequate procedure for the transformation of wet biomasses into biofuels and value-added chemicals.[5] Water is an essential reactant in the HTL reaction. Once the state of the reaction reaches the critical point, it intensely alters most of the physical and chemical properties of water, resulting in quick homogeneous and active reactions.[6] This procedure is performed at temperatures lower than 400 °C. At these temperatures, the exclusive features of hot condensed water are used to produce a biocrude with about 10-20% oxygen.[7] This temperature is adequate to start pyrolytic mechanisms in biopolymers while the pressure generated is adequate to sustain a liquid water processing phase.[8] The product yield and the physiochemical features of the HTL are mainly affected by the various kinds of feedstock, processing states (mainly reaction time and temperature) and the presence of a catalyst.[9] HTL can be a useful conversion technique for biomass as a result of the hydrophilic nature of the biomass and the ease with which it forms water slurries of biomass particles, which are generally pumped at concentrations of about 5-35% dry solids.[8]

Biomass contains different amounts of cellulose, hemicelluloses, and lignin. Its pyrolysis generates products that are equivalent to the total sum of the individual pyrolysis of the three constituents. Thus, the chemical nature of the bio-oil is firmly linked to the ratio of the components in the biomass. Bio-oil is a combination of over 300 compounds that are produced from the depolymerization of cellulose, hemicellulose and lignin. The oxygen and water composition of bio-oil ranges from 40-50% and 25-35%, respectively.[10] Hence, it is essential to comprehend the molecular composition of the products used to upgrade the bio-oils produced. The chemical constituents of the bio-oils help us to understand its features and its stability.

Additionally, producing the reaction networks could be a way to comprehend the arrangement of the synergy between pieces of the system, and this can have a significant effect on the behavior of the whole system. These networks illustrate the architecture of a complicated system by producing the link between the empirical data and the massive toolkit of the effectively analyzed system. This information is essential to develop the practical techniques required to enhance the selectivity of the products.

In this research, Monterey pine whole biomass underwent the HTL process in the presence of water under different conditions. Samples were characterized using Fourier transform infrared (FTIR) spectroscopy and Proton nuclear magnetic resonance (^{1}H NMR) spectroscopy. Generally, these on-line techniques are influenced by both physical process parameters like temperature, pressure, flow rate and liquid level, and molecular parameters that correlate to constituent concentrations, molecular structure and the degree of reaction.[11] Subsequently, the data fusion method was used to create the final fused data (FFD) as an input data set for producing the most likely reaction networks among the active groups described by these spectrometers. Data fusion makes use of techniques from other fields of study like signal processing, statistical estimation and pattern recognition.[12] The essence of data fusion is to link the data from several sensors to carryout deductions that cannot be acquired from a single sensor or source. Input data from various sources might involve parametric data linked to the object identity; consequently, lesser detection error probability and better accuracy are obtained by using data from numerous distributed sources.[13]

A major issue in computational fields like biology or chemical processes is the use of high dimensional data sets to study the network architecture of the variables accurately.[14, 15] Functional connectivity is often illustrated in terms of statistical reliance, and it is also seen as a practical theory that controls the discovery of a functional connection without any guarantee of how that connection was made. It could also be illustrated as a dependency test between two or more time series used to reject the hypothesis of statistical independence. [16] The Granger causality technique and the Bayesian network inference technique are two major procedures frequently used to evaluate spontaneous interactions among a set of elements.[17, 18] Several studies have been conducted on the systematic and computationally intensive comparisons between the two techniques on synthetic experimental data and it was inferred that when the size

of the data is small, Bayesian network (BN) inference is more preferable than the Granger causality approach.[19] This research generated a high dimensional data set of 3665 variables with a brief data length of 27 samples; thus, the BN approach was used. A BN is an analytical form that depicts a cumulative probability distribution among a set of random variables. The tabu[20] and Hill-climbing[21] techniques and a hybrid method (Max-min Hill-climbing)[15] were used to analyze the network that maximizes the score function showing how the network fits the data optimally.

Given the complexity of the thermochemical conversion of biomass and the difficulty in characterizing the physical constituents of the products with analytical instruments, a major thrust is aimed at discovering the link between the measurements that were easily acquired and labor-intensive measurements. Thus, the next objective of this study was to discover a good link to quickly foresee thorough and accurate measurement from the easily acquired measurements.[22] This study made use of a particular decomposition method known as Self-modeling multivariate curve resolution (SMCR) to straighten the two-way signals from instrumentally undetermined multi-constituents mixtures into determinants for single species.[23] This procedure exhibits the use of a recent quantitative means for the concurrent qualitative and quantitative recovery of the reaction network. Additionally, a quantitative SMCR strategy makes use of a regression constraint throughout the Alternating least squares (ALS) techniques that were utilized to measure each component of the reaction. This quantitative SMCR strategy could be utilized for procedures that are not required to produce mixtures of known compositions as a result of the lack of isolated reference material, cohesion problems and also whenever the production of the aforementioned samples are costly and time consuming. The main aim of the SMCR-ALS techniques is to break up unsettled multi-component and multivariate measurement matrices into unmixed determinants like spectral profiles, concentration profiles, and pH profiles for specific species without preexisting knowledge of the system.[24] Some applications of the SMCR procedures are quantification of trace analyses, [25] peak purity assessments,[26] or characterization of batch reactions.[27] The essence of this research is to formulate the SMCR-ALS algorithm for the monitoring of online reactions in a multi-compound reaction using compositional control techniques and quantitative monitoring of variations in the procedure.

5.2. MATERIALS AND HTL PROCESS

The data used for this research was acquired from the experimental survey of hydrothermal decomposition of Monterey pine whole biomass that was bought from Sigma Aldrich Canada. The refined biomass specimen was manufactured by thermally decomposition. A stainless-steel micro batch reactor of 24 cm in length and 2.1cm in width was used to conduct this experiment and the solvent used was subcritical water. The procedure for this experiment has been described in the preceding papers.[28] Twenty-seven liquid samples were examined in this research. This study was performed at different temperatures and reaction time intervals. The temperature was from 150-350 °C, while the reaction time was from 15-35 minutes. The initial pressure was fixed at 0.1 MPa. Catalysts are essential for HTL because they influence the rate of reaction and the structure of HTL products. Many homogeneous and heterogeneous catalysts were analyzed by other researchers for the catalysis of biomass HTL, even though the larger part of the work was centered around homogeneous catalysts (acid, alkali, and metal salts) because they are quite affordable.[29] A typical feature of homogeneous catalysts is also that the liquid products are not affected by coking.[30] Due to the aforementioned fact, 0.05M of sulfuric acid and 1M of sodium hydroxide was used to carry out this work. The temperature and residence time were obtained from the research and we concentrated on acquiring the maximum liquid with the specified features.[31] The volume ratio of biomass to the medium was 1:10, and the end products were placed in a glass beaker and analyzed.

5.3. SPECTROSCOPIC ANALYSIS AND DATA FUSION

The molecular-level description of the distinctive nature of bio-oil was acquired using FTIR and ^{1}H NMR spectroscopy. FTIR permits the characterization of the bio-oil [32] despite the high volatile constituents present. The IR spectra were obtained using an ABB MB 3000 FTIR spectrometer. The spectra were obtained at a resolution of 2 cm^{-1} as against the normal spectra range of 4000-600 cm^{-1}. All the spectra had a numerical mean of 120 scans. The investigation was conducted with the aid of a Pike miracle attenuated-total-reflectance attachment. The spectra obtained are shown in Figure 4-1. The hand book spectroscopic data was used to label the functional groups.[33]

^{1}H NMR was also used to label all functional groups that had hydrogen and carbon atoms in the specimen.[11] All ^{1}H NMR measurements were obtained at a frequency of 60MHz and a resolution of FWHM<1.0 HZ (20 ppb). Figure 4-2 illustrates the spectra for the conversion of biomass under nine different circumstances using sodium hydroxide as catalyst. Analysis was aided by consulting handbooks of spectroscopic data. [34, 35]

It is important to acquire only the essential information from the data obtained. This could be done with the aid of data fusion. [36] The multi-sensor data integrates information from different sensors and sources so as to draw conclusions that cannot be obtained from a single sensor or source and it also supports the principle for planning, decision-making and controls outcome. Multi-sensor data fusion can be conducted at four various processing levels and this is dependent at the phase the fusion occurs. They are signal level, pixel level, feature level, and decision level.[37] This work involves acquiring signals from various spectrometers. The data fusion used at the signal level implies that signals from various sensors are integrated to produce a new signal with a good signal-to noise ratio than the original signals. All the outlined procedures for data integration were analyzed in the previous papers.[38] Furthermore, the quantity of data was reduced by scaling and normalizing them to the minimum peak intensity while retaining the useful information; after which, newly developed data points were generated to form the final data set. This data set had a lower signal to noise ratio. This implies acquiring data that are more authentic unlike when the data is obtained from individual data sets.

5.4. BAYESIAN HIERARCHICAL CLUSTERING (BHC)

Clustering is defined as the process of categorizing data points in such a way that components that display some resemblance or innately belong to same class are placed in a specific group.[39] It plays a significant function in data analysis and inference and it is very common machine learning and data mining procedures. The Bayesian hierarchical approach offers a detailed fundamental approach to these kinds of analyses and is fast gaining grounds in other disciplines like clustering regions with various growth patterns in economics,[40] signal processing applications[41] , and computational biology and genetics.[42]

The BHC algorithm is identical to the traditional agglomerative clustering because it is a one-pass, bottom-up procedure that computes all the data points in its cluster and constantly joins

pairs of clusters. The algorithm utilizes a mathematical hypothesis test instead of distance to select the clusters to join.[43] The preceding hypothesis can be illustrated in such a way that the BHC effectively enumerates the probabilities of clustering's persistent with the commonly used Dirichlet method mixture model. The posterior probability of the incorporated or joined hypothesis is the derived using Bayes' rule.[44]

5.5. SELF-MODELING MULTIVARIATE CURVE RESOLUTION (SMCR)

SMCR has been used in various works of chemical reactions or equilibrium shifts[45] and a good number of its uses are in spectroscopy. Tauler et al. analyzed the use of SMCR to IR spectra obtained during numerous process runs and used for inspecting the Fujiwara reaction.[46] Gemperline et al. investigated UV /visible batch measurement. The principle of SMCR[47] utilizes bilinear decomposition instead of chemical or physical separation procedures in order to manage the instrumentally undetermined multicomponent signals and extract the pure variables of overlaid constituents. "Self-modeling" means that the SMCR do not need any specific information relating to the data to determine the pure variables. The only hypotheses are the bilinear model for the data and some specific generic awareness about the pure variables like non-negativity and operating in a single mode.[48]

This research used a rational procedure for spectral deconvolution for SMCR methods by identifying an objective decision in which the pure variables are compliant with the non-negativity and unimodality constraints.[49] A general analysis of the objective resolution is that it leads to a series of plausible solutions whose validity depend on the interaction or relationship between the real variables.[50] Therefore, the aim was to identify an objective decision in which the pure variables are arrived at using statistical foundations. This is possible only if the statistical principles are consistent with either the chemical or physical model of the data. Conventional procedures for unique resolution consist of some methods like the evolving factor analysis (EFA).[51] The resolution of the two-way data obtained from the on-line FTIR spectroscopic monitoring of HTL of biomass was illustrated to show the capability of the SMCR in spectroscopic studies. All required steps and developed algorithm were explained in the previous work.[52]

5.6. LEARNING THE BAYESIAN NETWORK (BN) STRUCTURE

The learning technique solves the optimization issue by applying the accepted heuristic search procedures. The search space is very big and such a technique tends to invest resources analyzing contenders that are not logical. This issue becomes demanding when we handle huge data sets either in number of instances or number of attributes.[14] Therefore, this research utilized the BN procedure to tackle this issue.[53] Studying the Bayesian networks are often perceived as an optimization issue where the function of data processing is to identify the network that increases a mathematically instigated score. Mainly, there are two procedures for identifying the BN structure. The first procedure presents learning through a constrained manner based on pairwise conditional independence tests between nodes. This procedure entails the estimation features of susceptible independence among the attributes found in the data. This is normally done by applying a statistical hypothesis test. A network that illustrates the analyzed dependencies and independences was then created. The second procedure presents learning as an optimization issue. We begin by explaining a numerical propelled score that describes the strength of each possible structure to the analyzed data. What the learner needs to do is to locate a structure that increases the score. Almost all learning techniques used standard heuristic search techniques like greedy hill-climbing and simulated annealing to locate high-scoring structures. The aforementioned "generic" search procedure does not use information concerning the anticipated structure of the network to be studied. For instance, greedy hill-climbing and tabu search techniques analyzes all the possible local changes in all the steps and utilizes the one that generates the largest improvement in the score. There is another method, the Max-min hill-climbing (MMHC) algorithm, which can be classified as a hybrid method, using concepts and techniques from both approaches. All the necessary procedures were described in the previous papers.[28]

5.7. RESULTS AND DISCUSSION

5.7.1. Spectroscopy Analysis and Data Fusion

Generally, biomass is made up of cellulose, hemicelluloses and lignin which are CHO chemical compounds. From Figure 5-1, we deduced that the wavenumbers of functional groups in bio-oil ranges from 3000 to 3500 cm^{-1} and also between 1000 and 1750 cm^{-1}. This proves the presence

of C=O bonds, C=C bonds, C–O bonds, C–H bonds, and O–H bonds in bio-oil. The functional groups depict that aldehydes, ketones, acids, aromatics, phenols, ethers, aliphatic compounds and alcohols in the bio-oil which exhibited the constant results acquired from the GC-MS results supported by the previous works.[54]

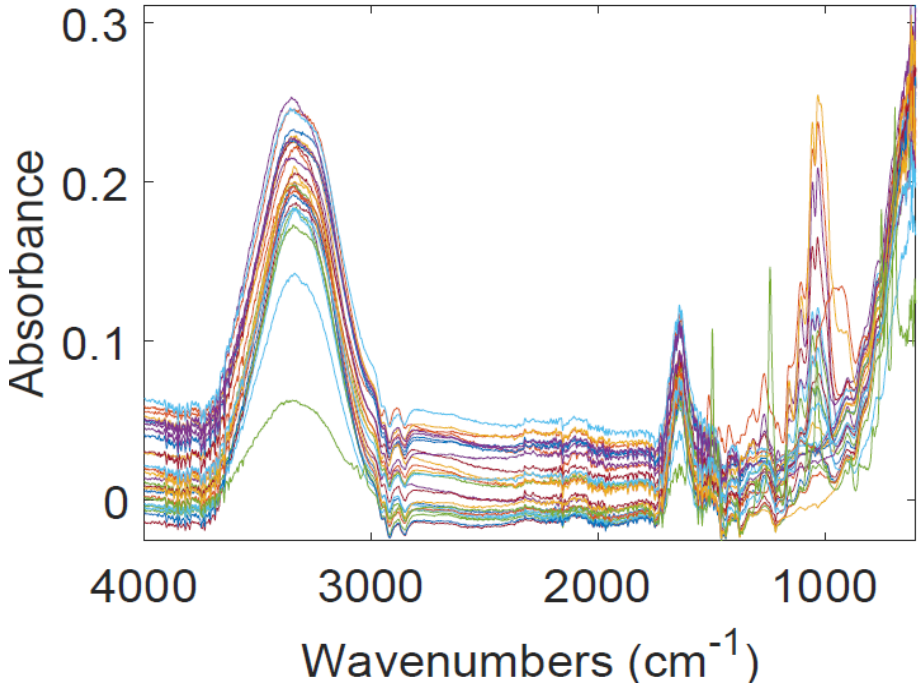


Figure 5-1. FTIR spectra biomass conversion in 27 different conditions.

Figure 5-2 depicts the ^{1}H NMR spectrum of the bio-oil derived in the presence of NaOH. The protons on the aliphatic carbons were observed between 0.5 – 1.5 ppm. The strength and percentage of proton in the subsequent phase had the highest peaks (1.5-3.0 ppm) in the ^{1}H NMR spectrum affirming the presence of aromatic or olefinic functional groups. The next phase of the spectrum had its peak at 4.0 ppm and this depicts protons of alcohols and carbon atoms next to an aliphatic alcohol (3.0-4.5 ppm). This phase has a lesser strength than that of the aromatic or olefinic carbon phase. In addition, the presence of phenolic and aromatic groups can be confirmed by peaks in the next regions (5.0-7.0 ppm) and (7.0 8.0ppm), respectively. Finally, the spectrum in the last phase (9.0-10.0 ppm) is identical to the aldehyde protons (-CHO). Hence, the presence of these functional groups in the whole phase (0.5–9.5 ppm) was proved by results obtained from FTIR.

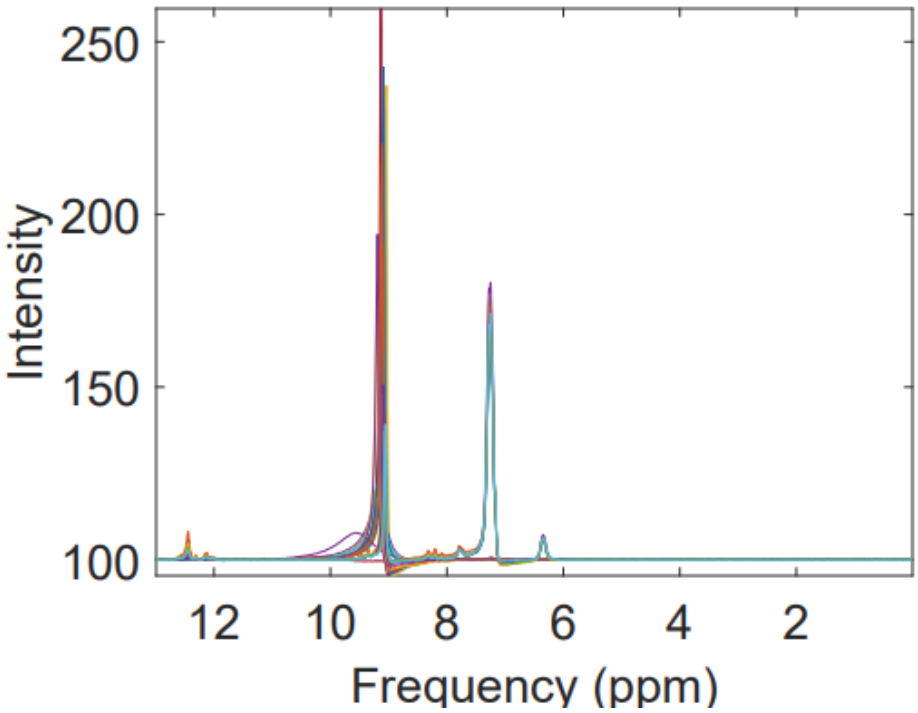


Figure 5-2. ^{1}H NMR results for biomass conversion (SCW and NaOH).

After studying the data gotten from FTIR and ^{1}H NMR, the next phase was implemented by utilizing a low-level fusion technique to create a credible and thorough data set combining information from the two sources. The primary essence of multiple sensor data fusion is to merge integral and unnecessary information to serve a composite representation that could be used to illustrate the whole scene or object recognition.

Another specification for data fusion is the normalization of signals with various intensities generated by FTIR and ^{1}HNMR. To create a better optimal model, various procedures like normalizing the unit length and the highest peak intensity could be used. The final fused signal in Figure 5-3 was produced by linking the recently acquired data points. The final spectrum consists of two main phases, the first phase is the IR spectrum and the next phase is the ^{1}H NMR spectrum. The final spectrum had a large number of variables (3665), hence a standard multivariate processing technique referred to as the principal component analysis (PCA) was applied so as to decrease it to a lesser but more significant set of variables and assign data points according to their contribution to the total variance of the data set.[55] 95% of all the variance was explained by the first two principal components. The procedures were conducted in MATLAB

version 2018b and R version 3.5.1. Figure 5-3 depicts the final data sets that were used for clustering by BHC.

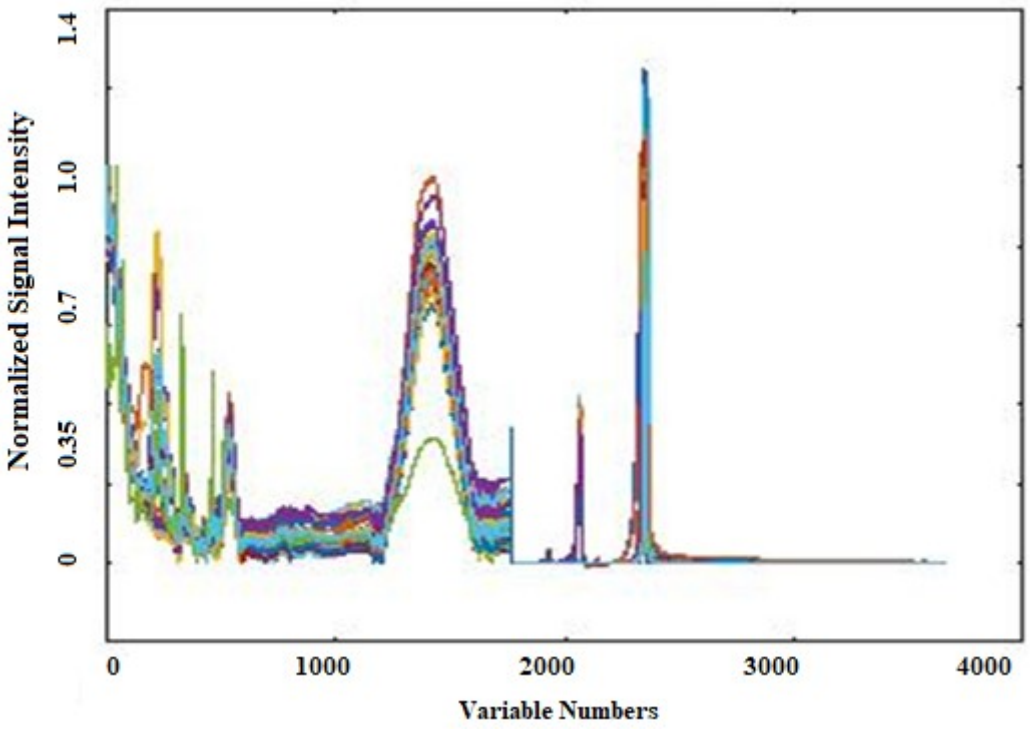


Figure 5-3. The final fused spectra for bio-oil.

5.7.2. Linear Cross Correlation Analysis

Linear cross correlation was also performed on the clusters (from SMCR-ALS result) result to better understand the variables obtained from clustering and was qualitatively and quantitatively present on the lower and upper panel, respectively (Figure 5-4). In this plot, the distribution of each variable is shown diagonally, the significance level presenting by stars and the values for correlation are shown on the top of the diagonal, and the bivariate scatter plots are displayed with a fit line at the base of the diagonal.

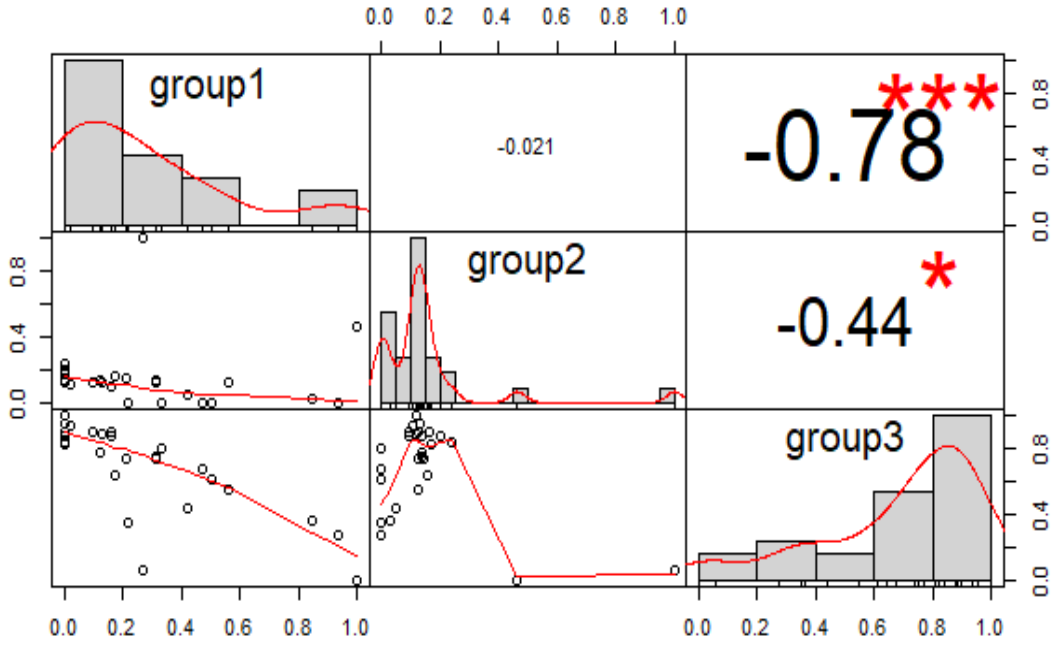


Figure 5-4. The linear cross-correlation of the groups.

The results indicate that group1 and group3 have the most correlated variables (0.78), while group2 and group3 have the least correlation (0.44). These results are consistent with findings from Table 5-2. Overall, the correlation matrix shows a smooth correlation between variables, but that alone does not indicate causality among variable, which is why Bayesian learning was chosen to assess any reaction network between variables.

5.7.3. Six-cluster BHC and Bayesian Network

This research work applied the FTIR and ^{1}H NMR interpretation for reviling functionalities of HTL bio-oil using spectra band task and analysis based on the previous studies. As stated earlier, the main constituents in the pyrolysis of biomass are alcohols, phenols, aromatics, carbonyls (acids, aldehydes and ketones), aliphatic hydrocarbons and gases.[8] Therefore, this work selected 3 to 6 clusters. After the production of the three various Bayesian learning procedures, the network that had 6 clusters had optimal solutions. This work focuses on the 6-cluster BN. In Figure 5-5, all the variables (1764 wavenumbers) are categorized into six clusters using the BHC method, and the tabu, HC and MMHC BN procedures were applied to outline the most credible network between these clusters (Figure 5-6).

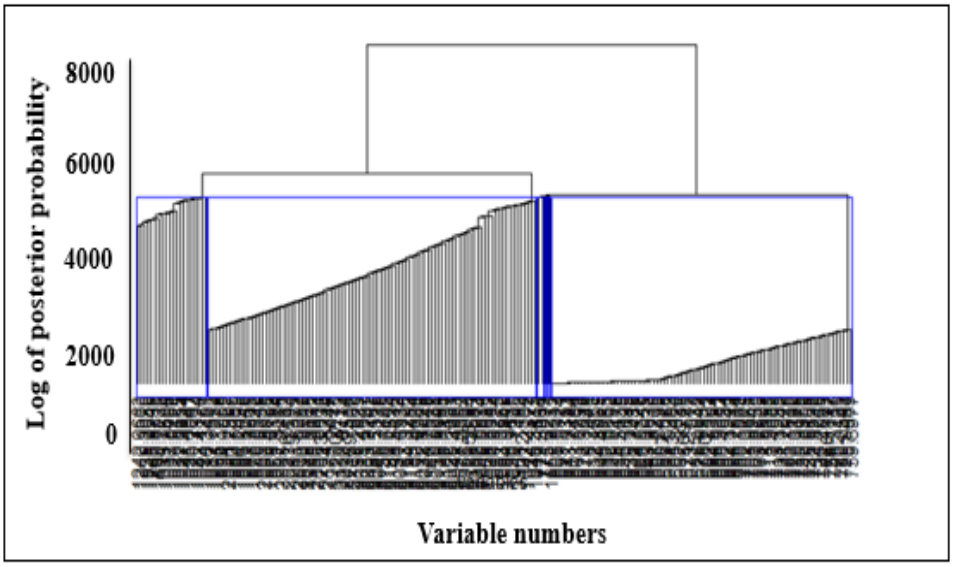


Figure 5-5. Six-cluster dendrogram for biomass conversion.

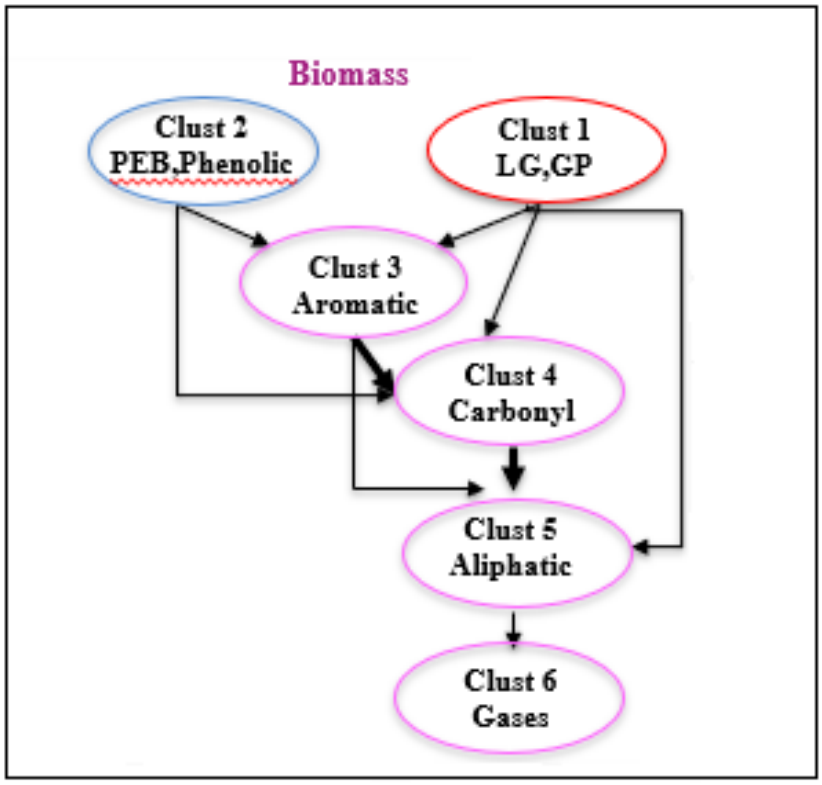


Figure 5-6. Six-cluster Bayesian network for biomass conversion.

It was deduced that the various score-based search algorithms generated the same results providing certainty in the efficacy of the reaction network. Although several feedstocks have been reformed into biocrude through HTL, the reaction mechanism of HTL is a complex one as a result of the complexity of the feedstock. A lot of research works have tried to explain in details the reaction mechanisms and it is assumed that there are three elementary HTL reaction pathways: (a) Depolymerization of the biomass into its main components, (b) Decarboxylisation, decarbonylation, dehydration, and decomposition of biomass monomers by cleavage, and (c) Recombination of reactive remains.[56] This research work proposes the following reaction mechanisms for the HTL of biomass.

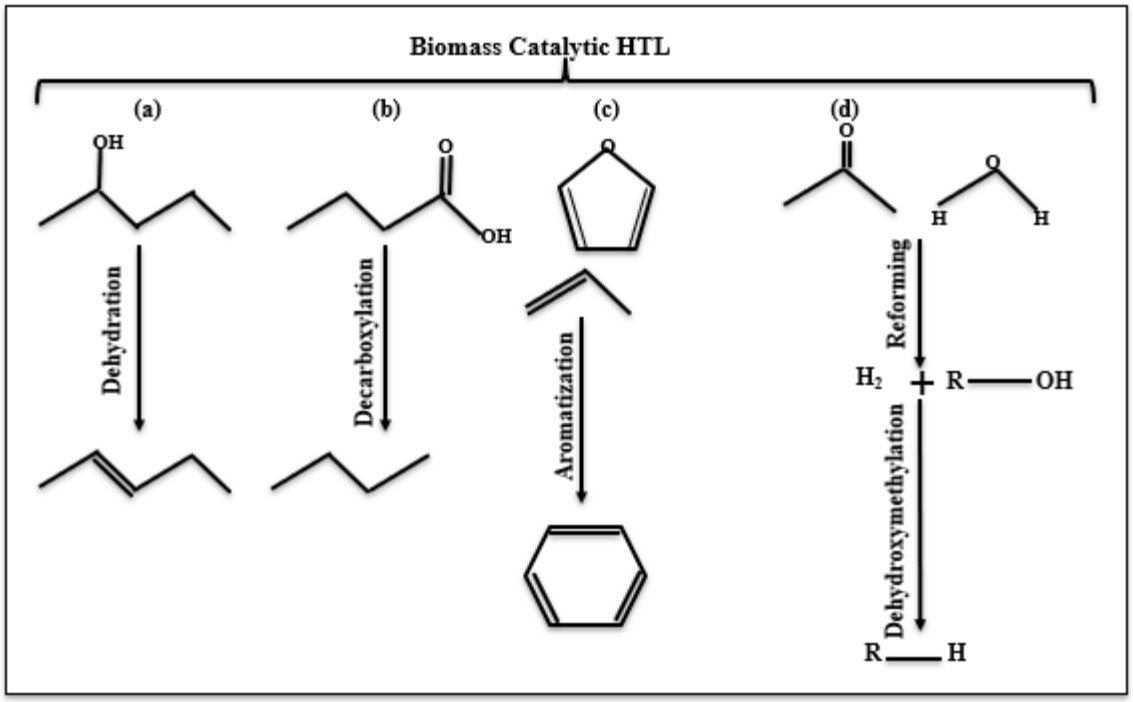


Figure 5-7. Proposed reaction mechanisms for biomass conversion.

The first phase of HTL that is the biomass feedstock is disintegrated into its primary building blocks of cellulose, lignin and hemicellulose. This phase does not exhibit any part of the actual pyrolysis reaction mechanism, but it is essential for adequate comprehension of the reaction modeling as shown in Figure 5-7. Cellulose and hemicelluloses are the most abundant carbohydrates in lignocellulosic biomass. Various carbohydrates have specific rate of hydrolysis. Hemicelluloses hydrolyze quicker than cellulose, as a result of the crystalline structure of

cellulose. The different hydrolyzed products exist in the aqueous fraction derived after hydrothermal liquefaction of biomass. Once carbohydrates are placed under hydrothermal conditions, they experience quick hydrolysis to form glucose and other saccharides. Alcohols are rarely recorded in HTL studies because single alcohols are found in small amounts with vapor pressures corresponding with a meaningful fraction of the biocrude leading to coelution.[57] Nevertheless, some alcohols and saccharides that were basically in the first cluster were discovered in this study (1010, 1012, and 1019 cm^{-1}). The most abundant alcohols were the straight long chain and branched long chain alcohols, and they were produced as a result of the hydrolysis reaction.

When hydrothermal liquefaction of lignin is performed, hydrolysis and cleavage of the ether bond, C-C bond, demethoxylation, alkylation, and condensation reactions take place, and there is counteraction between these chemical reactions.[58] Cleavage of the β-O-4 ether bond has the lead in the breakdown of lignin and its prototypical compounds, while the bond between Cα-Cβ is readily broken.[59] However, when the aromatic rings undergo hydrothermal reactions, they are unchanged and the biphenyl-type compounds were highly stable. Delicate conditions like low temperature and lesser reaction time are needed for the manufacturing of phenolic monomers and dimers from lignin. This is performed by preliminary cleavage of the ether bond and aliphatic C-C bond during hydrothermal liquefaction. An increase in temperature might result in the demethoxylation and alkylation of lignin derived phenolic compounds. Alkyl phenols can also be obtained at high temperature. [60] Lin et al. discovered that during lignin liquefaction, that intermediates with aliphatic side chains exhibited a huge reactivity and further reacted with phenol or with themselves to change to the multi-condensed product.[61] Phenols are highly abundant compounds from HTL of especially carbohydrate and lignin-rich feedstocks and depict compounds of potentially high value and a source of oxygen in the final biofuels. They are present in cluster 2 of the proposed reaction network (1099, 1100, and 1119 cm^{-1}).[62]

The significant wavenumbers in cluster 3 are linked to the oscillation of the benzene and aromatic skeleton (1223, 1254, 1279, and 1500 cm^{-1}). The production of oxygenated aromatics from lignocellulose refining is common. The three monolignols of lignin are the protypes of many aromatic compounds, together with the dehydration reactions of carbohydrates.[63] In contrast to the monofunctional ketones, the oxygenated aromatics naturally exhibit diversified

functionalities resulting in very complex compounds obtained from the complex and heterogeneous arrangement of lignin. The development of monomers originates through thermal breakdown and hydrolysis of ether bonds. [64] At higher temperatures, the decomposition and dehydrogenation reaction of cyclic compounds from alkenes results in the production of aromatic hydrocarbons.[65] Aromatic structures were labeled by absorptions around 1600 cm^{-1} and slight absorptions between 3000 and 3050 cm^{-1} in FTIR spectra (Figure 5-1).

Under subcritical conditions, alkaline water and carbohydrates are known to form carboxylic acids like acetic, propionic, formic, and lactic acid. They can also be subjected to homogeneous and heterogeneous ketonic decarboxylation generating a series of various ketones.[64] Figure 5-1 depicts carbonyl absorption at 1715 cm^{-1} and 1745 cm^{-1} showing six-membered and five-membered cyclic ketones, respectively. Furthermore, compounds categorized in the first cluster are then further degraded to produce several oxygenated hydrocarbons like formic acid, lactic acid, hydroxymethy furfural (HMF), and levullinic acid.[66] Carbonyls are said to be highly abundant in most HTL biocrudes. The most abundant carbonyls were indenones, acetophenones, and a wide range of alkylated chromenones.[67] A recent study states that FTIR spectra of bio-oil in the region of 1490–1850 cm^{-1} could provide comprehensive information on several carbonyl groups in the bio-oil.[68] This work recorded the presence of ketones, aldehydes, and carboxylic acids in cluster 4(1695, 1710, 1723, 1745, and 1749 cm^{-1}).

The development of short chained aliphatic hydrocarbons was noticed in cluster 5 (1332, 1420, 1573, and 1665 cm^{-1}), and this depicts the occurrence of C-C bond cleavage reactions. Glycerol conversion under near- and subcritical water conditions has been outlined to undergo C-C splitting via an ionic and a radical pathway.[64] These compounds result in decarboxylation, decarbonylating, and decomposition reactions.

We observed this from a thermodynamic point of view, the thermochemical conversion of biomass, glucose, and other organic components resulted in light constituents. CH_4 and CO_2 are thermodynamically preferred products with the CO and H_2 yields remaining low.[69] The existence of these molecules can be traced to the last cluster (cluster 6) in the generated BN. The proposed mechanism of HTL of biomass is illustrated below.

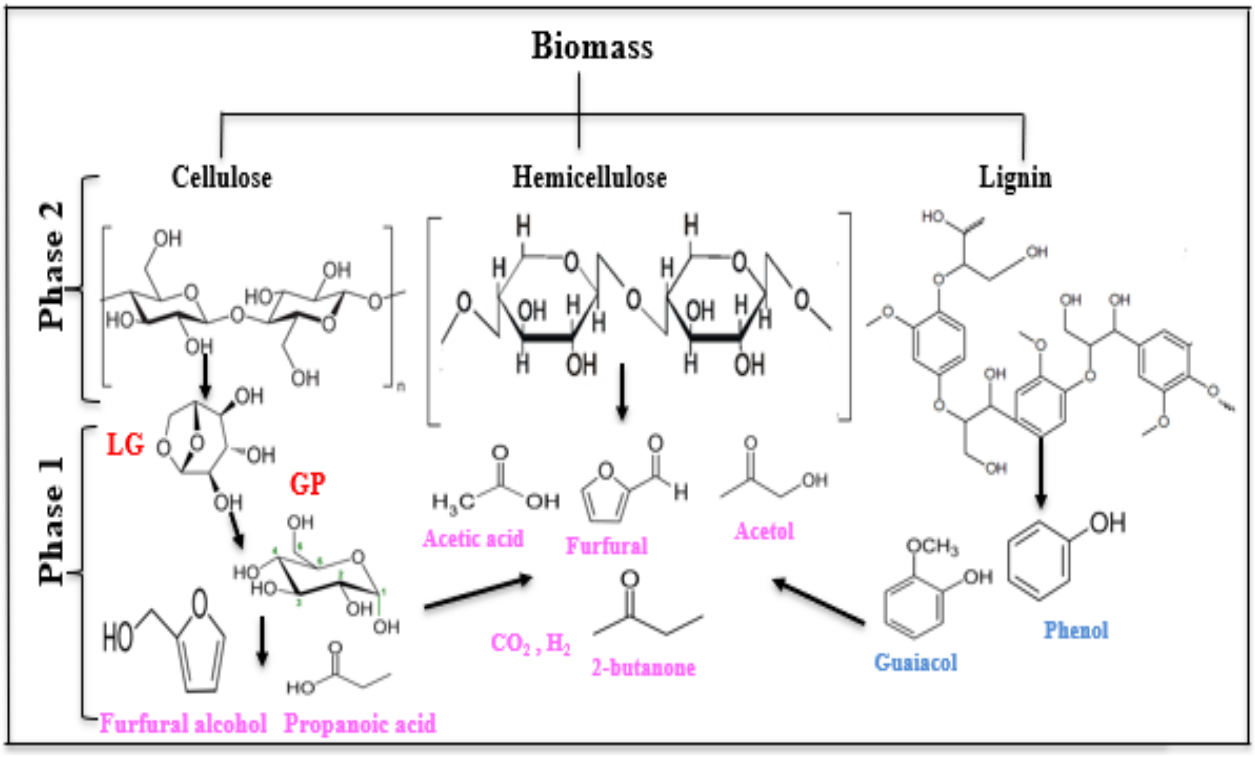


Figure 5-8. Two-phase reaction mechanism for biomass conversion.

To record the dependency and probable pathways between the clusters in a network, the strength value of all the arcs are calculated and provided in Table 5-1 (the strongest dependency belongs to the highest negative number). Hence, the arc from cluster 3 to cluster 4 shows the highest probability for updating itself (34.68) and this implies that once it is removed, the overall score of the network will be decreased automatically by 34.68; implying that ability to make causal inferences decreases in proportion, leading to higher uncertainty about the system.

Table 5-1. Arc strength for the reaction network shown in Figure 5-5 calculated by tabu, HC, and MMHC search methods.

From	To	Arc strength
Cluster 1	Cluster 3	-2.40
Cluster 1	Cluster 4	-8.46
Cluster 1	Cluster 5	-11.44
Cluster 2	Cluster 3	-0.761
Cluster 2	Cluster 4	-17.40
Cluster 3	Cluster 4	-34.68
Cluster 3	Cluster 5	-0.537
Cluster 4	Cluster 5	-33.88
Cluster 5	Cluster 6	-1.87

A causal map between clusters (DAG structure) was generated and the strength value for each arc was calculated. The next phase was studying the limitations which include measuring the uncertainty about the model. This was performed by applying the Markovian property in which the conditional probability of each child is only on its parents considering that in a single path or arc, a child is a descendant of his/her parents in the sequence of the ordered nodes.[70] Therefore, the conditional probability distribution of one group was calculated and presented as the equations 5(1-6) in which they were bases for calculating the model for the mean value of the intensity (root mean square value of the absorbance intensities, RMS) of that group (equations 5(7-12)). As a result, the last four equations were developed to display the mean intensity (X_i is the intensity value of the i_{th} variable (i= 1,2,3,4) and μ_i is the mean value of X_i) which depict to what extent the mean value of the probability distribution of each cluster is linked to others in the network. Therefore, they can be useful to monitor an online process in real time analysis by applying appropriate controls over the process variables.

$$P(X_1) \sim N(\mu_1, 0.041^2) \quad (5\text{-}1)$$
$$P(X_2) \sim N(\mu_2, 0.162^2) \quad (5\text{-}2)$$
$$P(X_3 | X_1, X_2) \sim N(\mu_3, 0.074^2) \quad (5\text{-}3)$$
$$P(X_4 | X_1, X_2, X_3) \sim N(\mu_4, 0.013^2) \quad (5\text{-}4)$$
$$P(X_5 | X_1, X_3, X_4) \sim N(\mu_5, 0.005^2) \quad (5\text{-}5)$$
$$P(X_6 | X_5) \sim N(\mu_6, 0.014^2) \quad (5\text{-}6)$$
$$\mu_1 = 0.195 \quad (5\text{-}7)$$
$$\mu_2 = 0.240 \quad (5\text{-}8)$$
$$\mu_3 = 0.608 + 1.02\,\mu_1 - 0.196\,\mu_2 \quad (5\text{-}9)$$
$$\mu_4 = 0.022 + 0.384\,\mu_1 - 0.153\mu_2 + 0.670\mu_3 \quad (5\text{-}10)$$
$$\mu_5 = -0.076 + 0.209\,\mu_1 - 0.078\mu_3 + 0.766\,\mu_4 \quad (5\text{-}11)$$
$$\mu_6 = 0.035 + 0.130\,\mu_5 \quad (5\text{-}12)$$

5.8. SMCR-ALS RESULTS

To recover reaction networks both qualitatively and quantitatively, this paper explores another approach called self-modeling multivariate curve resolution. Spectroscopic data (FTIR) was arranged into a two-way data type (the response matrix), where objects (spectrum of a sample) were represented in rows and wave numbers (or concentration profiles) where represented in columns. In accordance with the Beer-Lambert-Bouguer law, the matrix was the product of two matrices: C (concentration) and S (spectral absorbances), based on the profiles of the individual N components. The following figure shows the raw, de-noised and background-corrected signals

for these spectra. The removed residual noise is also shown. The convergence criterion for ALS was assessed using the lack of fit (LOF), which provides a measure of the relative fit between experimental data and ALS reconstructed data. The LOF was ultimately determined as 1.13% indicating a solid fit.

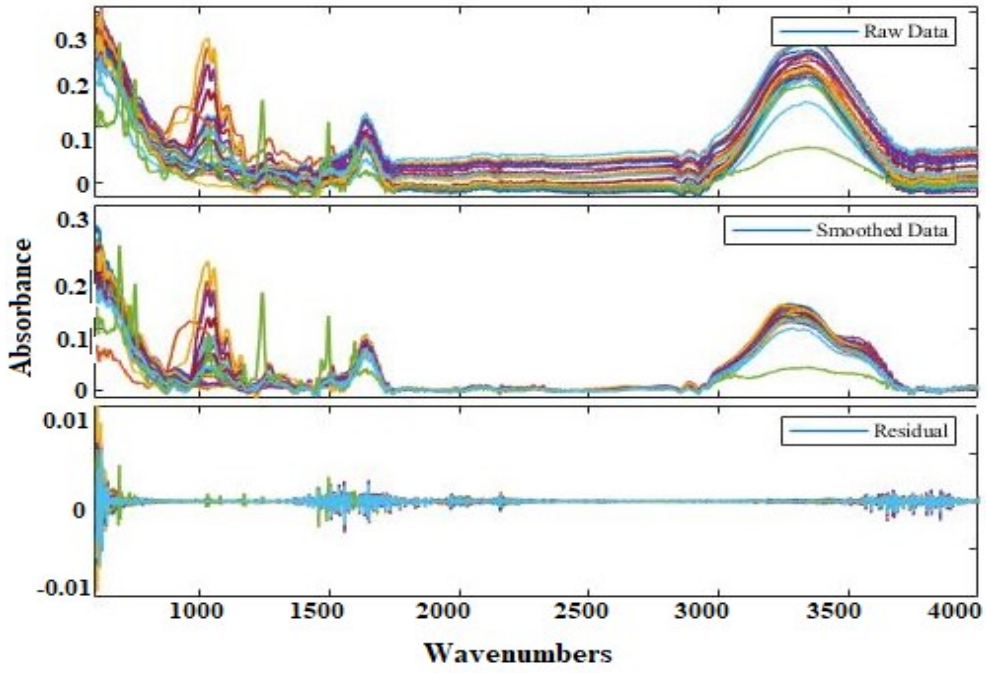


Figure 5-9. Raw, smoothed, and removed residual data for biomass conversion.

The next crucial step was to determine the number of changing chemical species/ranks (N) that occurred during the reaction, though this is difficult because of: (i) measurement noise and its non-assumed distributions; (ii) noise heteroscedasticity; and (iii) co-linearity in the measurement data.[23] Thus, to achieve a good chemical rank estimate, this study attempted to reduce the influence of measurement noise as much as possible by employing a smoothing technique called Savitsky-Golay filtering to directly increase the signal-to-noise ratio. The ratio of the second and third derivatives ROD[28] was then calculated and plotted where the maximum point was 3 indicating the number of chemical ranks (Figure 5-10).

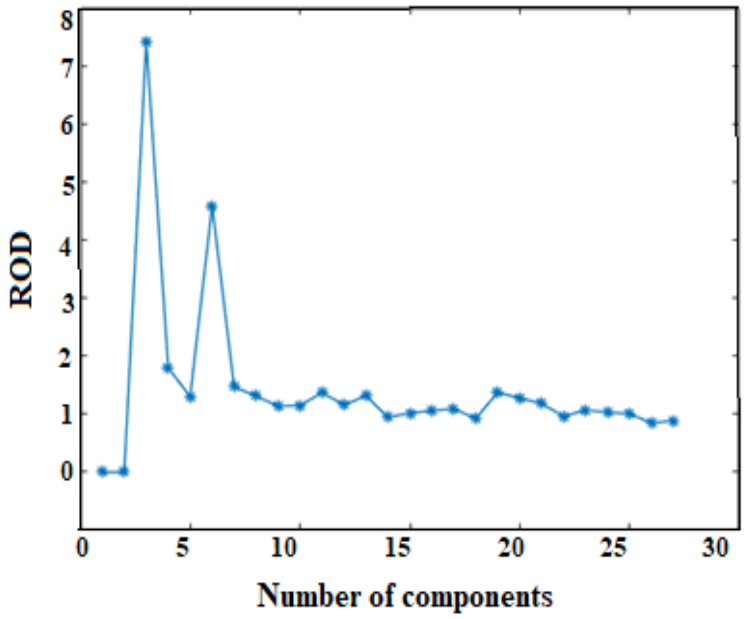


Figure 5-10. Calculating chemical rank using ROD.

In this study, the assumption is that the first, second, and third pseudocomponents represent the actual molecules of A_1, A_2, and A_3, respectively. As a result, to obtain the required necessary information related to these pseudo components, the whole spectrum resolved into spectral profile and concentration of these changing species during the reaction. Figure 5-11(a) demonstrates the overall resolved spectra for all three pseudo components for the entire region (4000-400 cm^{-1}), and Figures 5-11 (b-d) illustrate the regions of high absorbance peaks of these pseudo components.

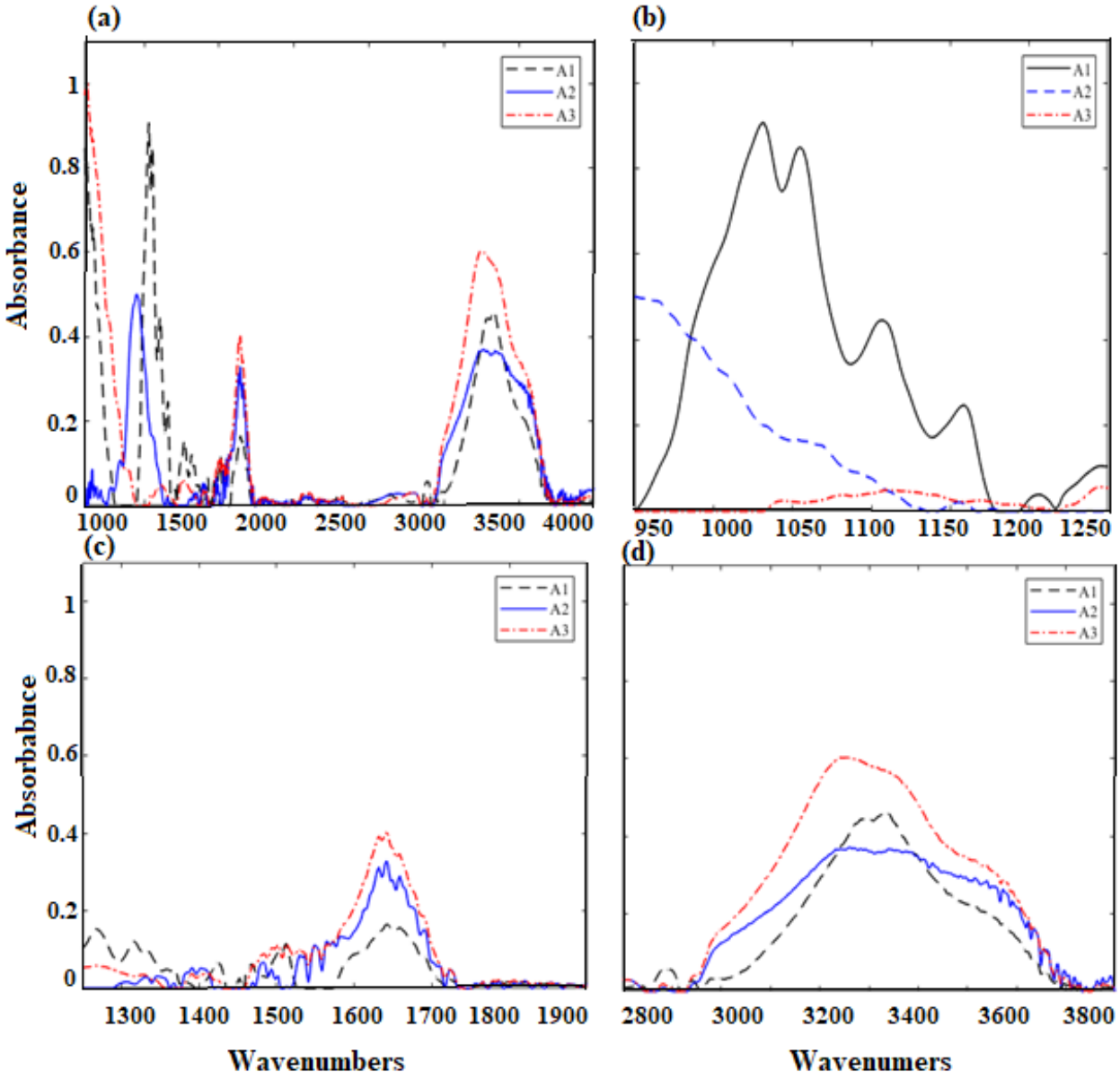


Figure 5-11. (a) Resolved spectra for the pseudocomponents over the whole region; and (b, c, and d) resolved spectra for the pseudocomponents by focusing on the major peaks.

This work also used an initialization method for the SMCR algorithm called Evolving factor analysis (EFA) to determine the initial estimate of concentration of pseudocomponents by exploiting information from global regions. The concentrations of A_1, A_2, and A_3 were used as an input to develop the reaction network. All three score-search heuristic optimization methods to learn BN structures produced the same networks (Figure 5-12). The A_1 pseudocomponents confirm the presence of primary and secondary alcohols (1075-1010 cm^{-1}, and 1120-1100 cm^{-1})

and aliphatic ethers(R-O-R) at 1150-1070 cm^{-1}. The spectrum for A_2 illustrates the presence of pseudocomponents associated predominantly with phenolic groups (broad absorbance between 3550-3200 cm^{-1}). A_3 pseudocomponents indicate the presence of carboxylic acids (3550-3500 cm^{-1}), ketones (3550-3205 cm^{-1}), and aryl aldehydes (1715-1695 cm^{-1}), as well as aromatic compounds due to the presence of C=C vibrations with absorbance peaks from 1625-1575 cm^{-1} or C-C in-ring stretching at 1500-1400 cm^{-1}.

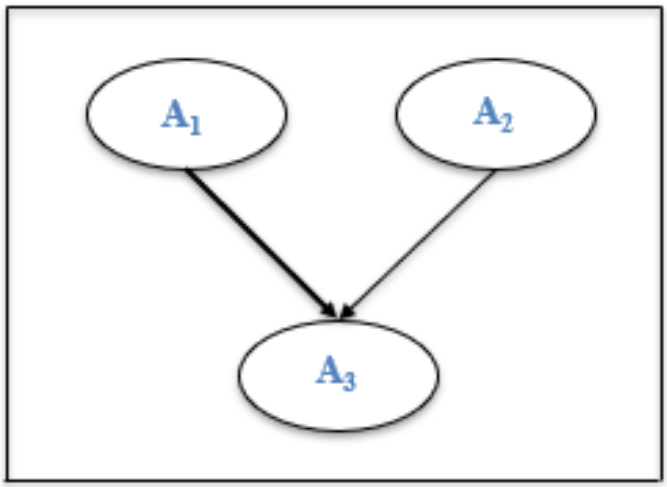


Figure 5-12. BN for pseudocomponents from SMCR-ALS.

Table 5-2. The strength values of each arc in BN shown in Figure 5-11.

From	To	Arc strength
Cluster 1	Cluster 3	-18.31
Cluster 2	Cluster 3	-8.634

To better understand parameter learning in BN, equations 5(13-15) represent each group's conditional probability, and the models presenting mean intensities are shown in equations 5(16-18).

$$P(X_1) \sim N(\mu_1, 0.38^2) \quad (5\text{-}13)$$
$$P(X_2) \sim N(\mu_2, 0.43^2) \quad (5\text{-}14)$$
$$P(X_3 | X1, X_2) \sim N(\mu_3, 0.51^2) \quad (5\text{-}15)$$
$$\mu_1 = 0.34 \quad (5\text{-}16)$$
$$\mu_2 = 0.27 \quad (5\text{-}17)$$
$$\mu_3 = 0.878 - 1.11\mu_1 - 0.51\mu_2 \quad (5\text{-}18)$$

5.9. DISCUSSIONS

Since all lignocellulosic biomass is largely composed of three basic independent structural components (cellulose, hemicellulose, and lignin), any aggregative behavior of these components during pyrolysis describes the behavior of any lignocellulosic feed.[71] Furthermore, since biomass pyrolysis follows a complex network of reaction mechanisms, biomass pyrolysis chemistry can be simplified by studying the independent pyrolysis reactions of each individual component. If synergistic effects occur, predicting a biomass feedstock's behavior would be considerably more complex. Hence, in the previous works, cellulose and lignin independently underwent HTL conversion using their model components: levoglucosan (LG) and 2-Phenoxyethyl benzene (PEB), respectively; and their most probable BNs were developed (Figure 5-13).

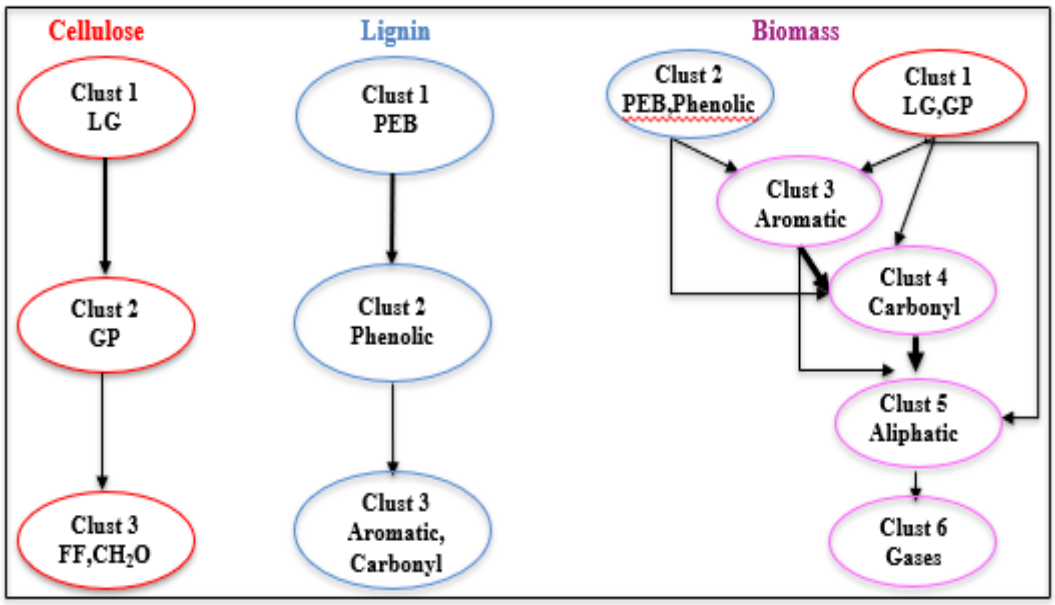


Figure 5-13. BNs for (a) LG, (b) PEB, and (c) biomass (data provided by data fusion).

After reviewing and comparing wave numbers in each cluster, it was inferred that the right side of biomass conversion from cluster 1 mostly represents hydrocarbon (cellulose or hemicellulose) conversion while the left side represents lignin conversion, though both have the same final products: aromatics, carbonyl groups, aliphatic hydrocarbons, and smaller molecules. In the previous work, one of the major products from the HTL of cellulose-related model compounds was formaldehyde, identified by FTIR, ^{1}H NMR, and GC-MS. However, after performing HTL

of biomass, the presence of formaldehyde was also confirmed by bands at 3308, 2982 and 2914 cm^{-1} for –CH stretch, a clear peak at 1636 cm^{-1} for –C=O for aldehyde, and bands at 1429, 1271, 1103, 1019, and 989 cm^{-1}. Figure 5-14 shows the mechanism of formaldehyde formation. Following glycosidic bond cleavage, hydrogen from the hydroxyl group of the carbon atom ^{6}C is transferred to ^{5}C. This is convoyed by cleavage of the ^{5}C–^{6}C bond, resulting in formaldehyde formation. These conclusions are consistent with the 2-phase reaction pathway proposed earlier (Figure 5-8).

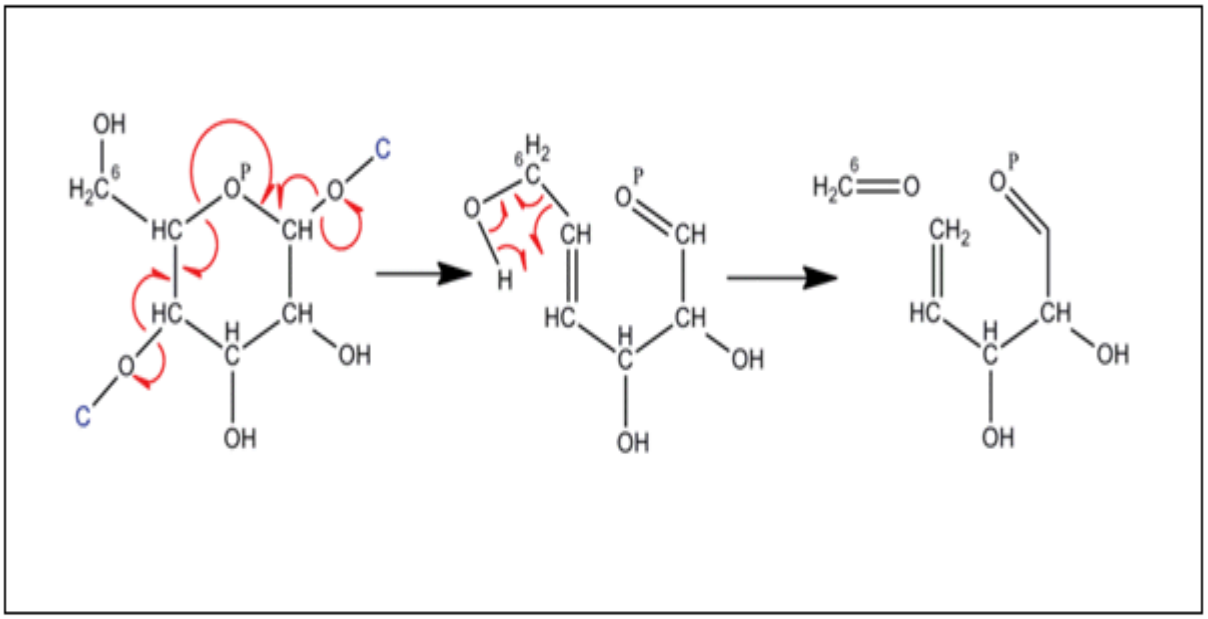


Figure 5-14. Cellulose HTL pathway to Formaldehyde.

The conclusion following development of the SMCR-ALS algorithm, detecting causality between the pseudocomponents of biomass conversion, and comparing these with the individual cellulose and lignin results was that the biomass and cellulose conversions closely follow the same causality map (Figure 5-15). These results are consistent previous studies using other methods. [72] For instance, it was confirmed by Muley et al. that the pyrolytic behavior of biomass can be affected more by cellulose rather than lignin, because cellulose can be more easily broken-down during pyrolysis at lower temperatures. This is due to the fact that cellulose has simpler molecular structure, lower thermal stability, and less fixed carbon content.[71] For cellulose conversion, we proposed that pseudo compounds presented by A_1 [(-OH), (-C=C-)] and A_2 (R-O-R')] can be converted to ortho-hydroxy aryl ketones (C=O)-CH=(C-OH) by alkyl and hydrogen transfers and oxidation reactions.[28] In fact, some signature wavenumbers identifying

alcohols, ethers, and alkenes collected in A_1 and A_2 from cellulose conversions can be traced in A_1 of the biomass conversion. In addition, in lignin conversion (Figure 5-15(b)), A_1 and A_2, which represent phenolic and aromatics, could be mapped to A_2 from biomass conversion. Overall, similarities between pseudo components were found in the last cluster of cellulose and lignin with A_3 from the biomass conversion.

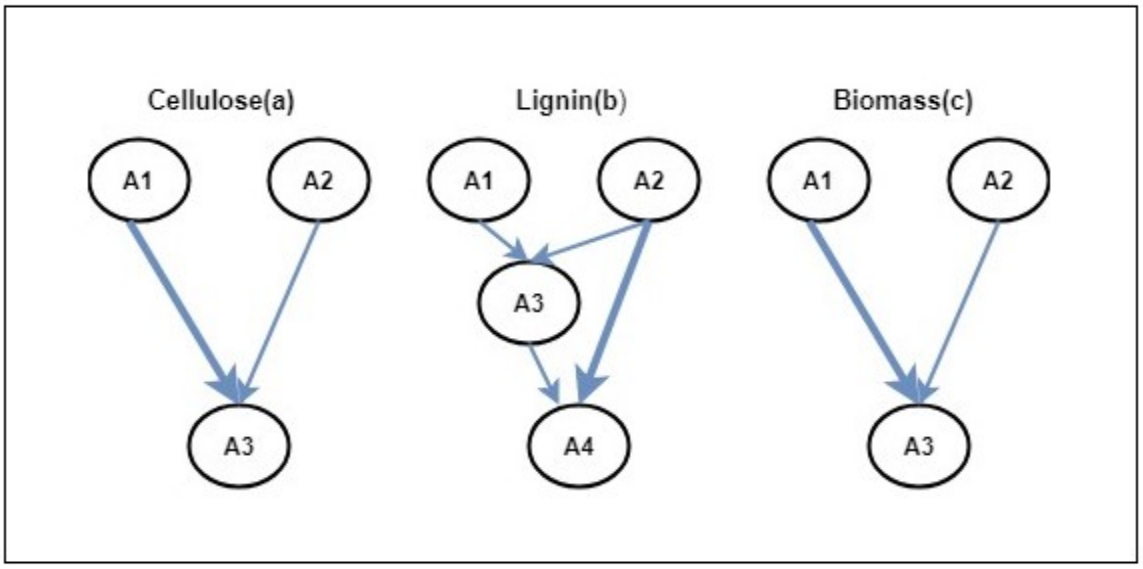


Figure 5-15. The BNs from SMCR-ALS method for (a) cellulose, (b) lignin, and (c) biomass conversion.

Another objective of this study which was developing an algorithm for online monitoring of species conversion in a complex reaction was attained by tracing the concentrations of A_1, A_2 and A_3 over the number of samples (Figure 5-16). Interestingly, the trends of these concentration profiles are with a good agreement with the BN which was developed earlier in this study (Figure 5-12); where the concentration of A_3 increases at the expense of the A_1 and A_2 concentrations as the processing time is increased. Hence, this plot can be useful by integrating it with a suitable control strategy that adjusts process conditions to maximize the yield of the desired product for online monitoring of a multi-spices process such as biomass conversion.

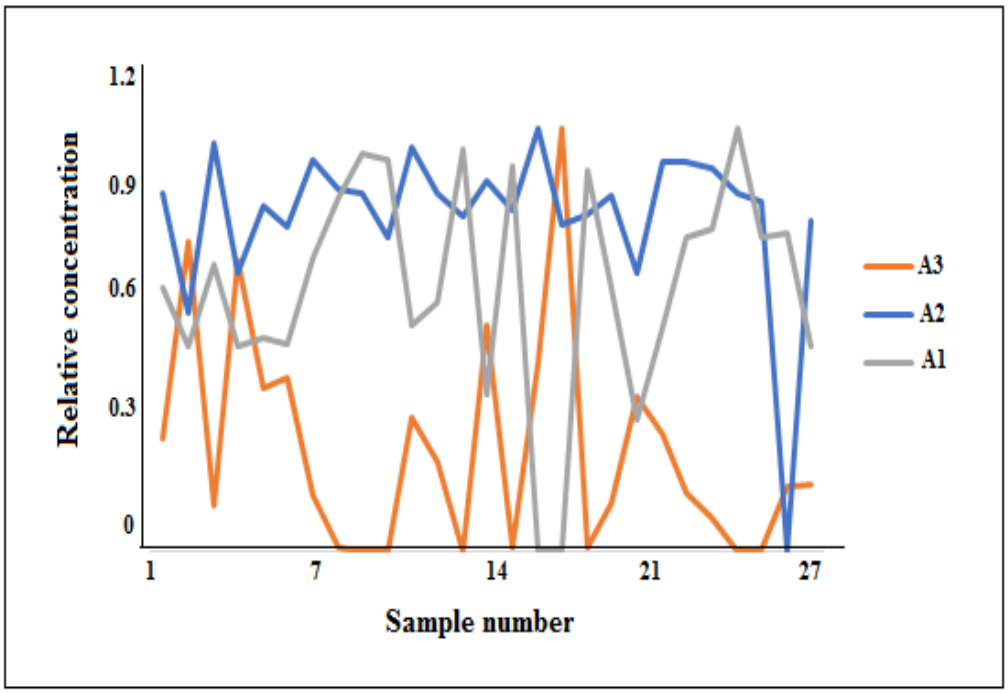


Figure 5-16. Corresponding concentration for pseudo components A_1, A_2, and A_3.

5.10. CONCLUSION

To overcome the limitations of high heat and vaporization in biomass conversion processes, this study applied HTL where water is an important reactant and catalyst, and thus the biomass can be directly converted without an energy consuming drying step, as is the case for pyrolysis. The reaction's chemistry is complicated and highly substrate dependent. Therefore, to recover the most probable biomass reaction mechanism of HTL, data fusion and a data mining approach called the BN learning were used. We used statistical variables to restrict the set of networks to be assessed. To obtain an optimal solution for a reaction network, three different BN learning techniques were applied, resulting in a final six-cluster BN mechanism. Wavenumbers collected in each cluster were identified and interpreted. The pathway to biocrude represents biomass main components' decomposition, large molecule hydrolysis, and reformation of produced molecules. In other words, producing molecules such as aromatics, carbonyl groups, alkanes, alkenes, alkynes, and gasses from larger molecules including alcohols, ethers, and phenolic groups. This study also developed a data-driven algorithm to describe the process and chemistry for ultimate use for real-time analysis and the optimizing hydrothermal cracking of biomasses in

150 to 350 °C using FTIR spectroscopy data. Only three pseudocomponents were involved in the reaction, so the whole spectra were resolved into profiles of these pseudocomponents and their concentrations were computed for use as an input for causality detection. These results produced new ways of generating reaction networks in biocrude fractions that are otherwise complex to distinguish.

5.11. REFERENCES

1. Kruse A, Krupka A, Schwarzkopf V, Gamard C, Henningsen T. Influence of proteins on the hydrothermal gasification and liquefaction of biomass. 1. Comparison of different feedstocks. *Ind Eng Chem Res*. 2005;44(9):3013-3020. doi:10.1021/ie049129y
2. Balat M, Balat M, Kirtay E BH. Main routes for the thermo-conversion of biomass into fuels and chemicals. Part 1: pyrolysis systems. *Energy Convers Manag*. 2009;50(12):3147–57.
3. Tiilikkala K, Fagernas L TJ. History and use of wood pyrolysis liquids as biocide and plant protection product history and use of wood pyrolysis liquids as biocide and plant protection product. *Open Agric J*. 2010;4:111-118.
4. Vitasari C, Meindersma G de HA. Conceptual process design of an integrated bio-based acetic acid, glycolaldehyde, and acetol production in a pyrolysis oil-based biorefinery. *Chem Eng Res*. 2015;95:133–43.
5. Lopez Barreiro D, Prins W, Ronsse F BW. Hydrothermalliquefaction (HTL) of microalgae for biofuel production: state of the artreview and future prospeects. *Biomass Bioenergy*. 2013;53:113–127.
6. Narayan S, Muldoon J, Finn M, Fokin VV, Kolb HC SK. On water: unique reactivity of organic compounds in aqueous suspension. *Angew Chem Int Ed*. 2005;44:3275–9.
7. Andrew A. Peterson, Frédéric Vogel, Russell P. Lachance, Morgan Fröling, Michael J. Antal J and JWT. Thermochemical biofuel production in hydrothermal media: A review of sub- and supercritical water technologies. *Energy and Envirometal Sicence*. 2008;(1):32-65.
8. Elliott DC, Biller P, Ross AB, Schmidt AJ, Jones SB. Hydrothermal liquefaction of biomass: Developments from batch to continuous process. *Bioresour Technol*. 2015;178:147-156. doi:10.1016/j.biortech.2014.09.132
9. Uihlein A, Schebek L. Environmental impacts of a lignocellulose feedstock biorefinery system: An assessment. *Biomass and Bioenergy*. 2009;33(5):793-802. doi:10.1016/j.biombioe.2008.12.001
10. E.J. Soltes TJE. Pyrolysis, in: I.S. Goldstein (Ed.), Organic Chemicals from Biomass. *CRC Press Boca Rat*. 1981:63-99.
11. Chalmers JM. Spectroscopy in Process Analysis. *Sheff Acad Press*. 2000.
12. David Lee Hall SAHM. *Mathemathical Techniques in Multisensor Data Fusion*. ARTEC House, INC; 2004.
13. Castanedo F. A review of data fusion techniques. *Sci World J*. 2013;2013. doi:10.1155/2013/704504
14. Friedman, Nir, Nachman, Iftach. Peer D. Learning Bayesian Network Structure from Massive Datasets: The “Sparse Candidate” Algorithm. In: *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*. ; 1999:206-215.
15. Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. *Springer Sci*. 2006:31-78. doi:10.1007/s10994-006-6889-7
16. Friston K, Moran R, Seth AK. Analysing connectivity with Granger causality and dynamic causal modelling. *Curr Opin Neurobiol*. 23(2):172-178. doi:10.1016/j.conb.2012.11.010
17. Guo S, Wu J, Ding M FJ. Uncovering interactions in the frequence domain. *PLoS Comput Biol*. 2008;4(5).
18. ansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M,

Greenblatt JF GMA. Bayesian Networks Approach for Predicting Protein-Protein Interactions from Genomic Data. *Science (80-)*. 2003:302: 449.
19. Zou C, Denby KJ, Feng J. Granger causality vs. dynamic Bayesian network inference: A comparative study. *BMC Bioinformatics*. 2009;10. doi:10.1186/1471-2105-10-401
20. Jia H, Li Y, Dong B, Ya H. An Improved Tabu Search Approach to Vehicle Routing Problem. *Procedia - Soc Behav Sci*. 2013;96(Cictp):1208-1217. doi:10.1016/j.sbspro.2013.08.138
21. Jiang T, Ren G, Zhao X. Evacuation Route Optimization Based on Tabu Search Algorithm and Hill-Climbing Algorithm. *Procedia - Soc Behav Sci*. 2013;96(Cictp):865-872. doi:10.1016/j.sbspro.2013.08.098
22. Awa K, Okumura T, Shinzawa H, Otsuka M, Ozaki Y. Self-modeling curve resolution (SMCR) analysis of near-infrared (NIR) imaging data of pharmaceutical tablets. *Anal Chim Acta*. 2008;619(1):81-86. doi:10.1016/j.aca.2008.02.033
23. Jiang JH, Liang Y, Ozaki Y. Principles and methodologies in self-modeling curve resolution. *Chemom Intell Lab Syst*. 2004;71(1):1-12. doi:10.1016/j.chemolab.2003.07.002
24. Beyramysoltan S, Abdollahi H, Rajkó R. Newer developments on self-modeling curve resolution implementing equality and unimodality constraints. *Anal Chim Acta*. 2014;827:1-14. doi:10.1016/j.aca.2014.03.019
25. R. Tauler, F.C. Sanchez DLM. Validation of alternating least squares multivariate curve resolution for chromatographic resolution and quantitation. *Trends Anal Chem*. 1996;15:279–286.
26. D. Bylund, R. Danielsson KEM. Peak purity assessment in liquid chromatography-mass spectrometry. *J Chromatogr*. 2001:43-52.
27. B. Ma, P.J. Gemperline, E. Cash, M. Bosserman E. Characterizing batch reactions with in situ spectroscopic measurements, calorimetry and dynamic modeling. *J Chemom*. 2003;17:470–479.
28. Sattari F, Tefera D, Sivaramakrishnan K, De Klerk A, Prasad V. Postulating Pseudo-reaction Networks for the Conversion of Levoglucosan in Hydrous Pyrolysis Using Spectroscopic Data & Self-Modeling Multivariate Curve Resolution. *Submiited to I&EC*. 2019.
29. Yeh TM, Dickinson JG FA et al. Hydrothermal catalytic production of fuels and chemicals from aquatic biomass. *J Chem Technol Biotechnol*. 2013;88(1):13-24.
30. PE S. A perspective on catalysis in sub- and supercritical water. *J Supercrit Fluids*. 2009;47(3):407-414.
31. Kabir G, Hameed BH. Recent progress on catalytic pyrolysis of lignocellulosic biomass to high- grade bio-oil and bio-chemicals. *Renew Sustain Energy Rev*. 2017;70(January):945-967. doi:10.1016/j.rser.2016.12.001
32. Staš M, Kubička D, Chudoba J, Pospíšil M. Overview of analytical methods used for chemical characterization of pyrolysis bio-oil. *Energy and Fuels*. 2014;28(1):385-402. doi:10.1021/ef402047y
33. Bassilakis R, Carangelo RM, Wojtowicz MA. TG-FTIR analysis of biomass pyrolysis. *Fuel*. 2001;80(12):1765-1786.
34. Mistry B. *A Handbook of Spectroscopic Data Chemistry (UV, IR, PMR, 13CNMR and Mass Spectroscopy)*. Oxford Book Company; 2009. doi:10.1063/1.2719251
35. Bonnefont CMD, Guerra A, Théron L, Molette C, Canlet C, Fernandez X. Metabolomic

study of fatty livers in ducks: Identification by1H-NMR of metabolic markers associated with technological quality. *Poult Sci*. 2014;93(6):1542-1552. doi:10.3382/ps.2013-03546
36. Llinas DLH and J. An introduction to multisensor data fusion. *Proc IEEE*. 1997;85(1):6-23.
37. Dong J, Zhuang D, Huang Y, Fu J. Advances in Multi-Sensor Data Fusion: Algorithms and Applications. 2009;(1):7771-7784. doi:10.3390/s91007771
38. Sattari F, De Klerk A, Prasad V. Application of data combination and data mining techniques to investigate the chemistry of cellulose and lignin derivatives in hydrous pyrolysis. *Submiited to I&EC*. 2019.
39. Hammer HL, Yazidi A, Oommen BJ. "Anti-Bayesian" flat and hierarchical clustering using symmetric quantiloids. *Inf Sci (Ny)*. 2017;418-419(August 2016):495-512. doi:10.1016/j.ins.2017.08.017
40. Frühwirth-Schnatter S KS. Model-based clustering of multiple time series. *J Bus Econ Stat*. 2008;26:78–89.
41. Jackson E, Davy M, Doucet A FW. Bayesian Unsupervised Signal Classification by Dirichlet Process Mixtures of Gaussian Processes. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*. ; 2007.
42. Beaumont M RB. The Bayesian revolution in genetics. *Nat Rev Genet*. 2004;(4):251–261.
43. Savage RS, Heller K, Xu Y, et al. R / BHC : fast Bayesian hierarchical clustering for microarray data. 2009;9:1-9. doi:10.1186/1471-2105-10-242
44. Tadesse MG, Sha N, Vannucci M. Bayesian variable selection in clustering high-dimensional data. *J Am Stat Assoc*. 2005;100(470):602-617. doi:10.1198/016214504000001565
45. Wentzell PD, Wang J-H, Loucks LF MK. Direct optimization of self-modeling curve resolution: application to the kinetics of the permanganate—oxalic acid reaction. *Can J Chem*. 1998;76:1-12.
46. Tauler R, Smllde AK, Henshaw JM, Burgess QLW, Kowalskl BR. Multicomponent Determination of Chlorinated Hydrocarbons Using a Reaction-Based Chemical Sensor . 2 . Chemical Speciation Using Multivariate Curve Resolution. 1994. doi:10.1021/ac00092a009
47. Gemperline PJ. Computation of the Range of Feasible Solutions in Self-Modeling Curve Resolution Algorithms. *Anal Chem*. 1999;71(23):5398–5404.
48. Jiang J-H, Liang Y-Z OY. On simplex-based method for self-modeling curve resolution of two-way data. *ChemomIntell Lab Syst*. 2003;65:51-65.
49. Jiang J, Ozaki Y. Self -modeling curve resolution (SMCR): principals, techniques, and applications T. *Ind Eng Chem Res*. 2017;4928(October). doi:10.1081/ASR-120014359
50. Jiang J, Ozaki Y. SELF-MODELING CURVE RESOLUTION (SMCR): PRINCIPLES , TECHNIQUES , AND APPLICATIONS. 2007;4928. doi:10.1081/ASR-120014359
51. Maeder M. Evolving factor analysis for the resolution of overlapping chromatographic peaks. *Anal Chem*. 1987;59(3):527-530.
52. Tefera DT, Agrawal A, Yañez Jaramillo LM, De Klerk A, Prasad V. Self-Modeling Multivariate Curve Resolution Model for Online Monitoring of Bitumen Conversion Using Infrared Spectroscopy. *Ind Eng Chem Res*. 2017;56(38):10756-10769. doi:10.1021/acs.iecr.7b01849
53. Tefera DT, Yañez Jaramillo LM, Ranjan R, Li C, De Klerk A, Prasad V. A Bayesian Learning Approach to Modeling Pseudoreaction Networks for Complex Reacting

Systems: Application to the Mild Visbreaking of Bitumen. *Ind Eng Chem Res*. 2017;56(8):1961-1970. doi:10.1021/acs.iecr.6b04437
54. Zhang L, Shen C, Liu R. GC – MS and FT-IR analysis of the bio-oil with addition of ethyl acetate during storage. 2014;2(January):1-6. doi:10.3389/fenrg.2014.00003
55. Dunteman BGH. *Principal Components Analysis*. SAGD publications, Inc; 1976.
56. Y. Z. Hydrothermal liquefaction to convert biomass into crude oil. In: Biofuels from agricultural wastes and byproducts. *Hoboken, NJ Wiley-Blackwell*. 2010:201-232.
57. Gai C, Zhang Y, Chen W-T, Zhang P DY. An investigation of reaction pathways of hydrothermal liquefaction using Chlorella pyrenoidosa and Spirulina platensis. *Energy Convers Manag*. 2015;96:330-339.
58. S. Kang, X. Li, J. Fan JC. Classified separation of lignin hydrothermal liquefied products. *Ind Eng Chem Res*. 2011;50:11288–11296.
59. H. Wang, Y. Zhao, C. Wang, Y. Fu QG. Theoretical study on the pyrolysis process of lignin dimer model compounds. *Acta Chim Sin*. 2009;9:893-900.
60. K. Ehara, S. Saka HK. Characterization of the lignin-derived products from wood as treated in supercritical water, J. Wood Sci. 48 (2002) 320–325. *J Wood Sci*. 2002;48:320–325.
61. L. Lin, M. Yoshioka, Y. Yao NS. Liquefaction mechanism of lignin in the presence of phenol at elevated temperature without catalysts. Studies on β-O-4 lignin model compound III. Multi-condensation. *Holzforschung*. 1997;51:333–337.
62. Madsen RB, Zhang H, Biller P, Goldstein AH, Glasius M. Characterizing Semivolatile Organic Compounds of Biocrude from Hydrothermal Liquefaction of Biomass. *Energy and Fuels*. 2017;31(4):4122-4134. doi:10.1021/acs.energyfuels.7b00160
63. Tompsett GA, Li N HG. Catalytic conversion of sugars to fuels. *John Wiley Sons*. 2011:273-279.
64. Pedersen TH, Jensen CU, Sandström L, Rosendahl LA. Full characterization of compounds obtained from fractional distillation and upgrading of a HTL biocrude. *Appl Energy*. 2017;202:408-419. doi:10.1016/j.apenergy.2017.05.167
65. Zhang X, Wang T, Ma L, Chang J. Vacuum pyrolysis of waste tires with basic additives. *Waste Manag*. 2008;28(11):2301-2310. doi:10.1016/j.wasman.2007.10.009
66. Wahyudiono, M. Sasaki MG. Conversion of biomass model compound under hydrothermal conditions using batch reactor. *Fuel*. 2009:1656–1664.
67. Huang X, Cheng D, Fengqiu Chen, Zhan X. Reaction pathways of hemicellulose and mechanism of biomass pyrolysis in hydrogen plasma: A density functional theory study. *Renew Energy*. 2016;96:490-497.
68. Lievens, C., Mourant, D., He, M., Gunawan, R., Li, X., and Li C-Z (2011). An FT-IR spectroscopic study of carbonyl functionalities in bio-oils. Fuel 90. *Fuel*. 2011;90(3417).
69. Kruse A, Gawlik A. Biomass conversion in water at 330-410 °C and 30-50 MPa. Identification of key compounds for indicating different chemical reaction pathways. *Ind Eng Chem Res*. 2003;42(2):267-279. doi:10.1021/ie0202773
70. Lee SM, Abbott PA. Bayesian networks for knowledge discovery in large datasets: Basics for nurse researchers. *J Biomed Inform*. 2003;36(4-5):389-399. doi:10.1016/j.jbi.2003.09.022
71. Muley PD, Henkel C, Abdollahi KK, Marculescu C, Dorin Boldor. A critical comparison of pyrolysis of cellulose, lignin, and pine sawdust using an induction heating reactor. *Energy Convers Manag*. 2016;117:273–280.

72. Stefanidis SD, Kalogiannis KG, Iliopoulou EF, Michailof CM, Pilavachi PA, Lappas AA. A study of lignocellulosic biomass pyrolysis via the pyrolysis of cellulose, hemicellulose and lignin. *J Anal Appl Pyrolysis*. 2014;105:143-150. doi:10.1016/j.jaap.2013.10.013

6. CONCLUSIONS AND FUTURE WORKS

6.1. CONCLUSIONS

Biomass has been considered as a potential replacement for fossil fuels because of its sustainability, availability, and environmentally friendly characteristics. In the past few decades, computational resources have quickly evolved; however, the extensive network modeling of a complex system such as biomass conversion remains a challenging task. To overcome this limitation, this study employed data fusion and data mining as two powerful techniques in the process of knowledge discovery from hidden interesting patterns in the large data sets provided by spectroscopic techniques. Therefore, several necessary steps were taken: data cleaning or de-noising, data integration, data selection, data transformation, data mining, and knowledge presentation. Taking these steps resulted in developing the most probable graphical models (BNs) to represent interactions between variables describing the biomass conversion using representative datasets and statistics founded on Bayes' rule of conditional probability. These developed BNs would be useful to enhance investigation of hydrous pyrolysis of biomass along with the capability of self-updating the arc dependencies in the network structure which makes them favorable for real-time analysis of the process. To achieve the main goal of this research which was develop working algorithms to enhance investigation of chemistry of biomass conversion, the following main objectives were also enhanced:

- Developing an algorithm for integrating multiple data sets provided by FTIR and ^{1}H NMR in order to increase the accuracy while reduce the overall uncertainty.
- Developing an algorithm that could combine expert knowledge and experimental data to develop the most likely reaction network for a complex reactive system such as hydrous pyrolysis of biomass. These plausible models have the advantages of providing a theoretical basis to handle uncertainty.
- Designing an SMCR-ALS algorithm for automatic resolution of spectra provided by online spectroscopic techniques which could be used for predicting the effect of intervention qualitatively and quantitatively.
- Demonstrating the efficacy of the developed algorithms in identifying the chemistry of model compounds and biomass.

These goals and objectives were accomplished by the conducting and investigating of hydrous pyrolysis of model compounds representative of the processing of cellulose and lignin, a physical mixture of the model compounds, and real biomass, as well as developing three different Bayesian network approaches and parameter learning for each process. Major conclusions from the study are summarized below.

The most probable pyrolytic pathway for HTL of levoglucosan (LG, an important primary product of cellulose decomposition) is a causal map with three clusters which can be explained as converting LG to formaldehyde (CH_2O) and furfural (FF) directly or by producing an intermediate compound such as glucose. Based on the pathway proposed for this reaction, the formation of formaldehyde and furfural is the result of the dehydroxymethylation reaction of the side chain of the furan-ring. The number of pseudocomponents found using an alternate algorithm (SMCR-ALS) was determined to be three. The causality map from this analysis indicated that pseudocomponents represented by A_1 [(-OH), (-C=C-)] and A_2 (R-O-R')] can be converted to ortho-hydroxy aryl ketones (C=O)-CH=(C-OH) by alkyl and hydrogen transfers and oxidation reactions.

When it comes to comparing the BHC and SMCR as methods to generate nodes among which the Bayesian networks are developed; the latter one provides better results. Applying the BHC technique groups wavenumbers in a cluster without any background constraints on how chemistry changes over different process conditions. On the other hand, the SMCR algorithm determines the concentration changes of pseudocomponents across different process conditions along with a full spectrum for each. This makes it more appropriate for the on-line monitoring of change in species during a process. However, in terms of mapping to real chemistry, interpreting the results from SMCR is more difficult.

After conducting HTL of 2-Phenoxyethyl benzene (PEB, a model compound for lignin) and LG, individually, and analyzing and comparing the BNs from FTIR technique and fusion method, it was confirmed that the latter method presented some major improvements. In other words, data fusion provides the result that is more consistent with the findings of density functional theory (DFT) in the literature. Regarding PEB decomposition, after applying data fusion and data mining, a three - cluster BN was identified the optimal network. This network presents the conversion of PEB (cluster 1) into final products (cluster 3) which were identified as aromatics,

alcohols, aldehydes, ketones, and alkenes by the cleavage of C-O-C and C-O bonds. Cluster 2 was identified as consisting mostly of compounds containing phenolic groups which could be the result of hydrogen abstraction by the phenoxy radicals from the homolysis of PEB.

In the search for a deeper understanding of biomass conversion, a physical mixture of cellulose and lignin underwent HTL. All three different score-based greedy search algorithms provided the same BNs while considering three and four clusters. The novelty of the three-cluster BN in this case was that the pathway from cluster 1 to 3 presents the conversion of lignin and phenolic groups into the final products with almost the same arc value (-7.47)., while the arc pointing out from cluster 2 to 3 shows the conversion of cellulose and glucose to the final products, again, presenting the same dependency strength (-43.99). The advantage of calculating these arc values is that they represent the dependency and probable pathways between the clusters in the network (the strongest dependency belongs to the highest negative number). Therefore, the arc from cluster 2 to cluster 3 shows the higher probability for updating itself (43.99) and this infers that once it is removed, the overall score of the network will be decreased automatically by 43.99; implying that ability to make causal inferences decreases in proportion, leading to higher uncertainty about the system.

In addition to these processes, HTL of real biomass was also conducted, and samples confirmed the presence of aldehydes, ketones, acids, aromatics, phenols, ethers, aliphatic compounds and alcohols in the bio-oil. To explain the reaction mechanisms, it was proposed that there are three elementary HTL reaction pathways (a) Depolymerization of the biomass into its main components, (b) Decarboxylisation, decarbonylation, dehydration, and decomposition of biomass monomers by cleavage, and (c) Recombination of reactive remains. Employing data mining resulted in a final six-cluster BN mechanism. After reviewing and comparing wavenumbers in each cluster, it was inferred that one part of the BN (cluster 1) mostly represents hydrocarbon (cellulose or hemicellulose) conversion while the other part (cluster 2) represents lignin conversion, though both have the same final products (aromatics, carbonyl groups, aliphatic hydrocarbons, and smaller molecules). Furthermore, detecting causality between the pseudocomponents (obtained from SMCR-ALS) of biomass conversion and comparing it with the individual results from LG and PEB conversion indicated that the pyrolytic behavior of biomass can be affected more by cellulose rather than lignin, because cellulose can be more

easily broken-down during pyrolysis at lower temperatures. In conclusion, the method used in this dissertation for causality detection could be a significant mining tool to comprehensively understand the behavior of the compounds present in a complex system such as biomass or bitumen conversion and in turn, allow the process to be controlled in a more precise way and the development of more targeted applications.

6.2. FUTURE WORKS

This work highlights some directions for future research. One possible direction could be employing an alternate strategy for self-modeling curve resolution (SMCR) by Particle swarm optimization (PSO) for computing better initial estimates to search concentration profiles or pure spectra of a multi-mixture solution. The literature indicates that SMCR using PSO is less sensitive to a local minimum in SMCR and it can be a new effective tool for curve resolution analysis.

Another possible route can be led by developing an algorithm for data fusion by integrating data provided from multi-spectroscopy techniques such as FTIR,^{1}H NMR, UV Vis, and Near IR to get a complete characterization of the process. Additionally, a mid-level data fusion[1] seems to be a useful technique not only to extract useful features but also to eliminate the unnecessary variables. Moreover, exploring some other statistical data fusion methods such as Kalman filtering[2] could be optimal under specific conditions. This technique model all of the events as probabilities, and typically has several parameters and a priori probabilities for false measurements and detection errors that are often difficult to obtain in a low-level data fusion method.

6.3. REFERENCES

1. Silvestri M, Elia A, Bertelli D, et al. Chemometrics and Intelligent Laboratory Systems A mid level data fusion strategy for the Varietal Classi fi cation of Lambrusco PDO wines. *Chemom Intell Lab Syst*. 2014;137:181-189. doi:10.1016/j.chemolab.2014.06.012

2. Zheng Z, Qiu H, Wang Z, Luo S, Lei Y. Data fusion based multi-rate Kalman filtering with unknown input for on-line estimation of dynamic displacements. *Measurement*. 2019;131:211-218. doi:10.1016/j.measurement.2018.08.057

6.4. WORKS SITED

1. A. Klein. Sensor and Data Fusion Concepts and Applications. SPIE Opt Eng Press Bellingham WA,. 1999.

2. Abella L, Nanbu S, Fukuda K, May R. A theoretical study on Levoglucosan pyrolysis reactions yielding aldehydes and a ketone in biomass. Kyushu University. 2007;67(2):67-74.

3. Abidi N, Manike M. X-ray diffraction and FTIR investigations of cellulose deposition during cotton fiber development. SAGE. 2018;(3). doi:10.1177/0040517516688634.

4. Agarwal V, Dauenhauer PJ, Huber GW AS. Ab initio dynamics of cellulose pyrolysis: nascent decomposition pathways at 327 and 600 degrees c. J Am Chem Soc. 2012;134(36):14958–72.

5. Alon U. Network motifs: Theory and experimental approaches. Nat Rev Genet. 2007;8(6):450-461. doi:10.1038/nrg2102.

6. Anca-Couce A. Reaction mechanisms and multi-scale modelling of lignocellulosic biomass pyrolysis. Prog Energy Combust Sci. 2016;53(2016):41-79. doi:10.1016/j.pecs.2015.10.002.

7. Andrew A. Peterson, Frédéric Vogel, Russell P. Lachance, Morgan Fröling, Michael J. Antal J and JWT. Thermochemical biofuel production in hydrothermal media: A review of sub- and supercritical water technologies. Energy and Envirometal Sicence. 2008;(1):32-65.

8. Ankerst M, Breunig MM, Kriegel H-P, Sander J. Optics: Ordering points to identify the clustering structure. ACM Sigmod Rec. 1999:49-60. doi:10.1145/304182.304187.

9. AntalJr. MJ. Biomass Pyrolysis: A Review of the Literature Part 1-Carbohydrate Pyrolysis. Adv Sol Energy. 1983:61-111.

10. Awa K, Okumura T, Shinzawa H, Otsuka M, Ozaki Y. Self-modeling curve resolution (SMCR) analysis of near-infrared (NIR) imaging data of pharmaceutical tablets. Anal Chim Acta. 2008;619(1):81-86. doi:10.1016/j.aca.2008.02.033.

11. Ayari I HJ. A framework for multi-sensor data fusion. Proc IEEE Symp Emerg Technol Fact Autom. 1995;2:51-59.

12. B. Ma, P.J. Gemperline, E. Cash, M. Bosserman E. Characterizing batch reactions with in situ spectroscopic measurements, calorimetry and dynamic modeling. J Chemom. 2003;17:470–479.

13. Baghbani G, Eskandari F. Calculating the required cash in bank branches: a Bayesian-data mining approach. Neural Comput Appl. 2018;30(9):2831-2841.

doi:10.1007/s00521-017-2888-9.

14. Balat M, Balat M, Kirtay E BH. Main routes for the thermo-conversion of biomass into fuels and chemicals. Part 1: pyrolysis systems. Energy Convers Manag. 2009;50(12):3147–57.

15. Bassbasi, M.; De Luca, M.; Ioele, G.; Oussama, A.; Ragno G. Prediction of the Geographical Origin of Butters by Partial Least Square Discriminant Analysis (PLS-DA) Applied to Infrared Spectroscopy (FTIR) Data. J Food Compos Anal. 2014;33(2):210-215.

16. Bassilakis R, Carangelo RM, Wojtowicz MA. TG-FTIR analysis of biomass pyrolysis. Fuel. 2001;80(12):1765-1786.

17. Bassilakis R, Carangelo RM, Wojtowicz MA. TG-FTIR analysis of biomass pyrolysis. Fuel. 2001;80(12):1765-1786.

18. Beaumont M RB. The Bayesian revolution in genetics. Nat Rev Genet. 2004;(4):251–261.

19. Bennett NM, Helle SS, Duff SJB. Extraction and hydrolysis of levoglucosan from pyrolysis oil. Bioresour Technol. 2009;100(23):6059-6063. doi:10.1016/j.biortech.2009.06.067

20. Beste A BA. Kinetic Analysis of the phenyl-shift reaction in b-O-4 lignin model compounds: a computational study. J Org Chem. 2011;76:2195–203.

21. Beyramysoltan S, Abdollahi H, Rajkó R. Newer developments on self-modeling curve resolution implementing equality and unimodality constraints. Anal Chim Acta. 2014;827:1-14. doi:10.1016/j.aca.2014.03.019

22. Binder JB, Gray MJ, White JF, Zhang ZC, Holladay JE. Reactions of lignin model compounds in ionic liquids. Biomass and Bioenergy. 2009;33(9):1122-1130. doi:10.1016/j. biombioe.2009.03.006.

23. Blackman SS. Association and fusion of multiple sensor data," in Multitarget-Multisensor: Tracking Advanced Applications. Artech House. 1990:187–217.

24. Blanco M, Castillo M, Peinado A, Beneyto R. Application of multivariate curve resolution to chemical process control of an esterification reaction monitored by near-infrared spectroscopy. Appl Spectrosc. 2006;60(6):641-647. doi:10.1366/000370206777670710.

25. Boblete O. Hydrothermal degradation of polymers derived from plants, progress in polymer science. Polym Sci. 1994;19(799).

26. Bonnefont C, Guerra A, Théron L, Molette C, Canlet C, Fernandez X. Metabolomic study of fatty livers in ducks : Identification by 1 H-NMR of metabolic markers associated with technological quality. 2011:1542-1552.

27. Bonnefont CMD, Guerra A, Théron L, Molette C, Canlet C, Fernandez X. Metabolomic study of fatty livers in ducks: Identification by1H-NMR of metabolic markers associated with technological quality. Poult Sci. 2014;93(6):1542-1552. doi:10.3382/ps.2013-0354.

28. Borràs, E.; Ferré, J.; Boqué, R.; Mestres, M.; Aceña, L.; Busto O. Data Fusion Methodologies for Food and Beverage Authentication and Quality assessment-A Review. Anal Chim Acta. 2015;891:1-14.

29. Bradbury AGW, Sakai Y, Shafizadeh F. A kinetic model for pyrolysis of cellulose. J Appl Polym Sci. 1979;23(11):3271-3280.

30. Britt PF, Buchanan AC, Cooney MJ MD. Flash vacuum pyrolysis of methoxy-substituted lignin model compounds. J Org Chem. 2000;65(5):1376-1389.

31. Cao F, Schwartz TJ, McClelland DJ, Krishna SH, Dumesic JA, Huber GW. Dehydration of cellulose to levoglucosenone using polar aprotic solvents. Energy Environ Sci. 2015;8(6):1808-1815. doi:10.1039/c5ee00353a.

32. Castanedo F. A review of data fusion techniques. Sci World J. 2013;2013. doi:10.1155/2013/704504.

33. Chalmers JM. Spectroscopy in Process Analysis. Sheff Acad Press. 2000.

34. Charles, M. Beck I. Classical analysis. A look at the past, present, and future. Anal Chem. 1994;66(4):224-239.

35. Cherubini F, Strømman AH. Principles of biorefining. Biofuels. 2011:3-24. doi:10.1016/B978-0-12-385099-7.00001-2.

36. Choi YS, Singh R, Zhang J, et al. Pyrolysis reaction networks for lignin model compounds: unravelling thermal deconstruction of β-O-4 and α-O-4 compounds. Green Chem. 2016;18(6):1762-1773. doi:10.1039/C5GC02268A.

37. Collard FX, Blin J. A review on pyrolysis of biomass constituents: Mechanisms and composition of the products obtained from the conversion of cellulose, hemicelluloses and lignin. Renew Sustain Energy Rev. 2014;38:594-608. doi:10.1016/j.rser.2014.06.013

38. Constantinou DA, Álvarez-galván MC, Luis J, Fierro G, Efstathiou AM. Applied catalysis B: environmental low-temperature conversion of phenol into CO, CO2 and H2 by steam reforming over La-containing supported Rh catalysts. Applied Catal B,

Environ. 2012;117-118:81-95. doi:10.1016/j.apcatb.2012.01.005.

39. Cooke E, Savage R, Kirk P, Darkins R WD. Bayesian hierarchical clustering for microar- ray time series data with replicates and outlier measurements. BMC Bioinformatics. 2011;12(399).

40. Cremer F, den Breejes E KS. Sensor fusion for anti-personnel land mines detection. Proc 3rd eurofusion Conf. 1998;4:63-70.

41. Crocker M. Thermochemical Conversion of Biomass to Liquid Fuels and Chemicals. Cambridge R Soc Chem. 2010.

42. D. Bylund, R. Danielsson KEM. Peak purity assessment in liquid chromatography-mass spectrometry. J Chromatogr. 2001:43-52.

43. D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi and JS-V. Handbook of chemometrics and qualimetrics: Part A. Chem Inf Comput Sci. 1998;38(6):1234–1254.

44. Darbeau RW. Nuclear Magnetic Resonance (NMR) spectroscopy: A review and a look at its use as a probative tool in deamination chemistry. Applied Spectroscopic Reviews. 2006;4928. doi:10.1080/05704920600726175.

45. David Lee Hall SAHM. Mathemathical Techniques in Multisensor Data Fusion. ARTEC House, INC; 2004.

46. Demirbaş A. Mechanisms of liquefaction and pyrolysis reactions of biomass. Energy Convers Manag. 2000;41(6):633-646. doi:10.1016/S0196-8904(99)00130-2.

47. Demirbas A. Pyrolysis of municipal plastic wastes for recovery of gasoline-range hydrocarbons. J Anal Appl Pyrolysis. 2004;72(1):97-102. doi:10.1016/j.jaap.2004.03.001.

48. DiLeo GJ, Neff ME, Savage PE. Gasification of guaiacol and phenol in supercritical water. Energy and Fuels. 2007;21(4):2340-2345. doi:10.1021/ef070056.

49. Dong J, Zhuang D, Huang Y, Fu J. Advances in Multi-Sensor Data Fusion: Algorithms and Applications. 2009;(1):7771-7784. doi:10.3390/s91007771.

50. Dorrestijn E, Laarhoven LJJ, Arends IWCE, Mulder P. Occurrence and reactivity of phenoxyl linkages in lignin and low rank coal. J Anal Appl Pyrolysis. 2000;54(1):153-192. doi:10.1016/S0165-2370(99)00082-0.

51. Dote YS, Sawayama SY, Inoue S, Minowa T. Recovery of liquid fuel from hydrocarbon-rich microalgae by thermochemical liquefaction. Sci dircet. 1994;73(12):1855-1857.

52. Dote YS, Sawayama SY, Inoue S, Minowa T. Recovery of liquid fuel from hydrocarbon-rich microalgae by thermochemical liquefaction. Sci dircet. 1994;73(12):1855-1857.

53. Dunteman BGH. Principal Components Analysis. SAGD publications, Inc; 1976.

54. E. Waltz JL. Multisensor Data Fusion. Artech House, Norwood MA. 1990.

55. E.J. Soltes TJE. Pyrolysis, in: I.S. Goldstein (Ed.), Organic Chemicals from Biomass. CRC Press Boca Rat. 1981:63-99.

56. Edwards I, Gross XE, Lowden DW SP. Fusion of NDT data. Br J NDT. 1993;35(12):710-713.

57. Elliott DC, Biller P, Ross AB, Schmidt AJ, Jones SB. Hydrothermal liquefaction of biomass: Developments from batch to continuous process. Bioresour Technol. 2015;178:147-156. doi:10.1016/j.biortech.2014.09.132.

58. Ester, M., Kriegel, H.P., Sander, J., Xu X. A Density-based algorithm for discovering clusters in large spatial databases with noise. AAAI Press. 1996:226-231.

59. Everitt, Brian, Landau S, Leese M SD. Cluster Analysis. Wiley Series in Probablity and Statistics; 2001.

60. F. Murtagh P. Contreras, Methods of hierarchical clustering. Department of Computer Science, Royal Holloway, University of London. 2011:CoRR abs/1105.0121.

61. Faltin F, Kenett R. Bayesian Networks. Encycl Stat Qual Reliab. 2007;1(1):4. doi:10.1002/wics.48.

62. Faulon JL, Sault AG. Stochastic Generator of Chemical Structure. 3. Reaction Network Generation. J Chem Inf Comput Sci. 2001;41(4):894-908. doi:10.1021/ci000029m.

63. Feng W, van der Kooi HJ, de Swaan Arons J. Phase equilibria for biomass conversion processes in subcritical and supercritical water. Chem Eng J. 2004;98(1):105-113.

64. Ferdosian CX and F. Conversion of lignin into bio-based chemicals and materials, green chemistry and sustainable technology. Springer-Verlag GmbH Ger. 2017;DOI 10.100.

65. Fernández C, Callao MP, Larrechi MS. Talanta UV – visible-DAD and 1 H-NMR spectroscopy data fusion for studying the photodegradation process of azo-dyes using MCR-ALS. Talanta. 2013;117:75-80. doi:10.1016/j.talanta.2013.08.004.

66. Fraley, C. and Raftery A. Model-Based Clustering, Disccriminant Analysis, and Density Estimation. Am Statitical Assoc. 2002;97(458):611-631.

67. Fraley, C. and Raftery A. Model-Based Clustering, Disccriminant Analysis, and Density Estimation. Am Statitical Assoc. 2002;97(458):611-631.

68. Francavilla M, Manara P, Kamaterou P, Monteleone M, Zabaniotou A. Cascade approach of red macroalgae Gracilaria gracilis sustainable valorization by extraction of phycobiliproteins and pyrolysis of residue. Bioresour Technol. 2014;184:305-313. doi:10.1016/j.biortech.2014.10.147.

69. Fränti VHCKK. Improving K-Means by Outlier Removal. In: Scandinavian Conference on Image Analysis; 2005:978-987.

70. Friedman, Nir, Nachman, Iftach. Peer D. Learning Bayesian Network Structure from Massive Datasets: The "Sparse Candidate" Algorithm. In: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence. ; 1999:206-215.

71. Friston K, Moran R, Seth AK. Analysing connectivity with Granger causality and dynamic causal modelling. Curr Opin Neurobiol. 23(2):172-178. doi:10.1016/j.conb.2012.11.010.

72. Frühwirth-Schnatter S KS. Model-based clustering of multiple time series. J Bus Econ Stat. 2008;26:78–89.

73. Gai C, Zhang Y, Chen W-T, Zhang P DY. An investigation of reaction pathways of hydrothermal liquefaction using Chlorella pyrenoidosa and Spirulina platensis. Energy Convers Manag. 2015;96:330-339.

74. Gemperline PJ. Computation of the Range of Feasible Solutions in Self-Modeling Curve Resolution Algorithms. Anal Chem. 1999;71(23):5398–5404.

75. Glover, F., Laguna M. Tabu Search. Kluwer Academic Publishers; 1998.

76. Guadix S, Meenakshisundaram M. Review on catalytic cleavage of C – C inter-unit linkages in lignin model compounds: Towards lignin depolymerisation. Top Catal. 2018;61(3):183-198. doi:10.1007/s11244-018-0909-2.

77. Guo S, Wu J, Ding M FJ. Uncovering interactions in the frequence domain. PLoS Comput Biol. 2008;4(5).

78. H. Wang, Y. Zhao, C. Wang, Y. Fu QG. Theoretical study on the pyrolysis process of lignin dimer model compounds. Acta Chim Sin. 2009;9:893-900.

79. Hammer HL, Yazidi A, Oommen BJ. "Anti-Bayesian" flat and hierarchical clustering using symmetric quantiloids. Inf Sci (Ny). 2017;418-419(August 2016):495-512. doi:10.1016/j.ins.2017.08.017.

80. Heckerman D. A tutorial on learning with Bayesian networks, in Learning in Graphical, M.J. Models. MIT Press Cambridge. 1997;1:79-119.

81. Heller K a., Ghahramani Z. Bayesian hierarchical clustering. Proc 22nd Int Conf Mach Learn. 2005:297-304. doi:10.1145/1102351.1102389.

82. Heller K GZ. Bayesian hierarchical clustering. . ICML '05 Proc 22nd Int Conf Mach Learn. 2005:297–304.

83. Hosoya T, Kawamoto H SS. Cellulose-hemicellulose and Cellulose-lignin Interactions in Wood Pyrolysis at Gasification Temperature. J Anal Appl Pyrolysis. 2007;80:118–25.

84. Hosoya T, Kawamoto H, Saka S. Different pyrolytic pathways of levoglucosan in vapor- and liquid / solid-phases. J Anal Appl Pyrolysis. 2008;83:64-70. doi:10.1016/j.jaap.2008.06.008.

85. Huang J, Liu C, Tong H, Li W, Wu D. A density functional theory study on formation mechanism of CO, CO2and CH4 in pyrolysis of lignin. Comput Theor Chem. 2014;1045:1-9. doi:10.1016/j.comptc.2014.06.009.

86. Huang X, Cheng D, Fengqiu Chen, Zhan X. Reaction pathways of hemicellulose and mechanism of biomass pyrolysis in hydrogen plasma: A density functional theory study. Renew Energy. 2016;96:490-497.

87. IEA. CO2 Emissions from Fuel Combustion 2016. IEA; 2016. doi:10.1787/co2_fuel-2016-en.

88. J Jaumot, A de Juan RT. MCR-ALS GUI 2.0: New Features and Applications. Chemom Intell Lab. (140):1-12.

89. J P. Causality: Models, Reasoning, and Inference. Cambridge Univ Press. 2000.

90. Jackson E, Davy M, Doucet A FW. Bayesian Unsupervised Signal Classification by Dirichlet Process Mixtures of Gaussian Processes. In: IEEE International Conference on Acoustics, Speech and Signal Processing. ; 2007.

91. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF GMA. Bayesian Networks Approach for Predicting Protein-Protein Interactions from Genomic Data. Science (80-). 2003:302: 449.

92. Jarvis MW, Daily JW, Carstensen H, et al. Direct detection of products from the pyrolysis of 2-Phenethyl Phenyl Ether. J. Phys. Chem. 2011:428-438.

93. Jia H, Li Y, Dong B, Ya H. An Improved Tabu Search Approach to Vehicle Routing Problem. Procedia - Soc Behav Sci. 2013;96(Cictp):1208-1217.

doi:10.1016/j.sbspro.2013.08.138.

94. Jiang J, Ozaki Y. Self -modeling Curve Resolution (SMCR): Principals, Techniques, and Applications. Ind Eng Chem Res. 2017;4928(October). doi:10.1081/ASR-120014359.

95. Jiang JH, Liang Y, Ozaki Y. Principles and methodologies in self-modeling curve resolution. Chemom Intell Lab Syst. 2004;71(1):1-12. doi:10.1016/j.chemolab.2003.07.002.

96. Jiang J-H, Liang Y-Z OY. On simplex-based method for self-modeling curve resolution of two-way data. ChemomIntell Lab Syst. 2003;65:51-65.

97. Jiang T, Ren G, Zhao X. Evacuation Route Optimization Based on Tabu Search Algorithm and Hill-Climbing Algorithm. Procedia - Soc Behav Sci. 2013;96(Cictp):865-872. doi:10.1016/j.sbspro.2013.08.098

98. Jin H, Zhao X, Wu Z, Cao C, Guo L. Supercritical water synthesis of nano-particle catalyst on TiO 2 and its application in supercritical water gasification of biomass. J Exp Nanosci. 2017;12(1):72-82. doi:10.1080/17458080.2016.1262066

99. K. Ehara, S. Saka HK. Characterization of the lignin-derived products from wood as treated in supercritical water, J. Wood Sci. 48 (2002) 320–325. J Wood Sci. 2002;48:320–325.

100. Kabir G, Hameed BH. Recent progress on catalytic pyrolysis of lignocellulosic biomass to high-grade bio-oil and bio-chemicals. Renew Sustain Energy Rev. 2017.

101. Kadiroğlu P. FTIR spectroscopy for prediction of quality parameters and antimicrobial activity of commercial vinegars with chemometrics. J Sci Food Agric. 2018;98(11):4121-4127. doi:10.1002/jsfa.8929.

102. Kantardzic M. Data Mining : Concepts, Models, Methods, and Algorithms. Wiley-IEEE Press; 2011.

103. Karimi K, Kheradmandinia S, Taherzadeh MJ. Conversion of rice straw to sugars by dilute-acid hydrolysis. Biomass and Bioenergy. 2006;30(3):247-253.

104. Keller, H.R., Massart, D.L., and De Beer JO. Window evoloving factor analysis for assesment of peak homogeneity in liquid chromatography. Chem. 1993;65(4):471-474.

105. Kim S, Chmely SC, Nimos MR, Bomble YJ, Foust TD, Paton RS BG. Computational study of bond dissociation enthalpies for a large range of native and modifi ed lignins. J Phys Chem Lett. 2011;2(22):2846–2852.

106. Klein MT VP. Model pathways in lignin thermolysis 1. Phenethyl phenyl ether. Ind

Eng Chem Fund. 1983;22(1):35-45.

107. Koller D FN. Probabilistic graphical models: principles and techniques. MIT Press Cambridge. 2009.

108. Kruse A, Gawlik A. Biomass conversion in water at 330-410 °C and 30-50 MPa. Identification of key compounds for indicating different chemical reaction pathways. Ind Eng Chem Res. 2003;42(2):267-279. doi:10.1021/ie0202773

109. Kruse A, Krupka A, Schwarzkopf V, Gamard C, Henningsen T. Influence of proteins on the hydrothermal gasification and liquefaction of biomass. 1. Comparison of different feedstocks. Ind Eng Chem Res. 2005;44(9):3013-3020. doi:10.1021/ie049129.

110. L. Lin, M. Yoshioka, Y. Yao NS. Liquefaction mechanism of lignin in the presence of phenol at elevated temperature without catalysts. Studies on β-O-4 lignin model compound III. Multi-condensation. Holzforschung. 1997;51:333–337.

111. Lange J. Lignocellulose conversion: an introduction to chemistry, process and economics. 2007;1(1):39–48.

112. Lee H, Dharma D, Hsie B, Tang G, Chakrabarti D, De Klerk A, Prasad V. Biomass conversion through hydrous pyrolysis. In: Innovation, Industry and Internationalization. London,ON; 2011:348-349.

113. Lee SM, Abbott PA. Bayesian networks for knowledge discovery in large datasets: Basics for nurse researchers. J Biomed Inform. 2003;36(4-5):389-399. doi:10.1016/j.jbi.2003.09.022.

114. Li G, Torraca G, Jing W, Wen Z qing. Applications of FTIR in identification of foreign materials for biopharmaceutical clinical manufacturing. Vib Spectrosc. 2009;50(1):152-159. doi:10.1016/j.vibspec.2008.10.016.

115. Li S, Lyons-Hart J, Banyasz J, Shafer K. Real-time evolved gas analysis by FTIR method: An experimental study of cellulose pyrolysis. Fuel. 2001;80(12):1809-1817. doi:10.1016/S0016-2361(01)00064-3.

116. Li, Y.; Zhang, J.; Li, T.; Liu, H. G.; Wang YZA. Comprehensive and Comparative Study of Wolfiporia Extensa Cultivation Regions by Fourier Transform Infrared Spectroscopy and Ultra-Fast Liquid Chromatography. PLoS One. 2016;11(12).

117. Li, Y.; Zhang, J.; Liu, H. G.; Jin, H.; Wang, Y. Z.; Li T. Discrimination of Storage Periods for Macrocybe Gigantea (Massee) Pegler & Lodge Using UV Spectral Fingerprints. Czech J Food Sci. 2015;33(5):441-448.

118. Libra JA, Ro KS, Kammann C, et al. Hydrothermal carbonization of biomass

residuals : a comparative review of the chemistry , processes and applications of wet and dry pyrolysis Hydrothermal carbonization of biomass residuals : a comparative review of the chemistry , processes and applicati. 2014;7269. doi:10.4155/bfs.10.81.

119. Lievens, C., Mourant, D., He, M., Gunawan, R., Li, X., and Li C-Z (2011). An FT-IR spectroscopic study of carbonyl functionalities in bio-oils. Fuel 90. Fuel. 2011;90(3417).

120. Liu WJ, Jiang H, Yu HQ. Thermochemical conversion of lignin to functional materials: a review and future directions. Green Chem. 2015;17(11):4888-4907. doi:10.1039/c5gc01054c.

121. Liu X, Zu X, Liu Y, Sun L. Conversion of waste water hyacinth into high-value chemicals by iron (III) chloride under mild conditions. BioResources. 2018;13(2):2293-2303. doi:10.15376/biores.13.2.2293-2303

122. Llinas DLH and J. An introduction to multisensor data fusion. Proc IEEE. 1997;85(1):6-23.

123. Long Y, Zhou H, Meng A, Li Q, Zhang Y. Interactions among biomass components during co-pyrolysis in (macro)thermogravimetric analyzers. Korean J Chem Eng. 2016;33(9):2638-2643. doi:10.1007/s11814-016-0102.

124. Lopez Barreiro D, Prins W, Ronsse F BW. Hydrothermalliquefaction (HTL) of microalgae for biofuel production: state of the artreview and future prospeects. Biomass Bioenergy. 2013;53:113–127.

125. Luo RC, Kay MG. Multisensor Integration And Fusion: Issues And Approaches. Proc SPIE. 1988;0931(August 1988):42-49. doi:10.1117/12.946646

126. Madsen RB, Zhang H, Biller P, Goldstein AH, Glasius M. Characterizing Semivolatile Organic Compounds of Biocrude from Hydrothermal Liquefaction of Biomass. Energy and Fuels. 2017;31(4):4122-4134. doi:10.1021/acs.energyfuels.7b00160.

127. Maeder M. Evolving factor analysis for the resolution of overlapping chromatographic peaks. Anal Chem. 1987;59(3):527-530.

128. Malaluan RM. A Study on Cellulose Decomposition in Subcritical and Supercritical Water. Ph.D. Dissertation, Tohoku University, Sendai, Japan, 1995.

129. Manya JJ, Velo E, Puigjaner L. Kinetics of biomass pyrolysis: a reformulated three-parallel-reactions model. Ind Eng Chem Res. 2003;42(3):434-441.

130. McLachlan, G. J and Peel DA. Finite Mixture Models. John Wiley & Sons; 2000.

131. McLaren J. The Technology Roadmap for Plant/Crop-Based Renewable Resources

2020. DIANE Publishing; 1999.

132. Mehdi Jalali-Heravi , Hadi Parastar HE-N. Self-modeling Curve Resolution Techniques Applied to Comparative Analysis of Volatile Components of Iranian Saffron from Different Regions. Anal Chim Acta. 2010;(662):143-154.

133. Minowa T, Zhen F, Ogi T. Cellulose decomposition in hot-compressed water with alkali or nickel catalyst. Journal of Super Critical Fluids. 1998; 13:253-259.

134. Mistry B. A Handbook of Spectroscopic Data Chemistry (UV, IR, PMR, 13CNMR and Mass Spectroscopy). Oxford Book Company. 2009. doi:10.1063/1.2719251.

135. Morris, S.A., Van der Veer Martens B. Mapping research specialties. Annu Rev Inf Sci Technol. 2008;42(1):213-295.

136. Muley PD, Henkel C, Abdollahi KK, Marculescu C, Dorin Boldor. A critical comparison of pyrolysis of cellulose, lignin, and pine sawdust using an induction heating reactor. Energy Convers Manag. 2016;117:273–280.

137. Nagarajan R, Scutari M, Lèbre S. Bayesian Networks in R with Applications in Systems Biology. Springer science.; 2013. doi:10.1007/978-1-4614-6446-4.

138. Nagarajan R, Scutari M, Lèbre S. Bayesian Networks in R.; 2013. doi:10.1007/978-1-4614-6446-4.

139. Narayan S, Muldoon J, Finn M, Fokin VV, Kolb HC SK. On water: unique reactivity of organic compounds in aqueous suspension. Angew Chem Int Ed. 2005;44:3275–9.

140. Neapolitan RE. Learning Bayesian Networks. Upper Saddle River, NJ: Pearson Prentice Hall; 2004.

141. Nia VP, Davison AC. High-Dimensional Bayesian Clustering with Variable Selection: The R Package bclust. J Stat Softw. 2012;47(5):1-22. doi:http://dx.doi.org/10.18637/jss.v047.i05.

142. Nimlos MR, Evans RJ. Levoglucosan Pyrolysis. Fuel Chem Div Prepr. 2002;47(1):393.

143. Nybacka L. FTIR spectroscopy of glucose FTIR spectroscopy of glucose. 2016.

144. Nygård HS, Olsen E. Review of thermal processing of biomass and waste in molten salts for production of renewable fuels and chemicals. Int J Low-Carbon Technol. 2012;7(4):318-324. doi:10.1093/ijlct/ctr045.

145. Paine III JBP, Pithawalla YB, Naworal JD. Carbohydrate pyrolysis mechanisms from isotopic labeling Part 2 . The pyrolysis of D -glucose : General disconnective analysis

and the formation of C 1 and C 2 carbonyl compounds by electrocyclic fragmentation mechanisms. J Anal Appl Pyrol. 2008;82(1):10-41. doi:10.1016/j.jaap.2008.01.002

146. Pan Y, Kong SC. Predicting effects of operating conditions on biomass fast pyrolysis using particle-level simulation. Energy and Fuels. 2017;31(1):635-646. doi:10.1021/acs.energyfuels.6b02445.

147. Pang S. Advances in thermochemical conversion of woody biomass to energy, fuels and chemicals. Biotechnol Adv. 2018;(November):0-1. doi:10.1016/j.biotechadv.2018.11.004.

148. PE S. A perspective on catalysis in sub- and supercritical water. J Supercrit Fluids. 2009;47(3):407-414.

149. Pedersen TH, Jensen CU, Sandström L, Rosendahl LA. Full characterization of compounds obtained from fractional distillation and upgrading of a HTL biocrude. Appl Energy. 2017;202:408-419. doi:10.1016/j.apenergy.2017.05.167

150. Peters JF, Banks SW, Bridgwater A V., Dufour J. A kinetic reaction model for biomass pyrolysis processes in Aspen Plus. Appl Energy. 2017;188:595-603. doi:10.1016/j.apenergy.2016.12.030

151. Piskorz J, Radlein D SD. On the mechanism of the rapid pyrolysis of cellulose. Anal Appl Pyrolysis. 1986;9:121-137.

152. Pourret O. Bayesian Networks: A Practical Guide to Applications. Chichester, West Sussex, Eng: John Wiley; 2008.

153. Pouwels, A.D, Eijkel , G.B, and Boon JJ. Pyrolysis. J Anal Appl. 1989;14(4):237–280.

154. R. Tauler, F.C. Sanchez DLM. Validation of alternating least squares multivariate curve resolution for chromatographic resolution and quantitation. Trends Anal Chem. 1996;15:279–286.

155. Rahimdoust Mojdehi N, Sawall M, Neymeyr K, Abdollahi H. Investigating the effect of flexible constraints on the accuracy of self-modeling curve resolution methods in the presence of perturbations. J Chemom. 2016;30(5):252-267. doi:10.1002/cem.2787.

156. Rass E-J, McKay B RG. Understanding catalytic biomass conversion through data mining. Top Catal. 2010;53:1202-1208.

157. Rissanen J. Information and Complexity in Statistical Models. Springer Publishing Company; 2007.

158. Robert M. Silverstein, Francis X. Webster DJK. Spectrometric identification of organic compounds. J Mol Struct. 2005:512. doi:10.1016/0022-2860(76)87024-X.

159. RóbertRajkó. Additional knowledge for determining and interpreting feasible band boundaries in self-modeling/multivariate curve resolution of two-component systems. Anal Chim Acta. 2010;661(2):129-132.

160. Ronaldus R. Bayesian belief networks: from construction to inference. 1995.

161. Ross, J.; Schreiber, I.; Vlad MO. Determination of Complex Reaction Mechanisms: Analysis of Chemical, Biological, and Genetic Networks. Oxford University Press: New York; 2006.

162. Roussel S, Bellon-Maurel VE, Roger JM, Grenier P. Authenticating white grape must variety with classification models based on aroma sensors, FT-IR and UV spectrometry. J Food Eng. 2003;60(4):407-419. doi:10.1016/S0260-8774(03)00064-5.

163. S. Kang, X. Li, J. Fan JC. Classified separation of lignin hydrothermal liquefied products. Ind Eng Chem Res. 2011;50:11288–11296.

164. Sanchez, F.C.; van den Bogaert, B.; Rutan, S.C.; Massart DL. Orthogonal projection approach applied to peak purity assessment. Chem. 68(1):139-171.

165. Santilli C, Makarov IS, Fristrup P, Madsen R. Dehydrogenative synthesis of Carboxylic acids from primary alcohols and hydroxide catalyzed by a Ruthenium N - heterocyclic carbene complex. JOC. 2016;. doi:10.1021/acs.joc.6b02105.

166. Sasaki M, Fang Z, Fukushima Y, Adschiri T, Arai K. Dissolution and hydrolysis of cellulose in subcritical and supercritical water. Ind Eng Chem Res. 2000;39(8):2883-2890.

167. Sattari F, De Klerk A, Prasad V. Application of data combination and data mining techniques to investigate the chemistry of cellulose and lignin derivatives in hydrous pyrolysis. Submiited to I&EC. 2019.

168. Sattari F, Tefera A, Sivaramakrishnan K, De Klerk A, Prasad V. Postulating Pseudo-reaction Networks for the Conversion of Levoglucosan in Hydrous Pyrolysis Using Spectroscopic Data & Self-Modeling Multivariate Curve Resolution. Submitted to I & EC. 2019.

169. Savage RS, Heller K, Xu Y, et al. R / BHC : fast Bayesian hierarchical clustering for microarray data. 2009;9:1-9. doi:10.1186/1471-2105-10-242.

170. Schreuder LT, Hopmans EC, Stuut JBW, Sinninghe Damsté JS, Schouten S. Transport and deposition of the fire biomarker levoglucosan across the tropical North Atlantic Ocean. Geochim Cosmochim Acta. 2018;227:171-185. doi:10.1016/j.gca.2018.02.020.

171. Scutari M, Vitolo C, Tucker A. Learning Bayesian Networks from Big Data with Greedy Search: Computational Complexity and Efficient Implementation. 2018. http://arxiv.org/abs/1804.08137.

172. Sequaris J-ML, Koglin E. Evolving Factor Analysis for the Resolution of Overlapping Chromatographic Peaks. Anal Chem. 1987;59(6):527-530. doi:Doi 10.1021/Ac00130a035.

173. Shafizadeh F. Introduction to pyrolysis of biomass. J Anal Appl Pyrolysis. 1982;3(4):283-305.

174. Sharma S. Applied Multivariate Techniques. John Wiley & Sons, Inc.; 1996.

175. Shen D, Xiao R, Gu S, Zhang H. The Overview of Thermal Decomposition of Cellulose in Lignocellulosic Biomass. INTECH. 2013. doi: 10.5772/51883.

176. Shen D, Xiao R, Luo K. The pyrolytic behavior of cellulose in lignocellulosic biomass : a review. RSC Adv. 2011:1641-1660. doi:10.1039/c1ra00534k.

177. Shen DK GS. The mechanism for thermal decompostion of cellulose and its main products. Bioresour Technol. 2009;100:6496-6504.

178. Shen, T. T.; Zou, X. B.; Shi, J. Y.; Li, Z. H.; Huang, X. W.; Xu, Y. W.; Chen W. Determination Geographical Origin and Flavonoids Content of Goji Berry Using Near-Infrared Spectroscopy and Chemometrics. Food Anal Methods. 2016;9(1):68–79.

179. Shinzawa H, Jiang JH, Iwahashi M, Noda I, Ozaki Y. Self-modeling curve resolution (SMCR) by particle swarm optimization (PSO). Anal Chim Acta. 2007;595(1-2 SPEC. ISS.):275-281. doi:10.1016/j.aca.2006.12.004.

180. Silvestri M, Elia A, Bertelli D, et al. Chemometrics and Intelligent Laboratory Systems A mid level data fusion strategy for the Varietal Classi fi cation of Lambrusco PDO wines. Chemom Intell Lab Syst. 2014;137:181-189. doi:10.1016/j.chemolab.2014.06.012.

181. Simoneit BRT, Schauer JJ, Nolte CG, et al. Levoglucosan, a tracer for cellulose in biomass burning and atmospheric particles. Atmos Environ. 1999;33(2):173-182. doi:10.1016/S1352-2310(98)00145-9.

182. Sirinukunwattana K, Savage RS, Bari MF, Snead DRJ, Rajpoot NM. Bayesian hierarchical clustering for studying cancer gene dxpression Data with unknown statistics. PLoS One. 2013;8(10). doi:10.1371/journal.pone.0075748.

183. Smolders K, Van de Velden M BJ. Operating parameters for the bubbling fluidized (BFB) and circulating fluidized bed (CFB) processing of biomass, In: Proceedings of ‘‘use of renewables.’ Achema: Frankfurt. 2006;(Paper 1030).

184. Soranzo N, Altafini C. ERNEST: A toolbox for chemical reaction network theory. Bioinformatics. 2009;25(21):2853-2854. doi:10.1093/bioinformatics/btp513.

185. Stagni A, Cuoci A, Frassoldati A, Faravelli T, Ranzi E. Lumping and Reduction of Detailed Kinetic Schemes: An Effective Coupling. *Ind Eng Chem Res*. 2014;53(22):9004-9016.

186. Staš M, Kubička D, Chudoba J, Pospíšil M. Overview of analytical methods used for chemical characterization of pyrolysis bio-oil. Energy and Fuels. 2014;28(1):385-402. doi:10.1021/ef402047y.

187. Stefanidis SD, Kalogiannis KG, Iliopoulou EF, Michailof CM, Pilavachi PA, Lappas AA. A study of lignocellulosic biomass pyrolysis via the pyrolysis of cellulose, hemicellulose and lignin. J Anal Appl Pyrolysis. 2014;105:143-150. doi:10.1016/j.jaap.2013.10.013.

188. T. Adschiri, S. Hirose, R. Malaluan KA. Noncatalytic conversion of cellulose in supercritical and subcritical water, J. Chem Eng Jpn. 1993;26(676).

189. Tadesse MG, Sha N, Vannucci M. Bayesian variable selection in clustering high-dimensional data. J Am Stat Assoc. 2005;100(470):602-617. doi:10.1198/016214504000001565.

190. Tauler R, Smllde AK, Henshaw JM, Burgess QLW, Kowalskl BR. Multicomponent Determination of Chlorinated Hydrocarbons Using a Reaction-Based Chemical Sensor . 2 . Chemical Speciation Using Multivariate Curve Resolution. 1994. doi:10.1021/ac00092a009.

191. Tefera DT, Agrawal A, Yañez Jaramillo LM, De Klerk A, Prasad V. Self-Modeling Multivariate Curve Resolution Model for Online Monitoring of Bitumen Conversion Using Infrared Spectroscopy. Ind Eng Chem Res. 2017;56(38):10756-10769. doi:10.1021/acs.iecr.7b01849.

192. Tefera DT, Yañez Jaramillo LM, Ranjan R, Li C, De Klerk A, Prasad V. A Bayesian Learning Approach to Modeling Pseudoreaction Networks for Complex Reacting Systems: Application to the Mild Visbreaking of Bitumen. Ind Eng Chem Res. 2017;56(8):1961-1970. doi:10.1021/acs.iecr.6b04437

193. Tiilikkala K, Fagernas L TJ. History and use of wood pyrolysis liquids as biocide and plant protection product history and use of wood pyrolysis liquids as biocide and plant protection product. Open Agric J. 2010;4:111-118.

194. Tompsett GA, Li N HG. Catalytic conversion of sugars to fuels. John Wiley Sons. 2011:273-279.

195. Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. Springer Sci. 2006:31-78. doi:10.1007/s10994-006-6889-7.

196. Ugi, I.; Bauer, J.; Bley, K.; Dengler, A.; Dietz A.; Fontain, E.; Gruber, B.; Herges R., Knauer, M.; Reitsam, K.; Stein N. Computer-Assisted Solution of Chemical Problems-The Historical Development and the Present State of the Art of a New Discipline of Chemistry. Chem Int Ed Engl. 1993;32:201-227.

197. Uihlein A, Schebek L. Environmental impacts of a lignocellulose feedstock biorefinery system: An assessment. Biomass and Bioenergy. 2009;33(5):793-802. doi:10.1016/j.biombioe.2008.12.001.

198. Uncu O and Ozen B. Prediction of various chemical parameters of olive oils with Fourier transform infrared spectroscopy. LWT – Food Sci Technol. 2015;63:978–984.

199. Vitasari C, Meindersma G de HA. Conceptual process design of an integrated bio-based acetic acid, glycolaldehyde, and acetol production in a pyrolysis oil-based biorefinery. Chem Eng Res. 2015;95:133–43.

200. W.H. Lawton EAS. Self modeling curve resolution. Technometrics. 1971;13(3):617-633.

201. W.S. Mok, M.J. Antal Jr., GV. Productive and parasitic pathways in dilute acid-catalyzed hydrolysis of cellulose,Ind. Eng Chem. 1992;32(94).

202. Wahyudiono, M. Sasaki MG. Conversion of biomass model compound under hydrothermal conditions using batch reactor. Fuel. 2009:1656–1664.

203. Walker TW, Motagamwala AH, Dumesic JA, Huber GW. Fundamental catalytic challenges to design improved biomass conversion technologies. J Catal. 2018;369:518-525. doi:10.1016/j.jcat.2018.11.028.

204. Wang J, Wei Q, Zheng J, Zhu M. Effect of pyrolysis conditions on levoglucosan yield from cotton straw and optimization of levoglucosan extraction from bio-oil. J Anal Appl Pyrolysis. 2016;122:294-303. doi:10.1016/j.jaap.2016.09.013.

205. Wang S, Guo X, Wang K, Luo Z. Influence of the interaction of components on the pyrolysis behavior of biomass. J Anal Appl Pyrolysis. 2011;91(1):183-189. doi:10.1016/j.jaap.2011.02.006.

206. Wang S, Zhongyang L. Pyrolysis of Biomass. Science Press Beijing; 2017.

207. Wang Y, He T, Liu K, Wu J, Fang Y. From biomass to advanced bio-fuel by catalytic pyrolysis/hydro-processing: hydrodeoxygenation of bio-oil derived from biomass catalytic pyrolysis. Bioresour Technol. 2012;108:280-284.

208. Wang Y, Li X, Mourant D, Gunawan R, Zhang S, Li C-Z. Formation of Aromatic Structures during the Pyrolysis of Bio-oil. Energy & Fuels. 2012;26(1):241-247. doi:10.1021/ef201155e.

209. Wentzell PD, Wang J-H, Loucks LF MK. Direct optimization of self-modeling curve resolution: application to the kinetics of the permanganate—oxalic acid reaction. Can J Chem. 1998;76:1-12.

210. Wong ASY, Huck WTS. Grip on complexity in chemical reaction networks. Beilstein J Org Chem. 2017;13:1486-1497. doi:10.3762/bjoc.13.147.

211. Wong, K. C (2015). Review of Spectrometric Identification of Organic Compounds. *Journal of Chemical Education* , 92, (10), 1602-1603.

212. Wu H, Fu Q, Giles R, Bartle J. Production of mallee biomass in Western Australia: Energy balance analysis. Energy and Fuels. 2008;22(1):190-198. doi:10.1021/ef7002969.

213. Wu S, Shen D, Hu J, Xiao R, Zhang H. Pyrolysis TG-FTIR and Py-GC – MS analysis of a model compound of cellulose – glyceraldehyde. J Anal Appl Pyrolysis. 2013;101:79-85. doi:10.1016/j.jaap.2013.02.009.

214. Wu W, Kawamoto K, Kuramochi H. Hydrogen-rich synthesis gas production from waste wood via gasification and reforming technology for fuel cell application. J Mater Cycles Waste Manag. 2006;8(1):70-77. doi:10.1007/s10163-005-0138-1.

215. Xu HY, Yue ZH, Wang C, Dong K, Pang HS, Han Z. Multi-source data fusion study in scientometrics. Scientometrics. 2017;111(2):773-792. doi:10.1007/s11192-017-2290-5.

216. Xu X, Matsumura Y, Stenberg J, Antal MJ. Carbon-Catalyzed Gasification of Organic Feedstocks in Supercritical Water †. Ind Eng Chem Res. 1996;35(8):2522-2530. doi:10.1021/ie950672b.

217. Y. Z. Hydrothermal liquefaction to convert biomass into crude oil. In: Biofuels from agricultural wastes and byproducts. Hoboken, NJ Wiley-Blackwell. 2010:201-232.

218. Yang H, Yan R, Chen H, Lee DH, Zheng C. Characteristics of hemicellulose, cellulose and lignin pyrolysis. Fuel. 2007;86(12-13):1781-1788. doi:10.1016/j.fuel.2006.12.013.

219. Ye X ning, Lu Q, Jiang X yan, et al. Interaction characteristics and mechanism in the fast co-pyrolysis of cellulose and lignin model compounds: A joint experimental and theoretical study. J Therm Anal Calorim. 2017;130(2):975-984. doi:10.1007/s10973-017-6465-3.

220. Yeh TM, Dickinson JG FA et al. Hydrothermal catalytic production of fuels and chemicals from aquatic biomass. J Chem Technol Biotechnol. 2013;88(1):13-24.

221. Yu Y, Bartle J, Mendham D, Wu H. Site Variation in Life Cycle Energy and Carbon Footprints of Mallee Biomass Production in Western Australia. Energy & Fuels. 2015;29(6):3748-3752. doi:10.1021/acs.energyfuels.5b00618.

222. Yuan X, Cheng G. From cellulose fibrils to single chains: understanding cellulose dissolution in ionic liquids. Phys Chem Chem Phys. 2015;17(47):31592-31607. doi:10.1039/C5CP05744B.

223. Zhang J, Choi YS, Yoo CG, Kim TH, Brown RC, Shanks BH. Cellulose–Hemicellulose and Cellulose–Lignin Interactions during Fast Pyrolysis. ACS Sustain Chem Eng. 2015;3(2):293-301. doi:10.1021/sc500664h.

224. Zhang L, Shen C, Liu R. GC – MS and FT-IR analysis of the bio-oil with addition of ethyl acetate during storage. 2014;2(January):1-6. doi:10.3389/fenrg.2014.00003.

225. Zhang S, Ding HUA, Wang X. Research and Application of Structure Learning Algorithim. In: Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an,. ; 2003:2-5.

226. Zhang X, Tauler R. Application of multivariate curve resolution alternating least squares (MCR-ALS) to remote sensing hyperspectral imaging. Anal Chim Acta. 2013;762:25-38. doi:10.1016/j.aca.2012.11.043.

227. Zhang X, Wang T, Ma L, Chang J. Vacuum pyrolysis of waste tires with basic additives. Waste Manag. 2008;28(11):2301-2310. doi:10.1016/j.wasman.2007.10.009

228. Zhang X, Yang W, Blasiak W. Thermal decomposition mechanism of levoglucosan during cellulose pyrolysis. J Anal Appl Pyrolysis. 2012;96:110-119. doi:10.1016/j.jaap.2012.03.012

229. Zheng Z, Qiu H, Wang Z, Luo S, Lei Y. Data fusion based multi-rate Kalman filtering with unknown input for on-line estimation of dynamic displacements. Measurement. 2019;131:211-218. doi:10.1016/j.measurement.2018.08.057.

230. Zou C, Denby KJ, Feng J. Granger causality vs. dynamic Bayesian network inference: A comparative study. BMC Bioinformatics. 2009;10. doi:10.1186/1471-2105-10-401.

Table 2-$A_{1\text{-}1}$. Wavenumbers in Cluster 1

609.4611	3369.3941	3394.4668	3334.6779	3288.3898	644.1772	713.6095	649.9633	3097.7421	730.9676
3415.6823	3425.3256	3363.608	3321.1772	3303.8191	3267.1743	707.8235	3218.9575	3199.6707	736.7536
640.3199	3429.183	3365.5367	3340.464	3296.1044	3265.2457	3236.3155	3224.7435	680.822	628.7478
3417.6109	3371.3227	3348.1787	3319.2485	3282.6037	669.25	3226.6722	3215.1001	**1042.3955**	732.8962
3398.3242	3373.2514	**3355.8934**	3317.3199	3294.1758	717.4669	3228.6008	3213.1715	676.9647	682.7507
3413.7536	3404.1102	3384.8235	3315.3912	3284.5324	3255.6023	3049.8163	3211.2428	705.8948	724.8249
3419.5396	3411.8249	3346.25	3338.5353	3280.6751	3238.2442	601.7464	617.1758	630.6765	646.1059
3400.2529	3409.8962	3353.9647	3309.6052	3292.2471	**2906.1729**	3247.8876	3209.3141	763.755	619.1045
3390.6095	3407.9676	3350.1073	3336.6066	**3272.9604**	659.6066	657.678	3207.3854	632.6052	738.6823
3421.4683	3406.0389	3344.3213	3311.5338	3286.4611	3263.317	771.4697	3205.4568	761.8264	648.0346
3200.2543	642.2486	2859.3912	3013.4625	3274.889	3251.7449	634.5339	3193.8847	723.2529	622.9618
3377.1088	3392.5382	3352.036	655.7493	3276.8177	667.3213	663.464	3195.8134	615.2471	702.0375
3379.0374	3367.4654	3325.0346	3299.9618	3278.7464	3257.531	767.6124	3220.8861	721.3242	613.3184
653.8206	3382.8948	3323.1059	3307.6765	661.5353	3261.3883	675.036	3222.8148	725.1816	605.6037
3375.1801	3359.7507	3332.7493	**1130.9073**	3271.0317	3259.4596	769.5411	3100.956	727.1102	740.6109
3380.9661	3431.1117	3326.9632	671.1787	715.5382	3244.0302	3234.3869	765.6837	621.0331	757.969
3388.6808	3357.822	611.3898	3301.8905	607.5324	711.6808	3230.5295	3203.5281	599.8177	684.6794
3402.1815	3361.6794	3328.8919	3298.0331	3269.103	3245.9589	3217.0288	678.8934	703.9661	722.5396
3423.397	3386.7521	3330.8206	3305.7478	709.7522	3242.1016	3232.4582	3201.5994	729.0389	686.6081
3396.3955	651.8919	3342.3926	3000.3184	3003.6736	636.4625	665.3926	603.6751	759.8977	**1400.1088**
756.0403	744.4683								

Table 2-$A_{1\text{-}2}$. Wavenumbers in Cluster 2

1051.1275	993.2673	2750.2896	2634.5691	3581.5482	2493.7759	2395.4135	3622.0504	3637.4798	1460.0065
1139.8465	2974.0158	2622.9971	2790.7918	2156.2579	2208.3321	2426.2723	1444.5771	1066.5569	3973.0691
1137.9179	1618.1578	2829.3653	2669.2853	2484.1326	1332.714	1406.0036	2281.6217	1159.1333	1539.0821
1049.1988	2964.3725	1189.9921	2678.9287	2540.0641	2325.9813	2237.2622	1847.67	3635.5511	3706.9121
923.835	2972.0872	3014.518	1562.2262	1004.8393	2092.6117	2102.255	1701.0908	1336.5713	3959.5684
1053.0562	3002.9459	2559.3508	2711.7161	2570.9229	3583.4769	3562.2615	2445.5591	1718.4488	3681.8393
921.9063	2339.482	1706.8768	1874.6715	1838.0266	3089.7363	1942.1751	1357.7867	1446.5058	3953.7824
2916.1556	1726.1635	1676.018	3571.9049	2327.9099	1984.6059	2256.549	1780.1664	1423.3617	3980.7838
1141.7752	2362.6261	2779.2197	2649.9985	1865.0281	1670.232	3070.4495	1355.8581	1868.8854	3762.8436
925.7637	3541.0461	2055.9668	2160.1153	2578.6376	1126.3458	3068.5209	1199.6354	1544.8681	3876.6354
2918.0843	2844.7947	2034.7514	1128.2745	2597.9243	2210.2608	2889.6275	2293.1938	3664.4812	3969.2118
1135.9892	2366.4834	2196.7601	2655.7846	1276.7824	3554.5468	1755.0936	2146.6145	1377.0735	3587.3343
948.9078	873.6895	2098.3977	2686.6434	1407.9323	1282.5684	3600.835	1685.6614	1220.8509	1596.9424
985.5526	2902.6549	2615.2824	2038.6088	2486.0612	1402.1462	1870.8141	1758.951	3660.6239	3938.353
929.621	887.1902	3018.3753	2709.7874	1577.6556	3087.8076	3060.8062	1317.6412	1479.2932	3718.4841
862.1174	3001.0173	2345.268	2785.0057	2538.1354	1890.1009	2451.3451	1535.2248	3633.6225	3566.1189
950.8364	1926.7457	1892.0295	2667.3566	2553.5648	1321.1419	1263.2817	1118.6311	1392.5029	3828.4186
860.1888	1132.1318	1188.0634	1963.3905	2370.3408	2424.3436	1404.0749	2403.1282	1174.5627	1525.5814
919.9777	2337.5533	1184.206	1834.1693	3579.6196	1272.9251	2079.1109	1741.5929	3662.5526	1542.9395
987.4813	3004.8746	2804.2925	1949.8898	2543.9215	3024.1614	3031.8761	1697.2334	1043.4128	3990.4272
927.6924	2900.7262	2734.8602	2088.7543	3558.4042	1008.6967	1284.4971	2216.0468	1166.848	3946.0677
983.6239	2993.3026	2744.5036	2694.3581	1812.9539	2250.763	1350.072	2385.7702	1533.2961	3699.1974
894.9049	2057.8955	2732.9315	2783.0771	2536.2068	1976.8912	2414.7003	3544.9034	3639.4085	3685.6967
2920.013	2999.0886	2727.1455	2624.9258	2551.6362	2422.415	2011.6073	1216.9935	1085.8437	3974.9978
946.9791	2904.5836	2117.6844	1818.7399	2514.9913	1178.42	3066.5922	1259.4243	1384.7882	3791.7738
2914.2269	2960.5151	2748.3609	1861.1707	3550.6895	2534.2781	3552.6181	2295.1225	3643.2658	1515.938
931.5497	2906.5122	2642.2838	1913.2449	1886.2435	2511.134	1352.0007	1342.3573	3589.2629	1324.7082
1998.1066	2962.4438	2705.9301	2119.6131	1695.3047	3074.3069	1334.6426	2432.0583	1151.4186	3809.1318
896.8336	1662.5173	2752.2183	2605.639	2939.2997	2144.6859	1317.2846	2300.9085	1481.2219	3830.3472
989.4099	2858.2954	2806.2212	1839.9553	2133.1138	1105.1304	1147.5612	1087.7723	1770.523	3961.4971
952.7651	2997.1599	2937.371	2692.4294	3575.7622	2420.4863	3058.8775	2376.1268	3641.3372	3907.4942
864.0461	1056.9135	2729.0742	2696.2867	2243.0483	1730.0209	2447.4877	1124.4171	1751.2363	3963.4258
954.6938	1851.5274	3016.4467	2518.8487	2497.6333	3078.1642	3093.5936	1296.0692	1600.7997	3895.9222
956.6225	2840.9373	877.5468	2653.8559	1278.7111	1575.7269	2254.6203	2405.0569	1014.4827	3716.5555
1034.9849	2958.5864	1946.0324	1924.817	1953.7471	2549.7075	2252.6916	1969.1765	1791.7385	1074.2716
892.9762	2021.2507	2663.4993	2189.0454	2478.3465	3082.0216	2412.7716	1257.4957	1514.0094	1463.8638
1658.6599	2842.866	2746.4323	2563.2082	1836.098	2393.4849	2399.2709	1735.8069	2163.9726	1251.7096
918.049	1616.2291	2561.2795	2893.0115	2495.7046	3028.0187	1319.2132	1799.4532	1483.1506	3782.1304
2360.6974	2995.2313	2715.5735	1849.5987	2036.6801	1569.9409	1207.3501	1957.6045	1747.379	3668.3386
1612.3717	1099.3444	2761.8617	2204.4748	1089.701	2162.0439	1753.165	1116.7024	3774.4157	3845.7766
981.6952	2812.0072	2661.5706	2588.281	1915.1736	2071.3962	1348.1434	2283.5504	3625.9078	3915.2089

865.9748	1573.7983	2640.3552	2651.9272	1546.7968	3026.09	1974.9625	1762.8083	1461.9351	1523.6527
2920.0416	2898.7975	2636.4978	1932.5317	2040.5375	2391.5562	1824.5259	1529.4388	3666.4099	3793.7024
973.9805	2929.6563	2742.5749	2032.8228	1909.3876	3560.3328	2129.2565	2009.6787	1068.4856	1504.366
1047.2702	2000.0353	2792.7204	1180.3487	2187.1167	2090.683	2374.1981	1201.5641	1434.9337	3934.4957
2358.7687	2954.7291	2590.2096	2135.0425	2094.5403	1787.8811	2464.8458	3606.621	1253.6383	3928.7096
975.9092	1571.8696	2638.4265	2050.1808	1552.5828	3062.7349	2082.9683	1681.804	1157.2046	1452.2918
970.1232	2368.4121	2619.1398	1324.9993	2509.2053	2489.9186	1205.4215	2177.4733	3658.6952	3679.9106
972.0519	995.196	2819.7219	1895.8869	3022.2327	3056.9488	2462.9171	3595.049	3620.1217	3926.781
891.0475	2059.8242	2704.0014	2788.8631	2879.5108	1973.0339	1442.6484	1714.5915	1560.2975	3940.2817
898.7622	1893.9582	1880.4575	2947.0144	1409.8609	2532.3494	1589.2277	2312.4805	1448.4344	3785.9877
1134.0605	3006.8033	2707.8588	2572.8516	1704.9481	3103.237	2441.7017	3598.9063	3631.6938	3784.0591
958.5511	2956.6578	2023.1794	2891.0828	2873.7248	1841.884	1967.2478	1587.299	3608.5497	3947.9964
2912.2983	1928.6743	2054.0382	1978.8199	2881.4395	2547.7788	2322.1239	1429.1477	1153.3473	3703.0547
933.4784	1564.1549	2673.1426	2945.0857	1274.8537	2173.616	2455.2024	1757.0223	1398.2889	3878.5641
945.0504	1091.6297	2219.9042	2786.9344	1832.2406	1888.1722	2181.3307	1292.2118	1373.2161	1517.8667
977.8379	1728.0922	1934.4604	2817.7932	1679.8754	2138.8998	1315.3559	1531.3674	3610.4784	3944.139
979.7666	1724.2349	2603.7104	2688.572	2530.4207	1737.7356	1776.3091	1959.5331	1512.0807	3764.7723
1620.0864	2935.4423	1940.2464	3473.8336	2499.5619	3569.9762	1213.1362	1795.5958	3955.7111	1016.4114
964.3372	875.6182	2675.0713	2096.469	**2850.0101**	3095.5223	2401.1996	1114.7738	1222.7795	3730.0562
991.3386	2335.6246	2644.2125	2576.7089	1703.0194	1919.031	2271.9784	1789.8098	1155.2759	3760.915
1614.3004	2086.8256	2048.2521	1668.3033	2206.4034	1803.3105	2198.6887	1830.3119	1683.7327	1502.4373
867.9035	1566.0836	885.2615	2331.7673	1907.4589	3085.8789	1583.4416	1120.5598	1363.5728	3920.9949
968.1945	1936.389	2835.1513	1045.3415	2044.3948	2460.9885	2453.2738	1760.8797	3568.0475	1467.7212
966.2658	1938.3177	2582.4949	2690.5007	1328.8566	2470.6318	2111.8984	1828.3833	1375.1448	3720.4128
904.5483	2813.9359	1863.0994	2125.3991	2487.9899	1265.2104	2081.0396	1108.9877	1168.7767	3951.8537
889.1189	2364.5547	2767.6477	1720.3775	2229.5475	3076.2356	2273.907	1365.5014	1527.5101	3826.4899
2850.5807	2952.8004	2759.933	1820.6686	3035.7334	2171.6873	3623.9791	1112.8451	1772.4517	3893.9935
2848.652	2810.0785	2740.6462	2052.1095	3049.2341	3072.3782	1215.0648	1299.9265	3909.4229	3807.2031
902.6196	1853.456	2671.214	3500.4755	2202.5461	1344.286	2439.773	2314.4092	3654.8379	3728.1275
962.4085	2933.5137	1672.1607	879.4755	2123.4704	2065.6102	1604.6571	2310.5519	1716.5202	3629.7651
912.263	1722.3062	1859.2421	2150.4719	2142.7572	1971.1052	3591.1916	1305.7125	1388.6455	3849.634
1143.7039	2352.9827	2800.4351	2190.9741	3604.6923	2457.1311	2397.3422	1367.4301	1340.4287	3778.273
943.1218	2821.6506	2594.067	1884.3148	1897.8156	2416.6289	2140.8285	1585.3703	1450.3631	3888.2075
1660.5886	3008.732	2221.8328	3000.4481	1413.7183	3033.8047	1427.219	1286.4258	3772.487	3832.2759
900.6909	2823.5792	2574.7802	2100.3264	2264.2637	2131.1851	2428.201	2383.8415	1417.5756	3708.8408
941.1931	2860.224	2758.0043	1581.513	1994.2493	2262.335	2410.8429	1297.9978	1598.871	3780.2017
960.4798	2030.8941	2702.0728	1687.59	3602.7637	3083.9503	1739.6643	1400.2176	3616.2644	3770.5583
2910.3696	2775.3624	1712.6628	1058.8422	2869.8674	2069.4676	2200.6174	2179.402	1733.8782	3949.925
2341.4106	2192.9027	2516.92	2526.5634	2476.4179	1267.139	2109.9697	3596.9776	3957.6397	1519.7954
2194.8314	1857.3134	2046.3235	2084.897	3577.6909	2449.4164	2291.2651	1801.3818	1419.5043	3814.9178
906.4769	2611.4251	1814.8826	2889.1542	2015.4647	2513.0627	1537.1534	1064.6282	1396.3602	1558.3689
935.4071	2839.0086	1965.3192	2223.7615	1807.1679	2279.6931	2408.9143	1380.9308	1070.4143	1081.9863
910.3343	1903.6016	2948.9431	3039.5908	2501.4906	1270.9964	1062.6996	1866.9568	3645.1945	3795.6311

916.1203	2825.5079	2684.7147	2632.6405	2347.1967	2106.1124	1440.7197	1768.5944	3618.1931	1226.6369
869.8321	1579.5843	2815.8645	2115.7558	2127.3278	2073.3249	2270.0497	1122.4885	1394.4315	3824.5612
1691.4474	2613.3537	2659.6419	2883.3681	2225.6902	2227.6189	2258.4777	1382.8595	3911.3516	3805.2745
2923.8703	1855.3847	2601.7817	2599.853	3047.3055	2067.5389	1766.6657	1425.2903	3998.1419	3922.9236
2852.5093	2794.6491	2682.786	2875.6534	1006.768	2289.3364	1826.4546	1699.1621	3697.2687	1076.2003
914.1916	2736.7889	2665.4279	2630.7118	3055.0202	2372.2694	1782.0951	1743.5216	3612.407	3834.2046
937.3357	2763.7903	2862.1527	1872.7428	2555.4935	1269.0677	2260.4063	1431.0764	3695.34	3766.701
1693.3761	1101.273	2700.1441	2329.8386	2217.9755	2503.4193	1107.0591	2308.6232	1172.634	3942.2104
939.2644	3539.1174	2713.6448	2584.4236	1822.5973	2491.8472	2443.6304	1218.9222	3994.2846	3701.1261
1622.0151	2931.585	1882.3862	1982.6772	2061.7529	2025.1081	2298.9798	1110.9164	1436.8624	3627.8364
1664.446	2777.291	1145.6326	2877.5821	2480.2752	3029.9474	3593.1203	2378.0555	1595.0137	3859.2774
1095.487	997.1246	2754.147	2541.9928	1280.6398	2435.9157	2387.6988	2214.1181	3652.9092	1041.4841
1656.7313	3010.6606	2677	2158.1866	3099.3797	2148.5432	2175.5447	2381.9128	4000.0706	3691.4827
2356.84	1899.7442	2565.1369	2885.2968	2027.0367	1785.9524	2104.1837	1433.005	1749.3076	3714.6268
2846.7233	1986.5346	883.3328	3041.5194	2507.2766	2437.8444	2318.2666	1843.8127	1361.6441	1541.0108
908.4056	2808.1498	1186.1347	2244.9769	2482.2039	2113.8271	2169.7586	1301.8552	3811.0605	3886.2788
2983.6592	1182.2774	2607.5677	2864.0814	2266.1923	3091.665	1353.9294	1290.2831	3996.2132	3897.8508
1666.3746	2827.4366	1816.8112	2595.9957	2351.054	2468.7032	1922.8883	1255.567	3683.768	3919.0663
2979.8019	1980.7485	2771.505	2628.7831	2867.9387	1550.6542	2007.75	2302.8372	3776.3444	3677.982
871.7608	2765.719	999.0533	1805.2392	3097.451	1809.0965	1203.4928	1593.085	1170.7053	1465.7925
2854.438	1990.3919	2617.2111	1323.0706	2520.7774	2075.2536	1209.2788	2165.9013	3967.2831	3924.8523
2966.3011	2609.4964	1951.8184	3037.6621	2121.5418	1944.1037	2241.1196	1369.3588	3984.6412	3930.6383
2981.7305	1130.2032	881.4042	1312.9107	1992.3206	2466.7745	2042.4661	2379.9841	3789.8451	3687.6253
1610.4431	2896.8689	2798.5065	1000.982	2248.8343	3064.6635	1917.1023	1149.4899	3978.8552	3768.6297
2968.2298	1677.9467	2005.8213	2524.6347	1548.7255	2275.8357	2406.9856	1556.4402	3847.7053	3672.1959
2977.8732	2621.0684	2657.7132	2231.4762	2277.7644	2167.83	2433.987	1379.0022	3965.3544	3755.1289
2989.4452	1608.5144	2837.08	1568.0122	**3350.3083**	1764.737	2239.1909	1845.7413	3986.5699	3874.7067
2985.5879	2731.0029	2781.1484	2183.2594	2246.9056	2324.0526	1311.4986	1161.062	1477.3645	1510.152
2925.799	2723.2882	2648.0699	2557.4222	1320.7853	1554.5115	1176.4913	1162.9906	3787.9164	3866.9921
1097.4157	2725.2168	2580.5663	1326.9279	2418.5576	2108.0411	2297.0511	2306.6945	3936.4243	3917.1376
2908.4409	1689.5187	2769.5764	1193.8494	3080.0929	2152.4006	1961.4618	3564.1902	3913.2802	1469.6498
1708.8055	2717.5021	1878.5288	3051.1628	2545.8501	1197.7068	1313.4272	1602.7284	3976.9265	3670.2673
2343.3393	2796.5778	2333.696	2626.8545	2233.4049	1778.2377	1294.1405	1288.3545	1485.0792	3884.3501
2028.9654	2802.3638	1674.0893	2586.3523	1411.7896	2001.964	2316.3379	1164.9193	3656.7665	3861.206
2991.3739	2950.8717	1191.9207	3045.3768	2522.706	2505.348	1211.2075	3585.4056	1083.915	3797.5598
1996.1779	1930.603	2680.8573	2887.2255	2013.536	1784.0238	1797.5245	1359.7154	3982.7125	3932.567
1988.4632	3012.5893	1876.6001	2003.8926	1060.7709	1012.554	2430.1297	1303.7839	3704.9834	3890.1361
1093.5584	2019.322	3020.304	2185.188	1010.6254	1811.0252	1410.647	1438.7911	1386.7169	3757.0576
2975.9445	2592.1383	2646.1412	2528.4921	1195.7781	2063.6815	2285.4791	1745.4503	1338.5	3857.3487
2927.7276	2833.2226	2894.9402	1606.5857	2459.0598	2287.4078	2212.1895	1371.2875	1072.3429	3843.848
1710.7341	1947.9611	2756.0756	1911.3163	2941.2284	1391.1563	3546.8321	1774.3804	3812.9892	3739.6995
2987.5166	2831.2939	3542.9748	1905.5302	2474.4892	2077.1823	1955.6758	1421.433	3693.4114	1521.7241
2354.9114	2773.4337	2136.9712	3500.7608	1920.9596	1346.2147	2320.1952	1390.5742	3988.4985	3722.3415

2856.3667	2738.7176	2698.2154	3053.0915	2871.7961	2349.1253	1261.353	1793.6671	3614.3357	3726.1988
1901.6729	2719.4308	2567.0655	2568.9942	2017.3934	2268.121	2154.3292	1731.9496	3992.3559	3816.8465
2970.1585	2721.3595	1103.2017	2943.157	2472.5605	2235.3336	1309.5699	2304.7658	3971.1405	3799.4885
1080.0576	3905.5655	3758.9863	3803.3458	3647.1232	3892.0648	1228.5656	3880.4928	3674.1246	1080.0576
1078.129	3836.1333	3818.7752	3491.9848	1405.4359	3741.6282	**1450.0079**	1018.3401	3863.1347	1100.129
3882.4214	1249.781	1454.2205	3822.6325	3865.0634	3724.2702	3650.9805	3841.9193	3737.7709	3882.4214
3868.9207	3801.4171	3712.6981	3676.0533	3820.7039	3710.7694	3753.2003	3689.554	3872.7781	

Table 2-A_{1-3}. Wavenumbers in Cluster 3

1631.6585	1633.5872	833.1873	837.0447	806.1859	1649.0166	777.2558	790.7565	3130.2385	3107.0944
1641.3019	3433.0403	3446.541	854.4027	810.0432	823.5439	725.1844	792.6852	3128.3098	3188.0987
1629.7298	3473.5425	3454.2557	827.4013	3512.116	1730.4484	**1484.6138**	1654.8026	3145.6678	3172.6693
1643.2305	3475.4712	1727.1635	3519.8307	852.4741	844.7594	723.3984	1647.0879	3151.4539	3137.9532
1639.3732	3450.3984	856.3314	825.4726	838.9733	796.5425	3529.474	3120.5951	3126.3811	3134.0958
1627.8011	3479.3285	3469.6851	**2784.8079**	811.9719	802.3285	3525.6167	3168.8119	3124.4524	3182.3127
3434.969	1720.7941	3490.9006	3517.902	808.1145	815.8292	781.1131	3112.8804	3190.0274	3161.0972
3436.8977	3440.755	1623.9438	3514.0446	798.4712	819.6866	3527.5454	3122.5238	3116.7377	3174.598
1625.8725	3452.3271	3494.7579	3508.2586	800.3999	850.5454	788.8278	3118.6664	1731.9496	3186.17
3458.1131	3448.4697	3498.6153	3487.0432	842.8307	813.9006	3537.1887	3170.7406	3141.8105	3136.0245
3460.0418	3456.1844	3485.1145	**1250.9452**	3533.3314	3502.4726	786.8991	3114.8091	3139.8818	3184.2413
3465.8278	3471.6138	3481.2572	829.33	3504.4013	1635.5158	783.0418	3166.8833	3163.0259	3180.384
858.2601	3442.6837	3496.6866	1645.1592	840.902	3521.7593	784.9705	3153.3825	3109.023	3155.3112
3438.8263	3467.7565	831.2586	3488.9719	3510.1873	775.3271	846.688	3143.7392	3132.1671	3159.1686
3463.8991	3492.8292	3506.33	3515.9733	817.7579	821.6153	3523.688	3110.9517	3149.5252	3176.5266
3477.3998	3444.6124	3483.1859	3500.5439	3531.4027	3535.2601	848.6167	3164.9546	3147.5965	3178.4553
3157.2399	1652.8739	856.3314	1730.0210	**1750.8867**	2949.1779				

Table 4-A1. Operating conditions for temperature and residence time for PM and Biomass conversion in a stainless-steel micro batch-reactor.

Medium	**T (°c)**	**$t_{1(min)}$**	**$t_{2(min)}$**	**$t_{3(min)}$**
Subcritical water	150	15	25	35
Subcritical water	250	15	25	35
Subcritical water	350	15	25	35
Sulfuric acid	150	10	20	30
Sulfuric acid	200	10	20	30
Sulfuric acid	250	10	20	30
Sodium hydroxide	150	10	20	30
Sodium hydroxide	200	10	20	30
Sodium hydroxide	250	10	20	30

T=Temperature, t=Time, min=minutes