

NIR and Py-mbms coupled with multivariate data analysis as a high-throughput biomass characterization technique : a review

Li Xiao, Hui Wei, Michael E. Himmel, Hasan Jameel and Stephen S. Kelley

Journal Name:	Frontiers in Plant Science
ISSN:	1664-462X
Article type:	Review Article
Received on:	16 May 2014
Accepted on:	22 Jul 2014
Provisional PDF published on:	22 Jul 2014
www.frontiersin.org:	www.frontiersin.org
Citation:	Xiao L, Wei H, Himmel ME, Jameel H and Kelley SS(2014) NIR and Py-mbms coupled with multivariate data analysis as a high-throughput biomass characterization technique : a review. <i>Front. Plant Sci.</i> 5:388. doi:10.3389/fpls.2014.00388
Copyright statement:	© 2014 Xiao, Wei, Himmel, Jameel and Kelley. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY) . The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

This Provisional PDF corresponds to the article as it appeared upon acceptance, after rigorous peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.

1 NIR and Py-mbms coupled with multivariate data analysis as a high-throughput
2 biomass characterization technique: a review

3 Li Xiao^{1,*}, Hui Wei², Michael E. Himmel², Hasan Jameel¹, Stephen S. Kelley^{1,*}

4 ¹North Carolina State University, Raleigh, NC 27606

5 ²Biosciences Center, National Renewable Energy Laboratory, Golden, Colorado 80401

6

7

8 *** Correspondence:**

9 Dr. Stephen S. Kelley or Li Xiao

10 Department of Forest Biomaterials

11 North Carolina State University

12 Raleigh, NC, 27606, USA

13 sskelley@ncsu.edu

14 lxiao3@ncsu.edu

15

16

17

18

19

20

21

22

23

24

25

1 *Abstract*

2 Optimizing the use of lignocellulosic biomass as the feedstock for renewable energy production
3 is currently being developed globally. Biomass is a complex mixture of cellulose, hemicelluloses,
4 lignins, extractives, and proteins; as well as inorganic salts. Cell wall compositional analysis for
5 biomass characterization is laborious and time consuming. In order to characterize biomass fast
6 and efficiently, several high through-put technologies have been successfully developed. Among
7 them, near infrared spectroscopy (NIR) and pyrolysis-molecular beam mass spectrometry (Py-
8 mbms) are complementary tools and capable of evaluating a large number of raw or modified
9 biomass in a short period of time. NIR shows vibrations associated with specific chemical
10 structures whereas Py-mbms depicts the full range of fragments from the decomposition of
11 biomass. Both NIR vibrations and Py-mbms peaks are assigned to possible chemical functional
12 groups and molecular structures. They provide complementary information of chemical insight
13 of biomaterials. However, it is challenging to interpret the informative results because of the
14 large amount of overlapping bands or decomposition fragments contained in the spectra. In order
15 to improve the efficiency of data analysis, multivariate analysis tools have been adapted to define
16 the significant correlations among data variables, so that the large number of bands/peaks could
17 be replaced by a small number of reconstructed variables representing original variation.
18 Reconstructed data variables are used for sample comparison (principal component analysis) and
19 for building regression models (partial least square regression) between biomass chemical
20 structures and properties of interests. In this review, the important biomass chemical structures
21 measured by NIR and Py-mbms are summarized. The advantages and disadvantages of
22 conventional data analysis methods and multivariate data analysis methods are introduced,
23 compared and evaluated. This review aims to serve as a guide for choosing the most effective
24 data analysis methods for NIR and Py-mbms characterization of biomass.

25 Keywords: Biomass; Lignocellulosic biofuel; Cellulose, Lignin; S/G-lignin; Near infrared
26 spectroscopy; Pyrolysis molecular beam; Mass spectrometry; Multivariate data analysis; High
27 throughput; Chemometrics

28

29

1 **Introduction for biomass chemical composition**

2 Biomass is a complicated mixture of organic and inorganic compounds. It is mainly composed of
3 cellulose, hemicelluloses and lignins, as well as minor components, such as proteins, extractives,
4 ash and other nonstructural mineral materials. Because of its renewable nature and chemical
5 composition, biomass is an attractive feedstock for energy and chemical products (Ragauskas et
6 al., 2006;Himmel et al., 2007;Wei et al., 2009;Sluiter et al., 2010). In order to provide an
7 effective guide for feedstock selection and process development, it is very important to measure
8 biomass chemical composition accurately and efficiently (Sluiter et al., 2010;Templeton et al.,
9 2010;Daystar et al., 2013). In this paper, we will review the use of two high-throughput
10 techniques, near infrared spectroscopy (NIR) and pyrolysis-molecular beam mass spectrometry
11 (Py-mbms) in biomass characterization. The advantages and disadvantages of different data
12 analysis methods, including band/peak assignment, tools for spectral treatments and resolution
13 enhancement and multivariate data analysis methods, are introduced, compared and evaluated.
14 Selected research publications are reviewed and categorized as ‘case studies’ according to the
15 ways they analyzed data and the specific biomass properties that are evaluated.

16

17 **Conventional biomass characterization relevant to biofuel production**

18 Traditional biomass compositional analysis, based on two-stage sulfuric acid hydrolysis followed
19 by gravimetric and instrumental analysis, has been used to measure lignin and carbohydrates for
20 more than 100 years. These methods have been used by researchers for studies of wood materials,
21 animal food, human health, bioenergy production, and many other areas related to biomaterials.
22 The history and uses of these methods were reviewed in detail elsewhere (Sluiter et al., 2010).
23 The analytical uncertainty for different methods was also evaluated by statistical analysis and
24 reported as the standard deviation of measurement for each component (Templeton et al., 2010).
25 Other wet chemical techniques also include: acidolysis, thioacidolysis, nitrobenzene oxidation,
26 transesterification, acetyl bromide method, orcinol method, Van Soest method, etc. Routine
27 procedures, a number of less common methods, and new analytical methods developed for
28 research purposes in the field of wood chemistry are described in books (Browning,
29 1967;Sjöström and Alen, 1999). These techniques quantify important chemical structure biomass,
30 but they are time consuming and laborious.

31 Separately, combustion-related properties are of interest for the utilization of biomass in biofuel
32 and biopower production. There are three types of combustion-related properties: morphological,
33 physical, and chemical properties (Braadbaart and Poole, 2008). Traditional fuel analysis of
34 biomass includes ultimate analysis, proximate analysis, and thermogravimetric analysis. In
35 addition, ash composition and sulfur can be determined and used to predict fuel indices,
36 especially for slagging behavior, aerosol formation, and corrosion related risks (Oberberger,
37 2014).

1 **Use of spectroscopic tools in biomass characterization as high throughput techniques**

2 Spectroscopic methods, such as Fourier transform infrared spectroscopy (FTIR), NIR, Raman
3 spectroscopy (Raman), and nuclear magnetic resonance (NMR), are widely used to measure
4 functional groups and chemical bonds in biomass. These measurements are faster and more
5 convenient than most conventional chemical methods used for biomass characterization and fuel
6 analysis. Besides, since there is no degradative chemical treatment used during analysis, the
7 information gained from these tools is more representative of the chemical structures in original
8 biomass. However, there are some drawbacks for using these spectroscopic tools. For example,
9 data interpretation for FTIR, Raman and NMR is relatively complicated, sample preparation can
10 be complex, and due to the mixed nature of biomass, peak assignment usually suffers from the
11 overlap of many compounds. A good summary of spectroscopic tools used as high throughput
12 techniques in biomass study can be found in a recent review (Lupoi et al., 2014)

13

14 **High throughput techniques coupled with multivariate statistical analysis**

15 Because of many chemical features included in a single spectrum, it is challenging to elucidate
16 data directly for a group of samples. Therefore, multivariate analysis tools have been widely used
17 in spectroscopic data analysis (Jin and Xu, 2011; Smith-Moritz et al., 2011; Xu et al., 2013; Lupoi
18 et al., 2014). Among them, the two multivariate tools that have been widely used are: **(1)**
19 Principal Component Analysis (PCA), and **(2)** Partial Least Square (PLS).

20

21 PCA is mainly used for identifying outliers, sample comparison and screening. It relies on
22 projecting original samples variables on several (usually less than six) reconstructed variables
23 which are representative of original sample variation. Those reconstructed variables are known
24 as principal components (PCs). Samples described with PCs can be plotted in scores plot, in
25 which similar samples cluster together while samples different from each other are separated in
26 two-, three- or n- dimensional coordinates. Together with the scores plot, PCA loadings plot
27 allows for the determination of important chemical features responsible for the sample grouping.
28 In the loadings plot, variables with large values are highly correlated with sample grouping
29 (Sykes et al., 2009).

30

31 PLS is used to build prediction correlation models between spectral data and the property of
32 interest. In the application of NIR and Py-mbms, spectral data is regarded as ‘predictors’ for the
33 biomass properties of interest. The properties of a new sample can then be estimated using a PLS
34 model built from spectral data taken on a set of similar samples with known characteristics. In
35 this way, time consuming experiments for new samples could be eliminated. Regression

1 coefficients are generated and can be used to relate chemical features in the spectra to the
2 specific sample properties (Labbe et al., 2006).

3 In summary, multivariate tools used in spectroscopic data analysis have three functions: **(1)**
4 comparing sample similarities and differences and discovering outliers; **(2)** building prediction
5 models between spectroscopic data and biomass properties of interest; and **(3)** discovering
6 correlations between property data and spectral data.

7

8 **Biomass characterization by NIR spectroscopy**

9 NIR is normally considered to be in the range of electromagnetic spectrum from 12000 cm⁻¹ to
10 4000 cm⁻¹ (Smith-Moritz et al., 2011). This wavelength region has two major advantages: first,
11 the speed of spectral acquisition is high, which facilitates the real-time data collection for process
12 control; secondly, the wide applicability to a diverse ranges of materials with little or no sample
13 preparation (Schwanninger et al., 2011). This allows NIR to be effective for online monitoring
14 and quality control of a wide variety of product properties and manufacturing processes
15 (Workman, 2001;Kelley et al., 2004a;Tsuchikawa, 2007;Jin and Xu, 2011). Because of this, NIR
16 has been extensively used as a high-throughput method to determine chemical, physical,
17 mechanical, and fuel properties of woody biomass during the past 20 years.

18 However, there are some disadvantages to NIR. Although NIR absorption spectra have similar
19 patterns to those in the mid-IR, they have wider separation, more anti-symmetry, and weaker
20 intensity due to the fact that it is the combination and overtone bands from fundamental
21 vibrations involved in NIR region. Therefore, the interpretation of NIR spectra are much harder
22 than mid-IR (Schwanninger et al., 2011;Lupoi et al., 2014).

23 The utility of band assignments depends on the purpose of specific research or application. There
24 is ongoing discussion around the necessity of interpreting NIR spectra in detail.

25 Chemical/physical information contained in the NIR spectra can be used for detailed analysis
26 (Schwanninger et al., 2011). However, it is not necessary to fully understand the chemical details
27 for NIR to be useful for quantitative analysis. If NIR is used as a fast tool in distinguishing
28 samples and in building prediction models for biomass properties, the detailed assignments are
29 generally not needed. Statistical analysis for extracting useful information is essential for this
30 purpose (Xu et al., 2013). Meaningful scientific insight of structural information could be better
31 gained with the help of both statistical analysis and band assignments.

32 **NIR band assignment and data processing**

33 In NIR analysis, data points are usually collected in reflectance form (R) and converted to
34 log₁₀(1/R) form, which is equivalent to an absorbance spectrum.

1 As stated above, knowledge regarding band assignment is important for the understanding of
2 chemical structures in biomass and there are several references on NIR band assignments
3 (Tsuchikawa et al., 2003; Schwanninger et al., 2011; Via et al., 2013). Commonly assigned
4 vibrations in the NIR spectra of woody biomass include (Schwanninger et al., 2011):
5

6 **1370nm – 1471nm:** First and second overtones of O-H stretching vibrations from free or
7 weakly bonded O-H in carbohydrates and first overtones of C-H, C_{aromatic}-H stretching
8 vibrations, such as first overtone of O-H stretching in free OH group or OH group with a
9 weak H-bond from cellulose, xylan and glucomannan (1386, 1414, 1428, 1471, 1477-
10 1484), first overtone of O-H stretching in phenolic hydroxyl groups from extractive or
11 lignin (1410, 1447, 1448), first overtone of C-H stretching and bending in aromatic
12 associated C-H from lignin (1417, 1440).

13 **1471nm – 1632nm:** first overtone of O-H stretching from strong O-H bonded group,
14 semi-crystalline and crystalline region of cellulose (1473-1632) or intramolecular H-bond
15 in glucomannan (1471, 1493).

16 **1666nm – 2000nm:** first overtone of aliphatic and aromatic C-H stretching vibrations and
17 O-H combination bands from extractives/lignin (e.g. 1668, 1674, 1684, 1726),
18 hemicellulose (e.g. 1720, 1724), cellulose (e.g. 1723, 1731), which are overlapped with
19 each other and water band (e.g. 1887-2000).

20 **Above 2000nm:** Assignment in this region is difficult due to high number of possibilities
21 for the coupling of vibrations.

22 There are a number of well-established NIR spectra preprocessing techniques that can be used to
23 achieve resolution enhancement and to more precisely locate band position. Methods for spectral
24 data preprocessing include: **(1)** smoothing and derivatization (Dodd, 2002; Rousset et al., 2011)
25 such as using the algorithm based method used by Savitzky and Golay (Savitzky and Golay,
26 1964), **(2)** calculation of differential spectra (Rousset et al., 2011), and **(3)** Fourier self de-
27 convolution, curve fitting (Ozaki et al., 2001) with more advanced techniques involving principal
28 component analysis (Fackler and Schwanninger, 2010) and two dimensional correlation analysis
29 (Ozaki et al., 2001; Schwanninger et al., 2011).

30 Among those preprocessing methods, derivatives are widely used to reduce the impact of
31 overlapping peaks and baseline variation. However, there is a concern that generating derivatives
32 can possibly generate false information. Both the shape of the spectrum and the data processing
33 algorithms have an impact on band shape and location. Differences between the location of the
34 bands between the raw and the second derivative spectrum can be more than 20 cm⁻¹ (5 nm).
35 Researchers have also reported that the second derivative form was not always more precise than
36 the normal form for the prediction of lignin in wood (Michell, 1995; Xu et al., 2013). Therefore,
37 when spectral data is processed with the second derivative, possible peak shifts should be taken

1 into consideration. The same consideration is also important for deriving conclusions from
2 processing spectra of PCA and regression coefficients from PLS (Schwanninger et al., 2011).

3 **NIR spectroscopy coupled with PCA**

4 The primary application of NIR coupled with PCA is to classify biomass samples of various
5 origins or from different pretreatments without conducting laborious traditional wet chemistry
6 techniques on all samples. Related areas of this application are summarized below:

7 (1) Related to species/plant fractions (Michell, 1995;Kelley et al., 2004a;Labbe et al.,
8 2008a;Labbe et al., 2008b;Nkansah et al., 2010);

9
10 (2) Related to genetic engineering of feedstock crops (Baillères et al., 2002;Sandak and
11 Sandak, 2011;Zhou et al., 2011);

12
13 (3) Related to chemical/thermal/biological treatments (Kelley et al., 2004b;Yang et al.,
14 2007;Houghton et al., 2009;Krongtaew et al., 2010).

15
16 For example, in order to evaluate the impact of biomass pretreatments (including acid and
17 alkaline pretreatments, some in combination with hydrogen peroxide) on the change of
18 cell wall compositions of wheat and oat straw, FT-NIR was utilized to characterize raw
19 and pretreated straw (Krongtaew et al., 2010). Second derivatives from NIR absorption
20 bands were generated and evaluated to show the changes in properties related to biomass
21 recalcitrance during subsequent bioethanol production. These properties include the
22 change of lignin, hemicelluloses; as well as amorphous, semi-crystalline, and crystalline
23 regions of cellulose moieties of pretreated sample. PCA of derivative data was efficiently
24 utilized to differentiate the alterations in chemical structure of straw due to different
25 pretreatment methods as shown in Figure 1. It was demonstrated that FT-NIR coupled
26 with PCA is a powerful tool to assess biomass digestibility, with a potential to be used in
27 process control in the area of biomass utilization or energy conversion.

28 29 **NIR spectroscopy coupled with PLS**

30 One of the main applications of NIR coupled with PLS is to build regression models for the
31 prediction of biomass properties, such as lignin content, S/G-lignin ratio, moisture content,
32 heating value (Kelley et al., 2004a;Rousset et al., 2011;Schwanninger et al., 2011).

33 Related areas of the application of NIR coupled with PLS in existing literatures are summarized
34 below:

35 (1) Prediction of cell wall components (Michell, 1995;Sanderson et al., 1996;Tucker et al.,
36 2001;Baillères et al., 2002;Kelley et al., 2004a;Lovett et al., 2004;Yeh et al., 2004;Jin and
37 Chen, 2007;Labbe et al., 2008b;Philip Ye et al., 2008;Wolfrum and Sluiter,
38 2009;Nkansah et al., 2010;Hou and Li, 2011;Sandak and Sandak, 2011;Smith-Moritz et
39 al., 2011;Zhou et al., 2011).

40

1 For example, in order to identify specific monosaccharide outliers from a plant mutant
2 population, FT-NIR coupled with PLS regression was utilized to analyze plant leaves of
3 Arabidopsis (Smith-Moritz et al., 2011). Various Arabidopsis cell wall mutants were
4 analyzed for prediction model building. PCA was performed on pre-processed and area-
5 normalized NIR spectra, followed by calculation of the Mahalanobis distance, a linear
6 discriminate analysis technique to identify outliers using PCA results. By using this
7 technique, a pilot study was conducted which consisted of 550 mutant lines (3590 leaf
8 samples), resulting in a set of 235 leaf samples as Mahalanobis outliers. Quantitative
9 information about monosaccharide composition is gained by means of PLS modeling
10 with known biochemical values and FT-NIR spectra. The correlation between predicted
11 and experiment determined monosaccharide composition (mol%) of 226 rice leaf samples
12 are shown in Figure 2 with $R^2 = 0.98$ (Smith-Moritz et al., 2011).
13

- 14 (2) Prediction of other physical properties (Thygesen, 1994; Hoffmeyer and Pedersen, 1995),
15 mechanical properties (Kelley et al., 2004a; André et al., 2006), fuel properties (Lestander
16 and Rhen, 2005; Labbe et al., 2008a):
17

18 For example, NIR coupled with PLS has been used to predict cell wall chemistry and
19 mechanical properties of loblolly pine from different radial locations and heights of trees
20 grown in Arkansas (Kelley et al., 2004a). Mechanical properties include three point
21 bending test and related microfibril angle. The correlation between experimental data and
22 predicted data from PLS modeling is very strong with correlation coefficients (r) as high
23 as 0.80. A reduced spectral range (650 nm – 1150 nm) usually available in handheld NIR
24 spectrometers was also demonstrated to be useful for predicting mechanical properties.
25

26 **Biomass characterization by Py-mbms**

27 Py-mbms has been intensively used for studies of biological and synthetic macromolecules, such
28 as wood, grasses, carbon in soil and chars. It has proved to be an efficient and powerful
29 analytical tool (Evans and Milne, 1987; Kelley et al., 2002; Labbe et al., 2005; Magrini et al.,
30 2007; Sykes et al., 2008; Mann et al., 2009; French and Czernik, 2010). Detailed description of
31 this technology is available in the above references. In short, the Py-mbms is composed of a
32 pyrolysis furnace and a free-jet molecular beam mass spectrometer (mbms). Typically the
33 furnace is preheated to 500°C before ground sample of biomass is inserted into the inert
34 atmosphere of the furnace. Pyrolysis products from biomass in the furnace are swept out of the
35 furnace into the mbms by an argon gas stream. Molecular fragments contained in the pyrolysis
36 vapor are expanded in a series of vacuum chambers to be quenched; so that intermolecular
37 collisions are prevented. A low-energy electron beam (17eV – 23 eV) in the triple quadruple
38 mass spectrometer is employed to produce a positive ion mass spectrum. The positive ion stream
39 is magnified and collected by the detector.

40 Mass peaks were assigned to chemical fragments produced from fast pyrolysis of biomass for
41 direct interpretation (Evans and Milne, 1987). The spectra from Py-mbms is also interpreted with

1 the help of multivariate analysis tools, especially PLS and PCA (Hoover et al., 2002;Kelley et al.,
2 2002;Kelley et al., 2004b;Labbe et al., 2005;Magrini et al., 2007;Mann et al., 2009).

3

4 **Py-mbms peak assignment and data processing**

5 During data acquisition of Py-mbms, amplified positive ions from biomass pyrolysis vapor are
6 scanned continuously; then the signal is collected by a computer. Approximate evolution time of
7 fast pyrolysis for a sample of 4 mg is less than 1min. During the evolution time there are
8 typically 50 single scans collected. Biomass with larger sample size will need longer evolution
9 time and more scans during fast pyrolysis. Together with single scan spectrum, time resolved
10 profile and averaged spectrum can be collected by the computer acquisition software (Evans and
11 Milne, 1987).

12 Average spectra are also known as spectral ‘fingerprints’. Spectral fingerprints gained at
13 analytical pyrolysis temperature of 500 to 550 °C and the molecular beam free jet expansion
14 represent primary products from biomass pyrolysis. Studies shown that at this temperature range,
15 molecular structure of the original biomass is well preserved and there is no interaction observed
16 among organic components during pyrolysis, although inorganics may alter the pyrolysis
17 pathways of the carbohydrates (Evans and Milne, 1987). Thus, with known peak assignment,
18 spectral ‘fingerprints’ generated could be used to depict the molecular structure of chemical
19 composition in biomass. A summary of important peak assignment in biomass is shown in Table
20 1 (Evans and Milne, 1987;Sykes et al., 2008). Characteristic spectral fingerprints of whole
21 biomass samples and separated constituents of biomass are shown in Figure 3 (Evans and Milne,
22 1987).

23 Py-mbms has been successfully applied in many biomass-related studies, including the research
24 of cellulose, cellulose with inorganics, many woods, xylan, milled wood lignin, bagasse, (Evans
25 and Milne, 1987), herbaceous biomass under different storage environments (Aglevor et al.,
26 1994), hardwood sawdust and its torrefaction products (Nimlos et al., 2003), and poplar grown
27 under different nitrogen conditions (Sykes et al., 2009).

28 For example, in the study of bark phenolysis conducted by Alma and Kelley, bark and its
29 phenolysis products from Calabrian pine, Lebanon cedar, acacia, and European chestnut were
30 characterized using Py-mbms (Alma and Kelley, 2002). From the results of Py-mbms averaged
31 spectra, it was shown that bark **(1)** has less common lignin peaks at m/z 180, 194, 210 assigned
32 to coniferyl alcohol/vinylsyringol, 4-propenylsyringol/ferulic acid, and sinapyl alcohol,
33 respectively; **(2)** has unique triplet of peaks at m/z of 96, 97, 98 assigned to furans; and **(3)** has
34 more phenols, such as peaks at m/z of 110, 124, 150, and 164 assigned to catechol, guaiacol,
35 vinyl guaiacol, and isoeugenol. In softwood bark, extractives and lignin dimers can be identified
36 at m/z of 298, 300, 302, and 272 assigned to didehydro abeitic acid, dehydro abeitic acid, abeitic

1 acid, and lignin dimer, respectively (Alma and Kelley, 2002). These results are consistent with
2 known differences between bark and wood.

3 **Selected peaks from Py-mbms raw data**

4 As summarized above, certain Py-mbms peaks can be unambiguously assigned to specific
5 biomass components. Lignin fragments are particularly easy to identify. Because of this, Klason
6 lignin content of biomass can be directly estimated from Py-mbms spectral fingerprints. Firstly,
7 spectral fingerprints of samples are area/mean normalized for the mass of the original sample.
8 Then, the total intensity of lignin related peaks from the normalized spectrum is calculated. After
9 that, a correction factor is calculated by dividing the known Klason lignin value by the summed
10 intensity of a NIST standard material. The correction factor can be used to convert the total
11 intensity of lignin related peaks to Klason lignin content (Davis and Lagutaris, 2002; Sykes et al.,
12 2008; Sykes et al., 2009; Ziebell et al., 2013). Similarly, S/G ratios were determined by dividing
13 the sum of S-lignin peaks by the sum of G-lignin peaks excluding peaks associated with both S
14 and G fragments (Davis and Lagutaris, 2002; Sykes et al., 2008; Mann et al., 2009; Sykes et al.,
15 2009; Ziebell et al., 2013).

16 For example, corrected lignin values and S/G-lignin ratio were determined from Py-mbms for
17 800 greenhouse-grown poplar trees grown under atmosphere containing different amount of
18 nitrogen (Sykes et al., 2009). Lignin contents ranged from 13 to 28% whereas S/G ranged from
19 0.5 to 1.5. It was shown that the variations in cell wall composition were larger in the plants
20 grown under high nitrogen conditions than those grown under low nitrogen conditions.

21 Similarly, 'within-tree' variability in lignin content and S/G ratio with increasing height and
22 increasing ring for poplars was determined by Py-mbms (Sykes et al., 2008). Wood disks from
23 seven different poplar trees, which were seven years old, were sampled at five different heights
24 of 0.3, 0.6, 1.2, 1.8, and 2.4 m from base to stem. Samples were collected from the north side of
25 each wood disk taken at height of 1.2 m to study difference between growth rings. According to
26 results from Py-mbms, ring effect on lignin content was significant while the effect of height was
27 small. Higher S/G ratio was observed with increasing ring size, whereas lignin content decreased.
28 S/G ratio was determined for switchgrass grown under different environment using the same
29 methodology (Mann et al., 2009).

30

31 **Py-mbms coupled with PCA**

32 Py-mbms coupled with PCA provides a fast analytical method to distinguish a large number of
33 biomass samples. It has been used to study biomass compositional variations due to species
34 (Evans and Milne, 1987; Agblevor et al., 1994; Alma and Kelley, 2002; Kelley et al., 2004b),
35 genetic engineering (Labbe et al., 2005; Davis et al., 2006), different growth environments (Mann
36 et al., 2009; Sykes et al., 2009), thermal (Nimlos et al., 2003)/chemical (Alma and Kelley,

1 2002;Kelley et al., 2004b)/biological (Kelley et al., 2002;Arantes et al., 2009) treatments, and
2 various storage/collection (Agblevor et al., 1994) methods.

3 For example, Py-mbms coupled with PCA has been used to measure the overall composition
4 between and within a series of original and transgenic aspens (Labbe et al., 2005). Two clones
5 were transformed with GRP-*iaaM* gene (N1-17-26 and N1-2-1) and GRP-*iaaM*/35S-ACCase
6 (N2-4-9 and N2-5-5). PCA analysis was conducted for data analysis with an attempt to identify
7 chemical differences between the modified and control aspens. Figure 4 shows PCA scores plots
8 with four replicate samples from five different aspen samples. Figure 4a shows a plot of PC1
9 versus PC2, while Figure 4b shows a plot of PC2 versus PC3. In Figure 4a, there is clear
10 separation between the two N1 samples while two N2 samples are indistinguishable. Moreover,
11 two N2 samples are clearly separated from each other along PC3 as shown in Figure 4b. The
12 loadings from PCA are shown in Figure 5. Using PC1 loadings as an example, C5 carbohydrates
13 (m/z 85 and 114) and lignin (m/z 137, 180, 210, and 272) are highlighted for PC1. This suggests
14 there are more C5 sugars and less lignin in controls than those in N1 and N2 samples (Labbe et
15 al., 2005).

16

17 Py-mbms had been also used to study the impact of storage environment on herbaceous material.
18 Weathered and unweathered fractions of three types of herbaceous biomass after storage at 18
19 different conditions for 6 to 9 months were analyzed by Py-mbms coupled with PCA (Agblevor
20 et al., 1994). Two major trends in the data were shown by PCA (factor analysis): major clusters
21 were distinguished by relative nitrogen contents between switchgrass and the other two
22 herbaceous biomass samples; subgroups of weathered and unweathered materials are clearly
23 separated as subgroups within the major clusters. According to the variance diagram (similar to
24 loadings plot), lower amount of carbohydrates constituted the major chemical difference between
25 weathered and unweathered samples (Agblevor et al., 1994). This observation is consistent with
26 results from traditional wet chemical analysis and Py-GC/MS.

27 In some cases, there is no separation of clusters in PCA scores plot. This indicates that there is no
28 comprehensive difference among samples for the specific chemical features included in those
29 particular PCs.

30 For example, three transgenic clones of populus wood were analyzed by Py-mbms, GC/MS,
31 and traditional wet chemical techniques to screen for possible variations in cell wall composition
32 due to genetic engineering (Davis et al., 2006). Various *Bacillus thuringiensis* (Bt) gene-
33 containing constructs were used to transform poplar genotypes. Transgenic poplar was then
34 compared with non-transgenic control. PCA results showed that there were generally no distinct
35 groupings of individual transgenic lines or non-transgenic controls, indicating no significant
36 differences in cell wall composition between control and transgenic poplars (Davis et al., 2006).

37 **Py-mbms coupled with PLS**

1 One of the primary applications of Py-mbms has been the development of prediction models for
2 biomass compositional properties. Results from conventional methods of cell wall compositional
3 analysis were used as references to build calibration models with capability for predicting the
4 composition for future samples. As a result, laborious wet chemistry techniques can be
5 eliminated. PLS regression is widely used in this arena for both woody (Tuskan et al.,
6 1999;Labbe et al., 2005) and herbaceous biomass (Agblevor et al., 1994;Kelley et al.,
7 2004b;Mann et al., 2009).

8 For example, the effectiveness of NIR and Py-mbms in predicting cell wall composition of
9 various agricultural residues was tested (Kelley et al., 2004b). 41 samples from 14 species with
10 known content of lignin and six individual sugars were analyzed by NIR and Py-mbms.
11 Prediction models were built between spectral data from both techniques and cell wall
12 compositional data. Correlation coefficient and root mean square error data for each calibration
13 and validation model was presented and compared. Good correlations between the predicted and
14 measured value of major components (lignin, glucose, xylose, and mannose) were obtained
15 (correlation coefficients of both calibration and validation model are above 0.80 for both NIR
16 and Py-mbms), while correlations for minor sugars (mannose, galactose, arabinose, and
17 rhamnose) were not as good. A summary of PLS prediction of chemical composition from Py-
18 mbms is presented in Table 2. According to the author, more samples for specific feedstocks are
19 needed for building improved models. This work also did a thorough comparison between NIR
20 and Py-mbms (Kelley et al., 2004b).

21

22 Other than being used to predict cell wall composition of biomass, PLS has been applied in
23 predicting other biomass properties and processing parameters. The acidic phenolysis condition
24 of bark (Alma and Kelley, 2002), weight loss during fungal degradation of spruce (Kelley et al.,
25 2002) and carbon content/fraction of different soils (Hoover et al., 2002;Magrini et al., 2007)
26 were also predicted by Py-mbms coupled with PLS.

27 For example, NIR and Py-mbms were utilized to monitor the chemical changes of wood
28 undergoing brown-rot degradation. In this case, spruce blocks were infected by *Postia placenta*
29 or *Glaoeophyllum trabeum* for 0, 2, 4, 8, and 16 weeks (Kelley et al., 2002). Weight losses over
30 the time period were monitored and recorded. PLS models were built to predict weight loss.
31 Strong correlation between recorded weight loss and predicted weight were obtained (correlation
32 coefficients of calibration model reached 0.98, while those of test model reached 0.96 for both
33 NIR and Py-mbms). The regression coefficients for PLS model from Py-mbms data show that
34 weight loss during decay is positively correlated to carbohydrates (m/z 85, 114, and 126) and
35 negatively correlated to monomethoxylated lignin fragments (m/z 123, 138, and 151) (Kelley et
36 al., 2002).

37 **Conclusions**

1 Compared to traditional techniques in biomass characterization, high-throughput analytical
2 techniques, such as NIR and Py-mbms have been proved to be efficient tools in exploring the
3 chemical features of different biomass samples with minimal sample preparation. These high-
4 throughput techniques coupled with multivariate analysis (MVA) have been demonstrated to be
5 efficient in identifying outliers, comparing samples (using PCA), and building prediction models
6 (using PLS). Both NIR and Py-mbms coupled with MVA could be used not only for
7 characterizing the cell wall chemistry, but also for predicting other chemical, physical,
8 mechanical, and fuel properties. In comparison with Py-mbms, NIR has the advantages of low
9 cost and simple instrumentation, field-portable, and nondestructive, whereas Py-mbms provides
10 superior information of molecular structural information.

11 Thus, we recommend that NIR and Py-mbms coupled with multivariate analysis should be
12 widely employed for biomass characterization. Additional fundamental work on assigning NIR
13 vibrations band and Py-mbms peaks for modified biomass or biomass related products are
14 recommended since current assignment are mainly based on the study of unmodified biomass.
15 Lack of assignments for new bands/peaks in modified biomass limit the application of these two
16 techniques in exploring the fundamental changes of chemical composition of modified biomass.
17 Also, comparison and correlation between analytical results from Py-GC/MS and Py-mbms
18 should be encouraged because of the important similarity and differences in these two techniques
19 are critical for using those techniques for the characterization of biomass molecular structure.

20

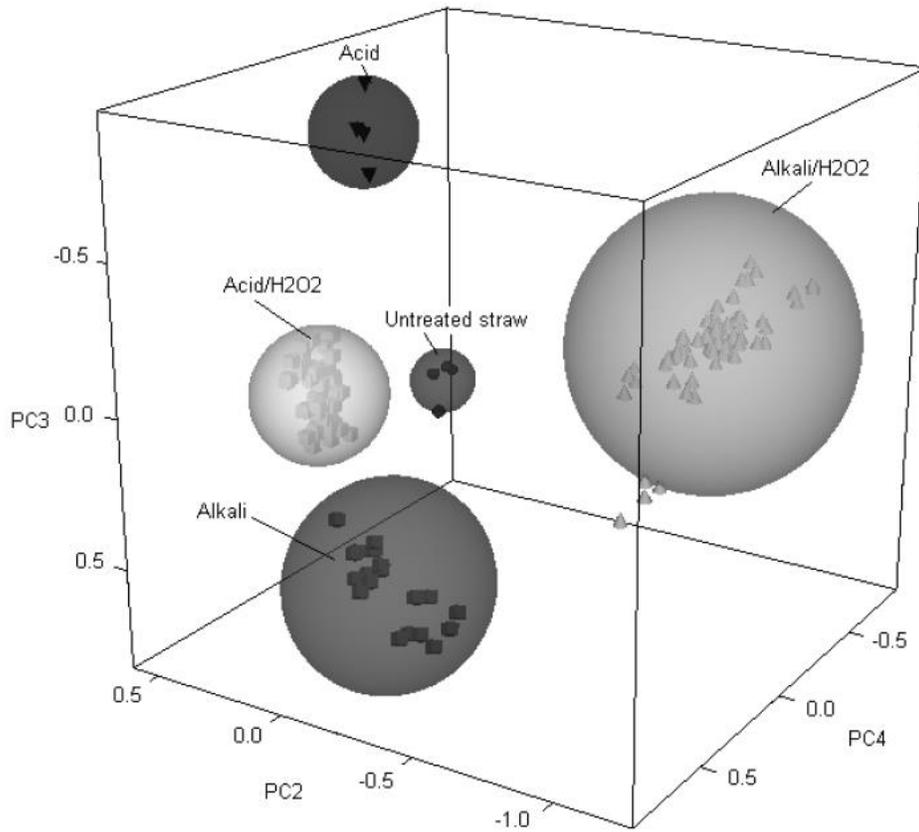
21 **Acknowledgements**

22 This project is supported by the project of Southeastern Partnership for Integrated Biomass
23 Supply Systems (IBSS). The IBSS project is supported by Agriculture and Food Research
24 Initiative Competitive Grant No. 2011-68005-30410 from the USDA National Institute of Food
25 and Agriculture. HW and MEH are sponsored by the Center for Direct Catalytic Conversion of
26 Biomass to Biofuels (C3Bio), an Energy Frontier Research Center funded by the U.S.
27 Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number
28 DE-SC0000997, and also by Laboratory Directed Research & Development (LDRD) program at
29 the National Renewable Energy Laboratory (NREL). Experimental data in figures and tables
30 were generated with the help of Robert Sykes and Mark Davis at National Renewable Energy
31 Laboratory (NREL). This work was conducted as part of the BioEnergy Science Center (BESC).
32 The BESC is a U.S. Department of Energy Bioenergy Research Center supported by the Office
33 of Biological and Environmental Research in the DOE Office of Science.

34

35

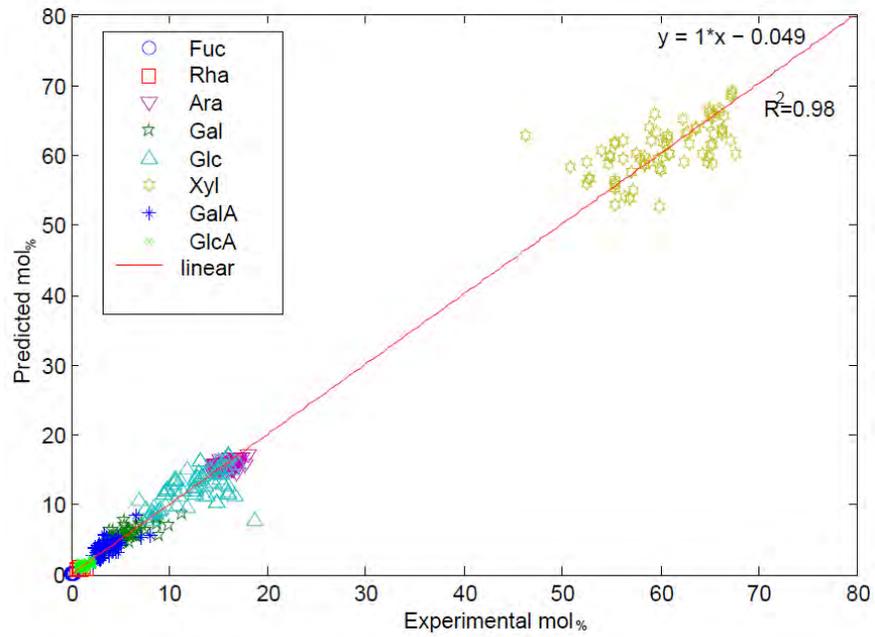
1
2



3
4
5
6
7
8
9
10
11
12
13

Figure 1. PCA scores plot of untreated wheat straw samples (●) and samples treated with acid (▼), alkali (■), acid/H₂O₂ (□), and alkali/H₂O₂ (Δ) as reproduced from literature (Krongtaew et al., 2010).

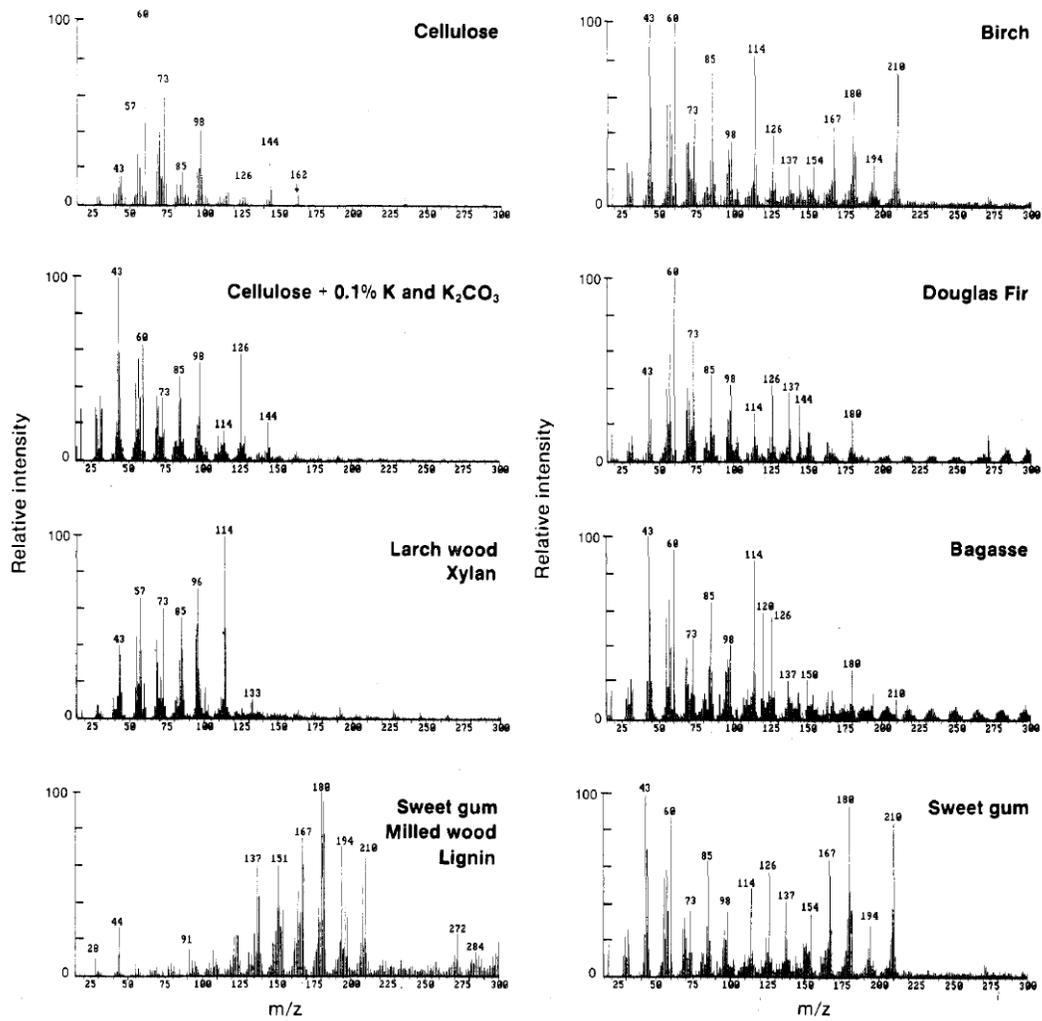
1
2



3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

Figure 2. A correlation analysis predicted (PLS model of FT-NIR) versus experimentally determined monosaccharide composition (mol%) of rice leaf samples. The correlation coefficient between experimental and predicted values was calculated to be $R^2 = 0.98$ as reproduced from literature (Aglevor et al., 1994;Smith-Moritz et al., 2011).

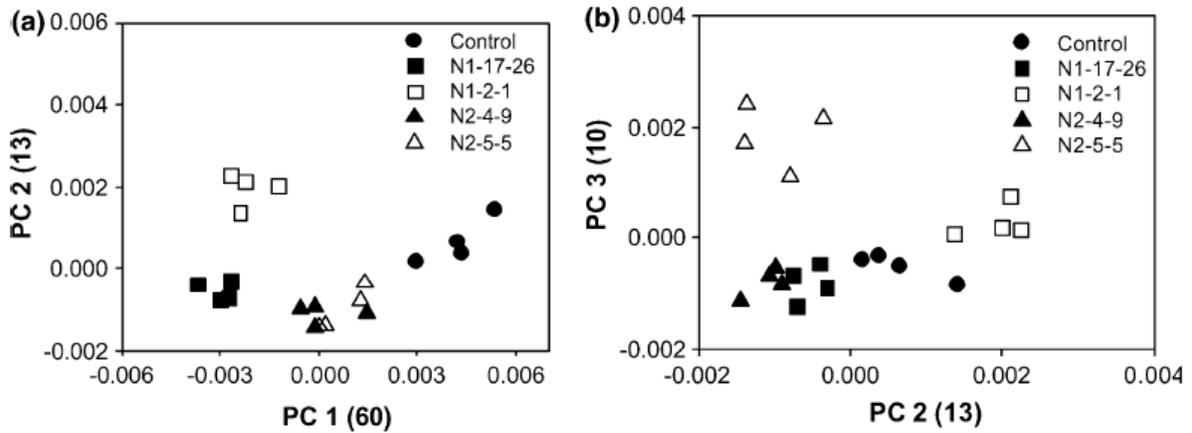
1
2



3
4
5
6
7
8
9
10
11
12

Figure 3. Characteristic mass spectral patterns of primary pyrolysis products for several whole biomass samples and for separated constituents of biomass (Evans and Milne, 1987).

1



2

3 **Figure 4.** Scores plot of PCA of Py-mbms data for original and transgenic aspens; (a) PC1
4 versus PC2; (b) PC2 versus PC3; N1 samples are clearly separated from control samples in (a)
5 while two N2 samples are not distinguishable; In (b) two N2 samples are clearly separated by
6 PC3 as reproduced from literature (Labbe et al., 2005).

7

8

9

10

11

12

13

14

15

16

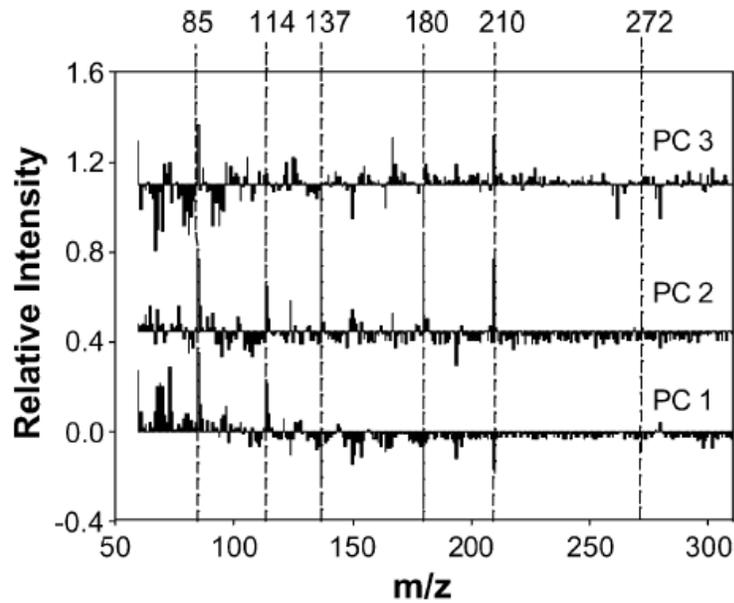
17

18

19

20

1



2

3 **Figure 5.** Loadings from PCA of Py-mbms data for original and transgenic aspens; from top to
4 bottom: PC3, PC2, PC1; C5 carbohydrates (m/z 85 and 114) and lignin (m/z 137, 180, 210, and
5 272) are highlighted for PC1 as reproduced from literature (Labbe et al., 2005).

6

7

8

9

10

11

12

13

14

15

16

17

1 **Table 1.** Peak assignments associated with Py-mbms spectrum for Populus wood based on
 2 literature (Evans and Milne, 1987; Sykes et al., 2008)

Mass peaks (m/z)	Assigned products	S or G precursor
57, 73, 85, 96, 114	from C5 sugar	
57, 60, 73, 98, 126, 144	from C6 sugar	
94	Phenol, dimethylcyclopentene	
108	Methyl phenol (<i>o</i> -cresol, <i>m/p</i> -cresol)	
110	Dihydroxybenzene, 5-methylfurfural	
120	Vinylphenol	
122	Ethylphenol, ethylphenol, benzoic acid	
124	Guaiacol (2-methoxyphenol), trimethylcyclopentenone	G
137*	Ethylguaiacol, homovanillin, coniferyl alcohol	G
138	Methylguaiacol	G
150	<i>p</i> -inylguaiacol, coumaryl alcohol	G
152	4-ethylguaiacol, vanillin	G
154	Syringol (2,6- dimethoxyphenol)	S
164	Isoeugenol, eugenol	G
167*	Ethylsyringol, syrinylacetone, propiosyringone	S
168	4-methyl-2,6-dimethoxyphenol	S
178	Coniferyl aldehyde	G
180	Coniferyl alcohol, syringylethene	S, G
182	Syringaldehyde	S
194	4-propenylsyringol	S
208	Synapyl aldehyde	S
210	Synapyl alcohol	S

* Fragment ion

m/z: mass to charge ratio

S: Syringyl lignin

G: Guaiacol lignin

3

4

5

6

1 **Table 2.** Summary of the PLS-2 predictions of chemical composition from Py-mbms (6 PCs)
2 (Kelley et al., 2004b)

	Lignin	Glucose	Xylose	Mannose	Galactose	Arabinose	Rhamnose
r(CALB)	0.85	0.85	0.87	0.92	0.83	0.70	0.80
r(VALD)	0.77	0.75	0.81	0.86	0.65	0.54	0.71
RMSEC	4.60	6.20	3.40	1.40	0.40	0.50	0.10
RMSEP	5.50	8.00	4.10	1.80	0.50	0.60	0.10

3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

1 References

- 2 Agblevor, F.A., Evans, R.J., and Johnson, K.D. (1994). MOLECULAR-BEAM MASS-SPECTROMETRIC
3 ANALYSIS OF LIGNOCELLULOSIC MATERIALS .1. HERBACEOUS BIOMASS. *Journal of Analytical*
4 *and Applied Pyrolysis* 30, 125-144. doi: 10.1016/0165-2370(94)00808-6.
- 5 Alma, M.H., and Kelley, S.S. (2002). The application of pyrolysis-molecular beam mass spectrometry for
6 characterization of bark phenolysis products. *Biomass & Bioenergy* 22, 411-419.
- 7 André, N., Labbé, N., Rials, T., and Kelley, S. (2006). Assessment of wood load condition by Near Infrared
8 (NIR) spectroscopy. *Journal of Materials Science* 41, 1879-1886. doi: 10.1007/s10853-006-4433-
9 6.
- 10 Arantes, V., Qian, Y.H., Kelley, S.S., Milagres, A.M.F., Filley, T.R., Jellison, J., and Goodell, B. (2009).
11 Biomimetic oxidative treatment of spruce wood studied by pyrolysis-molecular beam mass
12 spectrometry coupled with multivariate analysis and C-13-labeled tetramethylammonium
13 hydroxide thermochemolysis: implications for fungal degradation of wood. *Journal of Biological*
14 *Inorganic Chemistry* 14, 1253-1263. doi: 10.1007/s00775-009-0569-6.
- 15 Baillères, H., Davrieux, F., and Ham-Pichavant, F. (2002). Near infrared analysis as a tool for rapid
16 screening of some major wood characteristics in a eucalyptus breeding program. *Ann. For. Sci.*
17 59, 479-490.
- 18 Braadbaart, F., and Poole, I. (2008). Morphological, chemical and physical changes during
19 charcoalification of wood and its relevance to archaeological contexts. *Journal of Archaeological*
20 *Science* 35, 2434-2445. doi: 10.1016/j.jas.2008.03.016.
- 21 Browning, B.L. (1967). *Methods of wood chemistry*. Interscience Publishers, a Division of John Wiley &
22 Sons, NY.
- 23 Davis, M.F., and Lagutaris, R.M. (2002). Comparison of syringyl/guaiacyl (S/G) ratios measured by solid
24 state ¹³C NMR, pyrolysis molecular beam mass spectrometry and thioacidolysis. *Abstracts of*
25 *Papers of the American Chemical Society* 223, U126-U126.
- 26 Davis, M.F., Tuskan, G.A., Payne, P., Tschaplinski, T.J., and Meilan, R. (2006). Assessment of Populus
27 wood chemistry following the introduction of a Bt toxin gene. *Tree Physiology* 26, 557-564.
- 28 Daystar, J.S., Venditti, R.A., Gonzalez, R., Jameel, H., Jett, M., and Reeb, C.W. (2013). Impacts of
29 Feedstock Composition on Alcohol Yields and Greenhouse Gas Emissions from the NREL
30 Thermochemical Ethanol Conversion Process. *Bioresources* 8, 5261-5278.
- 31 Dodd, L.K.D.J.G. (2002). "Handbook of Vibrational Spectroscopy," in *Handbook of Vibrational*
32 *Spectroscopy*, ed. J.M.C. P.R.Griffiths. (Chichester, UK: John Wiley & Sons Ltd.), 12.
- 33 Evans, R.J., and Milne, T.A. (1987). MOLECULAR CHARACTERIZATION OF THE PYROLYSIS OF BIOMASS .1.
34 FUNDAMENTALS. *Energy & Fuels* 1, 123-137. doi: 10.1021/ef00002a001.
- 35 Fackler, K., and Schwanninger, M. (2010). Polysaccharide degradation and lignin modification during
36 brown rot of spruce wood: a polarised Fourier transform near infrared study. *Journal of near*
37 *Infrared Spectroscopy* 18, 403-416. doi: 10.1255/jnirs.901.
- 38 French, R., and Czernik, S. (2010). Catalytic pyrolysis of biomass for biofuels production. *Fuel Processing*
39 *Technology* 91, 25-32. doi: 10.1016/j.fuproc.2009.08.011.
- 40 Himmel, M.E., Ding, S.Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W., and Foust, T.D. (2007).
41 Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science* 315,
42 804-807. doi: 10.1126/science.1137016.
- 43 Hoffmeyer, P., and Pedersen, J.G. (1995). Evaluation of density and strength of Norway spruce wood by
44 near infrared reflectance spectroscopy. *Holz als Roh- und Werkstoff* 53, 165-170. doi:
45 10.1007/BF02716418.

- 1 Hoover, C.M., Magrini, K.A., and Evans, R.J. (2002). Soil carbon content and character in an old-growth
2 forest in northwestern Pennsylvania: a case study introducing pyrolysis molecular beam mass
3 spectrometry (py-MBMS). *Environmental Pollution* 116, S269-S275. doi: 10.1016/s0269-
4 7491(01)00258-5.
- 5 Hou, S., and Li, L. (2011). Rapid Characterization of Woody Biomass Digestibility and Chemical
6 Composition Using Near-infrared Spectroscopy Free Access. *Journal of Integrative Plant Biology*
7 53, 166-175. doi: 10.1111/j.1744-7909.2010.01003.x.
- 8 Houghton, T.P., Stevens, D.M., Pryfogle, P.A., Wright, C.T., and Radtke, C.W. (2009). The Effect of Drying
9 Temperature on the Composition of Biomass. *Applied Biochemistry and Biotechnology* 153, 4-10.
10 doi: 10.1007/s12010-008-8406-x.
- 11 Jin, L., and Xu, Q. (2011). Application of near infrared spectroscopy and multivariate analysis in the
12 forest products industry. *Adv. Mater. Res. (Durnten-Zurich, Switz.)* 236-238, 1098-1102. doi:
13 10.4028/www.scientific.net/AMR.236-238.1098.
- 14 Jin, S., and Chen, H. (2007). Near-infrared analysis of the chemical composition of rice straw. *Industrial*
15 *Crops and Products* 26, 207-211. doi: <http://dx.doi.org/10.1016/j.indcrop.2007.03.004>.
- 16 Kelley, S., Rials, T., Snell, R., Groom, L., and Sluiter, A. (2004a). Use of near infrared spectroscopy to
17 measure the chemical and mechanical properties of solid wood. *Wood Science and Technology*
18 38, 257-276. doi: 10.1007/s00226-003-0213-5.
- 19 Kelley, S.S., Jellison, J., and Goodell, B. (2002). Use of NIR and pyrolysis-MBMS coupled with multivariate
20 analysis for detecting the chemical changes associated with brown-rot biodegradation of spruce
21 wood. *Fems Microbiology Letters* 209, 107-111. doi: 10.1111/j.1574-6968.2002.tb11117.x.
- 22 Kelley, S.S., Rowell, R.M., Davis, M., Jurich, C.K., and Ibach, R. (2004b). Rapid analysis of the chemical
23 composition of agricultural fibers using near infrared spectroscopy and pyrolysis molecular
24 beam mass spectrometry. *Biomass & Bioenergy* 27, 77-88. doi: 10.1016/j.biomboie.2003.11.005.
- 25 Krongtaew, C., Messner, K., Ters, T., and Fackler, K. (2010). CHARACTERIZATION OF KEY PARAMETERS
26 FOR BIOTECHNOLOGICAL LIGNOCELLULOSE CONVERSION ASSESSED BY FT-NIR SPECTROSCOPY.
27 PART I: QUALITATIVE ANALYSIS OF PRETREATED STRAW. *Bioresources* 5, 2063-2080.
- 28 Labbe, N., Harper, D., and Rials, T. (2006). Chemical structure of wood charcoal by infrared spectroscopy
29 and multivariate analysis. *Journal of Agricultural and Food Chemistry* 54, 3492-3497. doi:
30 10.1021/jf053062n.
- 31 Labbe, N., Lee, S.H., Cho, H.W., Jeong, M.K., and Andre, N. (2008a). Enhanced discrimination and
32 calibration of biomass NIR spectral data using non-linear kernel methods. *Bioresource*
33 *Technology* 99, 8445-8452. doi: 10.1016/j.biortech.2008.02.052.
- 34 Labbe, N., Rials, T.G., Kelley, S.S., Cheng, Z.M., Kim, J.Y., and Li, Y. (2005). FT-IR imaging and pyrolysis-
35 molecular beam mass spectrometry: new tools to investigate wood tissues. *Wood Science and*
36 *Technology* 39, 61-U19. doi: 10.1007/s00226-004-0274-0.
- 37 Labbe, N., Ye, X.P., Franklin, J.A., Womac, A.R., Tyler, D.D., and Rials, T.G. (2008b). ANALYSIS OF
38 SWITCHGRASS CHARACTERISTICS USING NEAR INFRARED SPECTROSCOPY. *Bioresources* 3, 1329-
39 1348.
- 40 Lestander, T.A., and Rhen, C. (2005). Multivariate NIR spectroscopy models for moisture, ash and
41 calorific content in biofuels using bi-orthogonal partial least squares regression. *Analyst* 130,
42 1182-1189. doi: 10.1039/B500103J.
- 43 Lovett, D.K., Deaville, E.R., Mould, F., Givens, D.I., and Owen, E. (2004). Using near infrared reflectance
44 spectroscopy (NIRS) to predict the biological parameters of maize silage. *Animal Feed Science*
45 *and Technology* 115, 179-187. doi: <http://dx.doi.org/10.1016/j.anifeedsci.2004.02.007>.
- 46 Lupoi, J.S., Singh, S., Simmons, B.A., and Henry, R.J. (2014). Assessment of Lignocellulosic Biomass Using
47 Analytical Spectroscopy: an Evolution to High-Throughput Techniques. *Bioenergy Research* 7, 1-
48 23. doi: 10.1007/s12155-013-9352-1.

- 1 Magrini, K.A., Follett, R.F., Kimble, J., Davis, M.F., and Pruessner, E. (2007). Using pyrolysis molecular
2 beam mass spectrometry to characterize soil organic carbon in native prairie soils. *Soil Science*
3 172, 659-672. doi: 10.1097/ss.0b013e3180d0a3a5.
- 4 Mann, D.G.J., Labbe, N., Sykes, R.W., Gracom, K., Kline, L., Swamidoss, I.M., Burris, J.N., Davis, M., and
5 Stewart, C.N. (2009). Rapid Assessment of Lignin Content and Structure in Switchgrass (*Panicum*
6 *virgatum* L.) Grown Under Different Environmental Conditions. *Bioenergy Research* 2, 246-256.
7 doi: 10.1007/s12155-009-9054-x.
- 8 Michell, A.J. (1995). Pulpwood quality estimation by near-infrared spectroscopic measurements on
9 eucalypt woods. *Appita Journal* 48, 425-428.
- 10 Nimlos, M., Brooking, E., Looker, M.J., and Evans, R.J. (2003). Biomass torrefaction studies with a
11 molecular beam mass spectrometer. *Abstracts of Papers of the American Chemical Society* 226,
12 U536-U536.
- 13 Nkansah, K., Dawson-Andoh, B., and Slahor, J. (2010). Rapid characterization of biomass using near
14 infrared spectroscopy coupled with multivariate data analysis: Part 1 yellow-poplar
15 (*Liriodendron tulipifera* L.). *Bioresource Technology* 101, 4570-4576. doi:
16 <http://dx.doi.org/10.1016/j.biortech.2009.12.046>.
- 17 Obernberger, I. (2014). Strategy for the Application of Novel Characterization Methods for Biomass Fuels:
18 Case Study of Straw. *Energy & Fuels* 28, 1041-1052. doi: 10.1021/ef402249x.
- 19 Ozaki, Y., Šašič, S., and Jiang, J. (2001). REVIEW: How can we unravel complicated near infrared
20 spectra? Recent progress in spectral analysis methods for resolution enhancement and band
21 assignments in the near infrared region. *Journal of Near Infrared Spectroscopy* 9, 63-95.
- 22 Philip Ye, X., Liu, L., Hayes, D., Womac, A., Hong, K., and Sokhansanj, S. (2008). Fast classification and
23 compositional analysis of cornstover fractions using Fourier transform near-infrared techniques.
24 *Bioresource Technology* 99, 7323-7332. doi: <http://dx.doi.org/10.1016/j.biortech.2007.12.063>.
- 25 Ragauskas, A.J., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A., Frederick, W.J.,
26 Hallett, J.P., Leak, D.J., Liotta, C.L., Mielenz, J.R., Murphy, R., Templer, R., and Tschaplinski, T.
27 (2006). The path forward for biofuels and biomaterials. *Science* 311, 484-489. doi:
28 10.1126/science.1114736.
- 29 Rousset, P., Davrieux, F., Macedo, L., and Perre, P. (2011). Characterisation of the torrefaction of beech
30 wood using NIRS: Combined effects of temperature and duration. *Biomass & Bioenergy* 35,
31 1219-1226. doi: 10.1016/j.biombioe.2010.12.012.
- 32 Sandak, J., and Sandak, A. (2011). Fourier transform near infrared assessment of biomass composition of
33 shrub willow clones (*Salix* sp.) for optimal bio-conversion processing. *Journal of near Infrared*
34 *Spectroscopy* 19, 309-318. doi: 10.1255/jnirs.950.
- 35 Sanderson, M.A., Agblevor, F., Collins, M., and Johnson, D.K. (1996). Compositional analysis of biomass
36 feedstocks by near infrared reflectance spectroscopy. *Biomass and Bioenergy* 11, 365-370. doi:
37 [http://dx.doi.org/10.1016/S0961-9534\(96\)00039-6](http://dx.doi.org/10.1016/S0961-9534(96)00039-6).
- 38 Savitzky, A., and Golay, M.J.E. (1964). SMOOTHING + DIFFERENTIATION OF DATA BY SIMPLIFIED LEAST
39 SQUARES PROCEDURES. *Analytical Chemistry* 36, 1627-&. doi: 10.1021/ac60214a047.
- 40 Schwanninger, M., Rodrigues, J.C., and Fackler, K. (2011). A review of band assignments in near infrared
41 spectra of wood and wood components. *Journal of near Infrared Spectroscopy* 19, 287-308. doi:
42 10.1255/jnirs.955.
- 43 Sjöström, E., and Alen, R. (1999). *Analytical methods in wood chemistry, pulping, and papermaking*.
44 Springer.
- 45 Sluiter, J.B., Ruiz, R.O., Scarlata, C.J., Sluiter, A.D., and Templeton, D.W. (2010). Compositional Analysis of
46 Lignocellulosic Feedstocks. 1. Review and Description of Methods. *Journal of Agricultural and*
47 *Food Chemistry* 58, 9043-9053. doi: 10.1021/jf1008023.

- 1 Smith-Moritz, A.M., Chern, M., Lao, J., Sze-To, W.H., Heazlewood, J.L., Ronald, P.C., and Vega-Sanchez,
2 M.E. (2011). Combining multivariate analysis and monosaccharide composition modeling to
3 identify plant cell wall variations by Fourier Transform Near Infrared spectroscopy. *Plant*
4 *Methods* 7. doi: 10.1186/1746-4811-7-26.
- 5 Sykes, R., Kodrzycki, B., Tuskan, G., Foutz, K., and Davis, M. (2008). Within tree variability of lignin
6 composition in Populus. *Wood Science and Technology* 42, 649-661. doi: 10.1007/s00226-008-
7 0199-0.
- 8 Sykes, R., Yung, M., Novaes, E., Kirst, M., Peter, G., and Davis, M. (2009). "High-Throughput Screening of
9 Plant Cell-Wall Composition Using Pyrolysis Molecular Beam Mass Spectroscopy," in *Biofuels*, ed.
10 J.R. Mielenz. Humana Press), 169-183.
- 11 Templeton, D.W., Scarlata, C.J., Sluiter, J.B., and Wolfrum, E.J. (2010). Compositional Analysis of
12 Lignocellulosic Feedstocks. 2. Method Uncertainties. *Journal of Agricultural and Food Chemistry*
13 58, 9054-9062. doi: 10.1021/jf100807b.
- 14 Thygesen, L. (1994). Determination of dry matter content and basic density of Norway spruce by near
15 infrared reflectance and transmittance spectroscopy. *Journal of Near Infrared Spectroscopy* 2,
16 127-135.
- 17 Tsuchikawa, S. (2007). A review of recent near infrared research for wood and paper. *Appl. Spectrosc.*
18 *Rev.* 42, 43-71. doi: 10.1080/05704920601036707.
- 19 Tsuchikawa, S., Murata, A., Kohara, M., and Mitsui, K. (2003). Spectroscopic monitoring of biomass
20 modification by light-irradiation and heat treatment. *Journal of near Infrared Spectroscopy* 11,
21 401-405.
- 22 Tucker, M., Nguyen, Q., Eddy, F., Kadam, K., Gedvilas, L., and Webb, J. (2001). Fourier transform infrared
23 quantitative analysis of sugars and lignin in pretreated softwood solid residues. *Applied*
24 *Biochemistry and Biotechnology* 91-93, 51-61. doi: 10.1385/ABAB:91-93:1-9:51.
- 25 Tuskan, G., West, D., Bradshaw, H.D., Neale, D., Sewell, M., Wheeler, N., Megraw, B., Jech, K., Wiselogel,
26 A., Evans, R., Elam, C., Davis, M., and Dinus, R. (1999). Two high-throughput techniques for
27 determining wood properties as part of a molecular genetics analysis of hybrid poplar and
28 loblolly pine. *Applied Biochemistry and Biotechnology* 77-9, 55-65. doi: 10.1385/abab:77:1-3:55.
- 29 Via, B.K., Adhikari, S., and Taylor, S. (2013). Modeling for proximate analysis and heating value of
30 torrefied biomass with vibration spectroscopy. *Bioresource Technology* 133, 1-8. doi:
31 10.1016/j.biortech.2013.01.108.
- 32 Wei, H., Xu, Q., Taylor, L.E., Baker, J.O., Tucker, M.P., and Ding, S.Y. (2009). Natural paradigms of plant
33 cell wall degradation. *Current Opinion in Biotechnology* 20, 330-338. doi:
34 10.1016/j.copbio.2009.05.008.
- 35 Wolfrum, E., and Sluiter, A. (2009). Improved multivariate calibration models for corn stover feedstock
36 and dilute-acid pretreated corn stover. *Cellulose* 16, 567-576. doi: 10.1007/s10570-009-9320-2.
- 37 Workman, J.J., Jr. (2001). Infrared and Raman spectroscopy in paper and pulp analysis. *Appl. Spectrosc.*
38 *Rev.* 36, 139-168. doi: 10.1081/ASR-100106154.
- 39 Xu, F., Yu, J.M., Tesso, T., Dowell, F., and Wang, D.H. (2013). Qualitative and quantitative analysis of
40 lignocellulosic biomass using infrared techniques: A mini-review. *Applied Energy* 104, 801-809.
41 doi: 10.1016/j.apenergy.2012.12.019.
- 42 Yang, Z., Jiang, Z.H., Fei, B.H., and Qin, D.C. (2007). Discrimination of wood biological decay by soft
43 independent modeling of class analogy (SIMCA) pattern recognition based on principal
44 component analysis. *Spectroscopy and Spectral Analysis* 27, 686-690.
- 45 Yeh, T.-F., Chang, H.-M., and Kadla, J.F. (2004). Rapid Prediction of Solid Wood Lignin Content Using
46 Transmittance Near-Infrared Spectroscopy. *Journal of Agricultural and Food Chemistry* 52, 1435-
47 1439. doi: 10.1021/jf034874r.

1 Zhou, G., Taylor, G., and Polle, A. (2011). FTIR-ATR-based prediction and modelling of lignin and energy
2 contents reveals independent intra-specific variation of these traits in bioenergy poplars. *Plant*
3 *Methods* 7, 9.

4 Ziebell, A.L., Barb, J.G., Sandhu, S., Moyers, B.T., Sykes, R.W., Doepcke, C., Gracom, K.L., Carlile, M.,
5 Marek, L.F., Davis, M.F., Knapp, S.J., and Burke, J.M. (2013). Sunflower as a biofuels crop: An
6 analysis of lignocellulosic chemical properties. *Biomass & Bioenergy* 59, 208-217. doi:
7 10.1016/j.biombioe.2013.06.009.

8