Chemometrics: An Important Tool for the Modern Chemist, an Example from Wood-Processing Chemistry

Petteri Malkavaara*,†

Laboratory of Applied Chemistry, Department of Chemistry, University of Jyväskylä, P. O. Box 35, FIN-40351, Jyväskylä, Finland

Raimo Alén‡

Laboratory of Applied Chemistry, Department of Chemistry, University of Jyväskylä, P. O. Box 35, FIN-40351, Jyväskylä, Finland

Erkki Kolehmainen§

Laboratory of Organic Chemistry, Department of Chemistry, University of Jyväskylä, P. O. Box 35, FIN-40351, Jyväskylä, Finland

This study briefly outlines the idea of principal component analysis and cross-correlation calculations (applied chemometrics) and presents an illustrative example from wood-processing chemistry. The applicability of chemometric data analysis was demonstrated by investigating the various structural changes that take place in dissolved and degraded lignin ("kraft lignin") during laboratory-scale kraft pulping of Scots pine (*Pinus sylvestris*) and silver birch (*Betula pendula*). The structural data (^{31}P NMR and size exclusion chromatographic data) on kraft lignin were further processed by chemometric multivariate techniques (PCA and 2DCC), confirming, for example, that the cleavage of β-aryl ether structures, the most prominent linkages between monomeric units, is directly related to the decrease in the average molecular mass of lignin.

INTRODUCTION

The structural constituents of wood are polysaccharides (cellulose and hemicelluloses) and lignin.[1] During kraft pulping, native lignin is predominantly degraded by various cleavage reactions of aryl ether linkages into hydrophilic fragments with an increasing amount of phenolic hydroxyl groups present as phenolates. All these phenomena are responsible for the prominent dissolution of lignin material ("kraft lignin") in the cooking liquor ("black liquor"). Studies with model substances representing various structural units in lignin have largely clarified the delignification reactions in kraft pulping. In addition, a significant formation of carbohydrate-derived degradation products, aliphatic carboxylic acids, is formed by various alkali-catalyzed reactions.

Chemometrics, multivariate techniques in particular, have attracted increasing attention in chemical research. The number of citations in the literature has grown steadily during the past decade. For instance, the large number of recently published review articles clearly indicates the adoption of these modern methods as useful analytical tools.[2] Chemometrics are increasingly being used in some of the more novel branches of chemistry, such as synthetometrics, which deals with the use of chemometrics in organic synthesis and combinatorial chemistry.[3]

Principal component analysis (PCA) is the basis for a multitude of multivariate analysis techniques. In the matrix form, PCA can be expressed as

$$X = TP' + E$$

where X describes the original data, T is a so-called score matrix, P′ is a loading matrix, and E describes the residual error (i.e., noise). In this type of analysis, the score values represent the projections of the original objects (i.e., samples) in a given principal coordinate system defined by the loading vectors calculated from the original variables. If these variables represent, for example, NMR spectral data, then each loading vector can be plotted against an original NMR scale (in the form of a "loading subspectrum") and thus interpreted like a NMR spectrum. In addition, the resonances with the highest loading values mean in this case that there are changes in the original spectra in these areas. Samples obtaining high score values in that principal component can be concluded as having the largest amount of these changes.

Both 1D and 2D NMR spectroscopic methods have been widely used for the structural characterization of lignin.[4−6] Although the most common nuclei have been ^{1}H and ^{13}C, ^{31}P can also be used to obtain detailed information, especially with regard to the different type of hydroxyl groups in lignin.[7] The most spectacular development, which has revolutionized practically all branches of NMR spectroscopy, is without doubt the concept of two-dimensional (2D) NMR spectroscopy.[8] Several attempts have been made to extend this concept to the vibrational region of the spectrum.[9,10] A more general approach to multidimensional spectroscopy is to
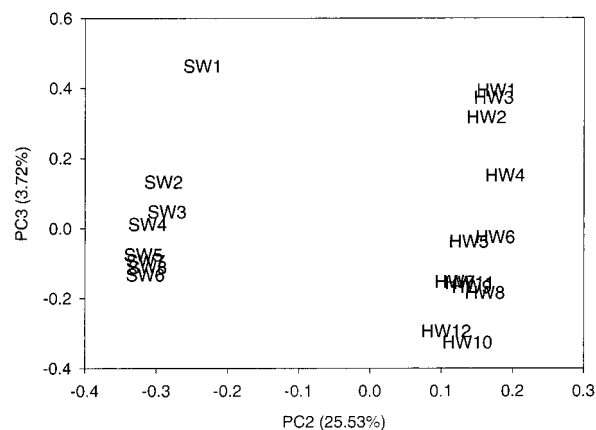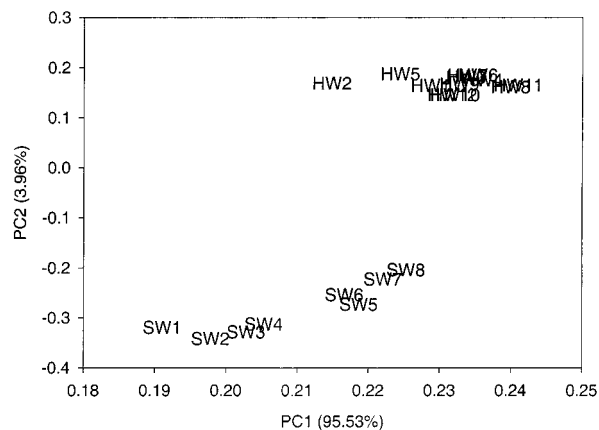
---

* To whom correspondence should be addressed.
† Phone: +358-14-2602566. Fax: +358-14-2602581. E-mail pmalkava@jyu.fi.
‡ Phone: +358−14−2602562. Fax: +358-14-2602581. E-mail: ralen@jyu.fi.
§ Phone: +358-14-2602670. Fax: +358-14-2602501. E-mail: ekolehma@jyu.fi.

CHEMOMETRICS AND WOOD-PROCESSING CHEMISTRY

J. Chem. Inf. Comput. Sci., Vol. 40, No. 2, 2000 **439**

Table 1. Cooking Data

Pine Kraft Pulping

heating-up time (from 85 to 175 °C): 90 min
liquor-to-wood ratio: 5.0 L/kg
alkali charge: 35% active alkali (as NaOH) on wood, sulfidity 30%

time at max T (min)	10	40	70	100	130	160	190	210
sample	SW1	SW2	SW3	SW4	SW5	SW6	SW7	SW8

Birch Kraft Pulping

heating-up time (from 80 to 170 °C): 90 min
liquor-to-wood ratio: 4.5 L/kg
alkali charge: 23.2% active alkali (as NaOH) on wood, sulfidity 33.4%

time at max T (min)	0	5	10	20	30	40	50	60
sample	HW1	HW2	HW3	HW4	HW5	HW6	HW7	HW8
time at max T (min)	80	100	140	200				
sample	HW9	HW10	HW11	HW12				
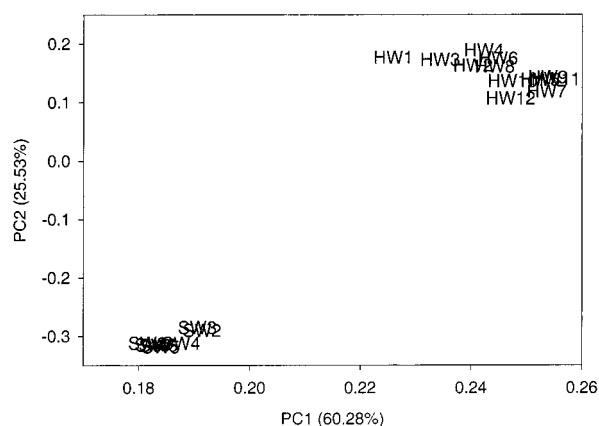
Figure 1. Score plot of the first and second PCs, NMR data.

Figure 2. Score plot of the first and second PCs, SEC data.

Figure 3. Score plot of the second and third PCs, NMR data.
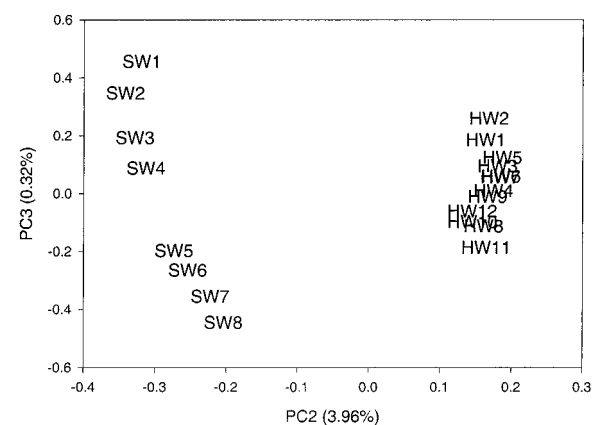
Figure 4. Score plot of the second and third PCs, SEC data.

apply statistical cross-correlation calculations, based on the chemometric correlation calculations in one dimension, to the data. The results from these 2D cross-correlation calculations would highlight the simultaneous changes in the correlated data and offer a unique tool to assign correlations to the spectral region. In addition, the statistical cross-correlation approach is not limited to spectroscopy: any relevant analysis data, regardless of its type, can be subjected to the calculations.

Multivariate analysis techniques such as PCA, principal component regression (PCR), and projection to latent structures (PLS) have already in some applications proved to be useful tools for the evaluation of spectral and chemical data on different pulps.[11−13] In the previous investigations,[14−17] the lignin content and the carbohydrate composition of both

intact wood and pulp samples were obtained by applying multivariate analysis techniques (PCA, PCR, and PLS) to various spectral data obtained mainly on the basis of infrared (IR) and cross-polarization/magic angle spinning solid-state ^{13}C nuclear magnetic resonance (CP/MAS ^{13}C NMR) determinations.

A multivariate chemometric analysis has been performed to find out what relationships subsist between the relevant combustion properties (burning time and swelling) of softwood and hardwood kraft black liquors and the chemical structure of the dissolved lignin in these liquors.[18] The lignin material was characterized in terms of NMR and average molecular mass data. ^{13}C and ^{31}P NMR spectra were used to reveal the total amount of different aliphatic and phenolic hydroxyl groups. On the basis of all the analytical data, a
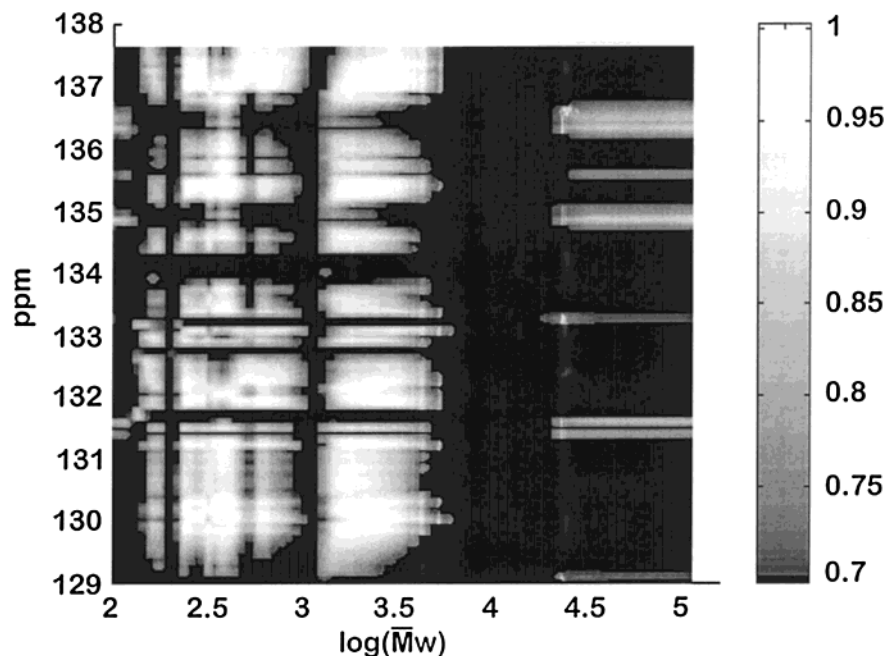
Figure 5. 2D cross-correlation image with a correlation threshold of 0.7.

relatively accurate model for predicting the combustion properties of kraft black liquors was developed.

To analyze the different types of linkages connecting the phenylpropane units in kraft lignins, the detailed ^{31}P NMR spectral multivariate analysis was performed.[19] The results clearly confirmed the cleavage of *β*-aryl-ethers (*β-O*-4 structures with adjacent α-hydroxyl group) in native lignin during kraft pulping. In addition, it could be demostrated that the *erythro* isomers were more amenable to alkaline degradation than the *threo* forms.

The aim of this study was to demonstrate the usefulness of the chemometric approach in analyzing simultaneous changes in the chemical structure and the average molecular masses of kraft lignins during delignification. Lignin fragments were characterized by means of ^{31}P NMR spectroscopy and size exclusion chromatography (SEC).

EXPERIMENTAL SECTION

Samples. Laboratory-scale pulping experiments on screened industrial chips of Scots pine (*Pinus sylvestris*) and silver birch (*Betula pendula*) were carried out in a battery of stainless steel autoclaves, each charged with 130 g of oven-dry chips. The autoclaves were rotated in a decene bath under the conditions shown in Table 1. For the analytical determinations, the kraft lignin sample in each case was prepared as follows.[20] First, the sample was precipitated from the black liquor with 2 M hydrochloric acid (pH to 2), separated from the solution by centrifugation, washed thoroughly with acidified water, and finally air-dried at room temperature.

NMR Spectroscopy. Kraft lignin samples were phosphitylated, and their ^{31}P NMR spectra were measured according to the earlier published procedures[18,21] using 2,2-bis(4-hydroxyphenol)propane (bisphenol A) as an internal standard. Measurements were carried out at 273 K with a Bruker Avance DRX 500 FT NMR spectrometer operating at 202.47 MHz using an inverse-gated decoupling sequence.

Size Exclusion Chromatography (SEC). A SEC system was used to determine the average molecular masses of the

lignin preparations. For these determinations, a small amount of each sample (1−2 mg) was dissolved in tetrahydrofuran and analyzed as described elsewhere.[22]

Data Analysis. ^{31}P NMR and SEC data were subjected to the PCA using a singular value decomposition.[23] Prior to the analysis, the NMR data (20 spectra, 1400 points each) were mean-centered and scaled to the unit variance. The SEC data (20 chromatograms, measured at 245 nm, 770 points each) were normed to the unit integral.

For the point-for-point 2D cross-correlation (2DCC) calculations, four PCA models (pine NMR, pine SEC, birch NMR, and birch SEC) were first computed and the data then reconstructed from the significant principal components (PCs), two in the case of the NMR data and one in case of the SEC data. The coefficient of determination for the models in each case was 95.10% and 95.16% for the pine and birch NMR, respectively, and 98.83% and 99.68% for the pine and birch SEC, respectively. A 2DCC calculation[24] was applied to these combined NMR (pine plus birch) and SEC (pine plus birch) PC data reconstructs, and absolute coefficient values above 0.7 were taken into account. The correlation coefficients (r_{ij}) were calculated by the following equation[24]

$$R_{ij} = \frac{n\sum x_j y_i - (\sum x_j)(\sum y_i)}{\sqrt{[n\sum x_j^2 - (\sum x_j)^2][n\sum y_i^2 - (\sum y_i)^2]}}$$

where R is the matrix consisting of correlation coefficients r_{ij}, x is the SEC, and y is the NMR PC data reconstruct. All the computations were carried out on a personal computer using a MATLAB software package.[25]

RESULTS AND DISCUSSION

PCA of the ^{31}P NMR and SEC Data. The score plots, describing the classification of the samples in the principal coordinates (explained variance of each principal component

CHEMOMETRICS AND WOOD-PROCESSING CHEMISTRY

J. Chem. Inf. Comput. Sci., Vol. 40, No. 2, 2000 **441**

is presented in parentheses), are presented in Figures 1−4. The results obtained from both types of data were in close agreement and yielded a valid classification of the samples, thereby confirming the applicability of this kind of analysis. In addition, both sets of results indicated feedstock-related differences in the delignification rate, seen in the different groupings of the pine and birch samples.

2DCC of ³¹P NMR and SEC PC Data Reconstructs. A 2D cross-correlation image (SEC data as abscissa and ³¹P NMR data as ordinate) with a correlation threshold of 0.7 is presented in Figure 5. Although it is safe to conclude that changes in the lignin molecular mass within the mass range 1000−3200 Da are highly correlated to changes in all types of lignin structural units, it is also clear that the only marked correlations in the higher-molecular-mass region were with resonances around 135 and 137 ppm. These resonances are typical of those of *threo-* and *erythro-β*-aryl ether structures, respectively.[19] Cross-correlations around 133 ppm were due to the impurities in the phosphitylating reagent. There were also two thin cross-correlation lines at 131.4 and 131.6 ppm that could be attributed to a minor molecular mass increase induced by the condensed lignin structures (approximative chemical shift range 131.2−131.9 ppm).

CONCLUSIONS

This kind of chemometric approach based on several relevant analytical procedures provided a versatile tool for understanding the simultaneous changes in lignin structure and molecular mass. These simultaneous changes, although very widely accepted as a fact, have been experimentally very difficult to detect. It was confirmed, for example, that in the kraft cooking the cleavage of *β*-aryl ether structures of lignin was indeed directly related to the decrease in its average molecular mass. The structural differences between the softwood and hardwood lignin structures were not only responsible for the notable differences in the structural changes but also for the different rate of changes in the molecular mass as well.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Sjöström, E. *Wood Chemistry—Fundamentals and Applications*, 2nd ed.; Academic Press: San Diego, CA, 1993; p 145.
(2) Lavine, B. Chemometrics. *Anal. Chem.* **1998**, *70*, 209R−228R and references therein.
(3) Carlson, R. Synthetometrics: Recent Developments. *Chemom. Intell. Lab. Syst.* **1992**, *14*, 41−56.
(4) Lundqvist, K. *Proton (¹H) NMR Spectroscopy*; In *Methods in Lignin Chemistry*; Lin, S. Y., Dence, C. W., Eds.; Springer-Verlag: Berlin, 1992; pp 242−249 and references therein.
(5) Robert, D. *Carbon-13 Nuclear Magnetic Resonance Spectrometry*; In *Methods in Lignin Chemistry*; Lin, S. Y., Dence, C. W., Eds.; Springer-Verlag: Berlin, 1992; pp 250−273 and references therein.
(6) Ede, R.; Ralph, J. Assignment of 2D TOCSY Spectra of Lignins: The Role of Lignin Model Compounds. *Magn. Reson. Chem.* **1996**, *34*, 261−268 and references therein.
(7) Argyropoulos, D. S. ³¹P NMR in Wood Chemistry: A Review of Recent Progress. *Res. Chem. Intermed.* **1995**, *21* (3−5), 373−395 and references therein.
(8) Günther, H. *NMR Spectroscopy—Basic Principles, Concepts, and Applications in Chemistry*, 2nd ed.; John Wiley & Sons: Chichester, 1995; pp xvii, 274.
(9) Noda, I. Two-Dimensional Infrared Spectroscopy. *J. Am. Chem. Soc.* **1989**, *111*, 8116−8118.
(10) Noda, I. Two-Dimensional Infrared (2D IR) Spectroscopy: Theory and Applications. *Appl. Spectrosc.* **1990**, *44(4)*, 550−561.
(11) Lennholm, H.; Wallbäcks, L.; Iversen, T. A ¹³C-CP/MAS-NMR-Spectroscopic Study of the Effect of Laboratory Kraft Cooking on Cellulose Structure. *Nord. Pulp Pap. Res. J.* **1995**, *10* (1), 46−50.
(12) Broderick, G.; Paris, J.; Valade, J. L.; Wood, J. Applying Latent Vector Analysis to Pulp Characterization. *Pap. Puu* **1995**, *77* (6−7), 410−418.
(13) Michell, A. Vibrational Spectroscopy—a Rapid Means of Estimating Plantation Pulpwood Quality. *Appita* **1994**, *47* (1), 29−37.
(14) Wallbäcks, L.; Edlund, U.; Nordén, B.; Berglund, I. Multivariate Characterization of Pulp Using Solid-State ¹³C NMR, FTIR, and NIR. *Tappi J.* **1991**, *74* (10), 201−206.
(15) Schimleck, L.; Wright, P.; Michell, A.; Wallis, A. Near-Infrared Spectra and Chemical Compositions of *E. globulus* and *E. nitens* Plantation Woods. *Appita* **1997**, *50* (1), 40−46.
(16) Malkavaara, P.; Alén, R. A Spectroscopic Method for Determining Lignin Content of Softwood and Hardwood Kraft Pulps. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 287−292.
(17) Olsson, R.; Tomani, P.; Karlsson, M.; Joseffson, T.; Sjöberg, K.; Björklund, C. Multivariate Characterization of Chemical and Physical Descriptors in Pulp Using NIRR. *Tappi J.* **1995**, *78* (10), 158−166.
(18) Malkavaara, P.; Alén, R.; Kolehmainen, E. Multivariate Correlation between ¹³C and ³¹P NMR Spectral Data on Dissolved Lignin and the Combustion Properties of Kraft Black Liquors. *Magn. Reson. Chem.* **1999**, *37* (6), 407−412.
(19) Malkavaara, P.; Alén, R.; Kolehmainen, E. Multivariate Analysis of ³¹P NMR Data: A Potential Analytical Method for the Wood Processing Industry. *Proceedings of the Pre-Symposium Workshop to the 9th Intl. Symposium on Wood and Pulping Chemistry: Advances in Microscopy and NMR Spectroscopy of Lignocellulosic Materials*; Québec City, Canada, 1997; pp 179−181.
(20) Alén, R.; Hartus, T. UV Spectrophotometric Determination of Lignin from Alkaline Pulping Liquors. *Cellul. Chem. Technol.* **1988**, *22* (6), 613−618.
(21) Argyropoulos, D. S. Quantitative Phosphorus-31 NMR Analysis of Lignins, a New Tool for the Lignin Chemist. *J. Wood Chem. Technol.* **1994**, *14* (1), 45−63.
(22) Hyötyläinen, J.; Knuutinen, J.; Malkavaara, P.; Siltala, J. Pyrolysis-GC-MS and CuO-oxidation-HPLC in the Characterization of HMMs from Sediments and Surface Waters Downstream of a Pulp Mill. *Chemosphere* **1998**, *36* (2), 297−314.
(23) Jolliffe, I. T. *Principal Component Analysis*; Springer: New York, 1986.
(24) Barton, F.; Himmelsbach, D.; Duckworth, J.; Smith, M. Two-Dimensional Vibration Spectroscopy: Correlation of Mid- and Near-Infrared Regions. *Appl. Spectrosc.* **1992**, *46* (3), 420−429.
(25) *MATLAB for Windows*, version 4.2c; The MathWorks, Inc.: Natick, MA, 1994.

CI990444I