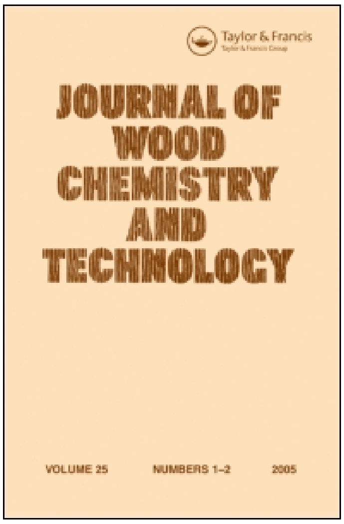
This article was downloaded by:
On: *25 January 2011*
Access details: *Access Details: Free Access*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Wood Chemistry and Technology

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597282

Multivariate Data Analysis of *In Situ* Pulp Kinetics Using <i>13</i>C CP/MAS NMR

Lars Wallbäcks[a]; Ulf Edlund[a]; Bo Nordén[a]
[a] NMR Research Group, Department of Organic Chemistry, Umeå University, Umeå, Sweden

To cite this Article Wallbäcks, Lars , Edlund, Ulf and Nordén, Bo(1989) 'Multivariate Data Analysis of *In Situ* Pulp Kinetics Using <i>13</i>C CP/MAS NMR', Journal of Wood Chemistry and Technology, 9: 2, 235 — 249
To link to this Article: DOI: 10.1080/02773818908050296
URL: http://dx.doi.org/10.1080/02773818908050296

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

JOURNAL OF WOOD CHEMISTRY AND TECHNOLOGY, 9(2), 235-249 (1989)

MULTIVARIATE DATA ANALYSIS OF IN SITU PULP KINETICS USING ^{13}C CP/MAS NMR.

Lars Wallbäcks, Ulf Edlund and Bo Nordén
NMR Research Group, Department of Organic Chemistry
Umeå University, S-901 87, Umeå, Sweden.

ABSTRACT

^{13}C CP/MAS NMR spectra of birch pulp samples were measured at regular intervals during the kraft pulping process. General multivariate data analysis methods based on principal components were used to extract the spectral information content and to relate these components to chemical descriptors such as lignin content. It was shown that this approach has an improved predictive ability relative to traditional methods. Model parameters could also be used to construct subspectra with independent information.

INTRODUCTION

A variety of spectroscopic techniques have earlier been used in order to separate concepts of crystallinity, morphology, mobility, lignin content etc. in cellulose samples. However, all measurements are hampered by the fact that contributions from lignin, cellulose and other carbohydrates are superimposed in overlapping bands. An important facet of wood research would be to design specific experiments that could afford subspectra of the different chemical components. ^{13}C CP/MAS NMR is a relatively new technique that has shown great promise in elucidating polymeric order/disorder. In the cellulose chemistry field, several reports have recently appeared where ^{13}C CP/MAS NMR data have been interpreted in terms of morphology changes[1].

As expected for complicated samples, resolution is a strongly limiting factor for interpreting spectroscopic results and there is also a controversy how to assign the broad signals, i e what areas in the biopolymer are responsible for the broad components[2]. To partly overcome these problems, variation in molecular dynamics has been proposed as an approach to separate signals from crystalline and non-crystalline areas. Horii[3] showed that ^{13}C T_1's were considerably longer for crystalline than for non-crystalline cellulose. A similar dynamic separation, based on $T_{1\rho}^{H}$ differences, has recently been suggested for the construction of subspectra of various wood products[4].

Due to the low resolution in ^{13}C CP/MAS NMR of pulp samples, all studies reported so far have only discussed spectral changes in regions of low signal overlap. In a recent study of the delignification during pulping of white spruce, the Klason lignin content was correlated to the normalized intensity of the aromatic C3, C4 carbon signals of lignin[5]. Due to the low relative intensity of this signal, pulp samples having lignin content below 5 % did not have measurable resonances at that spectral region. It is obvious that if spectral changes due to the methoxy carbons, as well as from the C_α, C_β and C_γ carbons could be taken into account, a better description of the delignification process could be achieved. This approach, however, would demand the use of multivariate data analysis methods.

Multivariate approaches, based on principal components data analysis, have earlier been applied on experimental data of complex samples of this kind[6,7]. Partial least squares (PLS) data analysis have been used to the multivariate calibration problem of determining concentrations of ligninsulfonate, humic acid and a detergent containing whitener, by molecular fluorescence[8]. This two-block approach has also been used to give valuable information about properties of peat using FT-IR intensity values[9]. FT-IR data at selected wavenumbers were correlated with various peat descriptors, like calorific value, degree of decomposition, content of amino acids etc. using *PLS* modelling. These data analytical methods have been developed to handle causal and complicated relations within a data set, and they are also well suited to handle large data sets, where relatively few objects are described by many variables.

If the multivariate methodology is applied on the ^{13}C CP/MAS NMR data of pulp samples, we will have a most efficient tool to extract information concerning the pulping process. In this first report we will demonstrate the methodology using birch kraft pulp samples where Klason lignin content is used as the only pulp descriptor.

The major concern and also the strength of the presented data analysis approach, is to eliminate factors that are associated with spectral processing. In future reports we will include "many-block" PLS techniques, using FT-IR diffuse reflectance data, NIR, an extended descriptor matrix containing carbohydrate analysis data etc. The purpose of this project is to get as a detailed description as possible of the major pulping stages, by identifying the spectral regions which are of interest and how these observed signal changes are related to other accessible descriptors used to probe the pulping process.

METHODS

Samples

The kraft pulping process (birch) is probed by 40 samples collected at regular time intervals (5-320 min), the zero point is 80 °C with an increase by 0.8 °C/minute until the maximum temperature of 170 °C is reached. The samples were prepared at two different occasions (with 20 samples each), at MoDo research laboratory, Örnsköldsvik, Sweden. To confirm reproducibility, eight of the first measurements were repeated once, using new material from the given batch. Twenty additional batches were produced in order to get a more detailed description of the main delignification period. The second series then contains the first 28 samples in addition to the 23 samples from 20 new pulp batches. The lignin content was analyzed according to standard methods for Klason-lignin.

NMR

^{13}C CP/MAS NMR spectra were obtained by using a Bruker MSL 100 spectrometer operating at 25.2 MHz. Air-dried samples were packed into perdeuterated PMMA rotors and the spinning rate was approximately 3.5-4 KHz. The contact time was 1 ms and the recycle time was set to 2.5 sec. The number of scans varied between 2000 and 30 000. A 20 Hz line-broadening was introduced in each spectrum, and data were analyzed using both manually phase corrected spectra and using magnitude spectra[10]. When duplicate samples were run it was found that the use of magnitude spectra gave more consistent results. In order to overcome the problem of intensity differences caused by differences in sample volume and number of scans, spectra were normalized to a constant integration sum. For the data analysis the region from 5 to 184 ppm was used for all samples. Intensity values in 93 evenly distributed data points from the 51 spectra (11 duplicates) constituted the ^{13}C CP/MAS NMR data matrices. In the calculation based on data from the carbohydrate spectral region (60-115 ppm), a separate normalization of these 30 variables was performed. Spectra of samples withdrawn at two different stages in the pulping process (5 min and 220 min, respectively) are shown in figure 1.

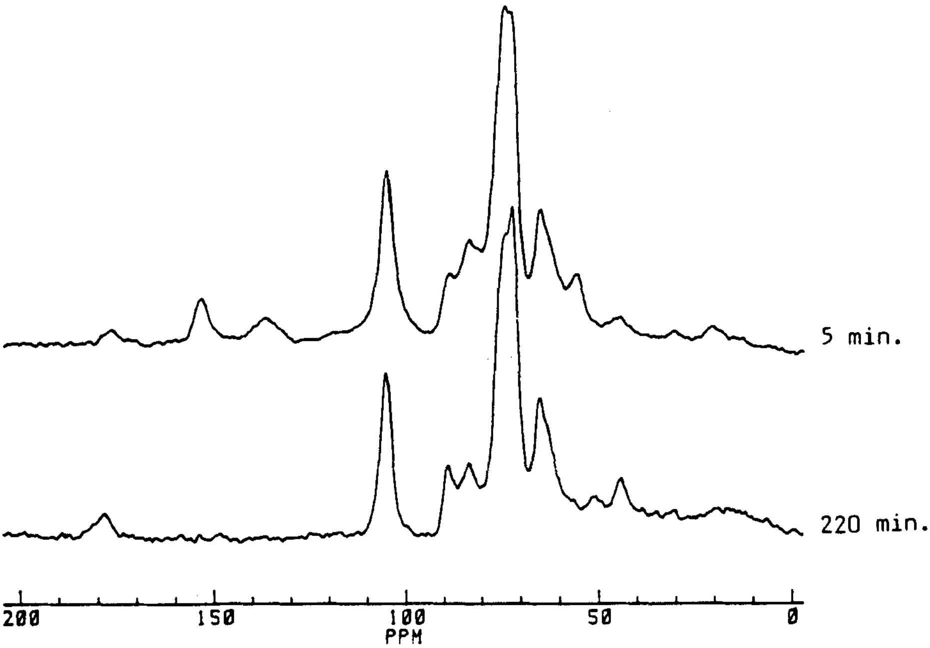


FIGURE 1 ^{13}C CP/MAS NMR spectra of birch pulp samples taken at 5 min. and 220 min.

Data analysis

Spectroscopic data are usually more complex than ordinary "wet" chemical analyses data. This might lead to difficulties in the assignments of spectral regions which are of relevance for the actual problem. In our approach the intensity values at several chemical shift values (I) of a set of N ^{13}C CP/MAS spectra are measured. Thereafter a multivariate method is used to create a calibration model. The calibration model can then be used to predict lignin content (Y) of new pulp samples, which have not been chemically analyzed. The Y vector could also be expanded to other pulp descriptors of relevance like viscosity, R5, kappa value, carbohydrate analysis data etc.

Traditional methods.

Usually the calibration problem is tackled by applying multiple regression (MR), where separate MR are made for each of the dependent variables. However, spectroscopic data are not suitable for such an analysis since the variables are often strongly correlated, leading to an unstable statistical solution. Alternatively one can reduce the X matrix to a set of uncorrelated principal components and thereafter apply normal multiple regression using the component scores as X variables.

Principal Components/ Partial Least Squares Data Analysis.

The experimental data were analyzed using principal components (PC) and partial least squares (PLS) data analysis[11,12,13,14]. The descriptor matrix X contains the elements x_{ik} (intensities) where index i is used for the descriptors (chemical shifts) and index k for the samples. From this data matrix the number of product terms, A, and the parameters, b_{ia} and t_{ak} in eqn. 1 are estimated by minimizing the squared residuals e_{ik}.

$$x_{ik} = \bar{x}_i + \sum_{a=1}^{A} b_{ia} * t_{ak} + e_{ik} \qquad (1)$$

In this model $\bar{x}_i$ and b_{ia} are constants, which only depend on the ^{13}C intensities. The t_{ak} term is the sample dependent parameter. Deviations from the model are expressed by the residuals e_{ik}.

First a model A=0 is fitted to the data, which means that each intensity is given as its mean value $\bar{x}_i$. Then the $\bar{x}_i$ value for each variable is subtracted from the matrix element x_{ik} thus giving residuals of dimension zero. If these residuals now contain systematic information, $b_{ia} * t_{ak}$ is estimated. Whether the residuals contain information or not is determined by cross-validation[15], which is an efficient strategy to determine if a component is significant or not. With this method, the predictive strength of the a:th cross-term is checked by first deleting approximately 25 % of the basis data set. The y values of the deleted samples are then predicted and the sum of squared difference (ss) between the observed and calculated y values is calculated. Another 25 % is then omitted, an additional ss is calculated etc. until every sample has been kept out just once. If the sum of the four separate ss is smaller than the complete ss of the total Y matrix, the component is taken as significant.

The "sensitivity" of the shift variables, b_{ia}, determines the direction of the hyperplane in the A-dimensional space. By plotting the loadings against the shift it will be possible to create "subspectra"[7], determined by the systematic variation in the data matrix. A geometrical interpretation of PCA is presented in figure 2.

The correlation between a ^{13}C CP/MAS NMR matrix and a chemical analysis (CA) matrix is quantified by using a soft modelling method called partial least squares data analysis (PLS). The PLS method is an extension of PCA in that two matrices are analyzed simultaneously, one matrix (X) containing the independent set of variables (NMR measurements), and the other matrix (Y) containing the dependent variables (lignin content). Subsequently, PCA-like algorithms are used for both matrices and a regression relationship is established between the matrix components. The relationship is later used for predictive purposes for new samples, where only the independent matrix variables are measured.

RESULTS AND DISCUSSION

Principal Components Data Analysis

As an initial analysis, a ^{13}C intensity matrix including 28 spectra from the first 20 samples (spanning the time domain 5-320

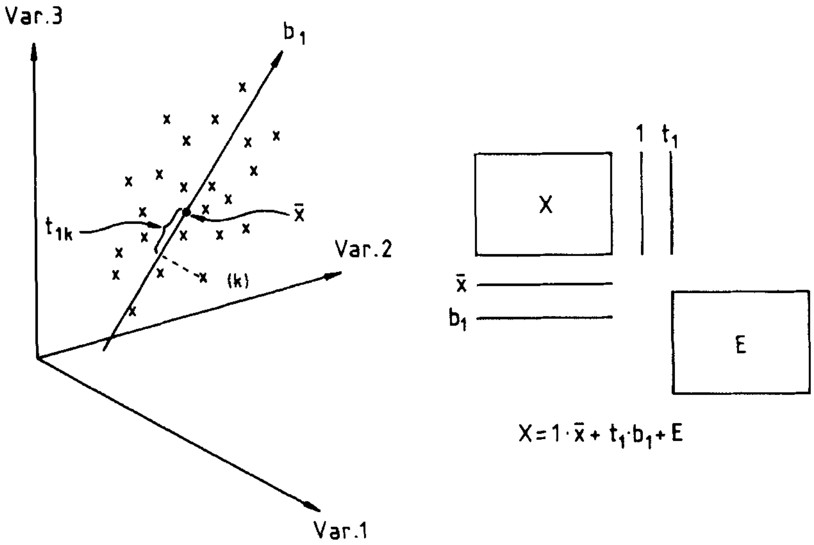


FIGURE 2 The data are least squares modelled by a line through the average $\bar{x}$. This line is the first principal component (PC) and its equation is defined by the loading vector b_1. The projection of point k on the line lies on the distance t_{1k} from $\bar{x}$.

min.) was compressed using principal components data analysis. At this stage, no attempts were made to optimize the sampling interval or to make any variable selection i.e. choice of shift values. The result of the first analysis was that a model having two components could explain 90 % of the total variance within the data set. The major component, t_1, described 85 % and was mainly accounting for the grouping of samples from the initial cooking and the final cooking period. As could be seen from Fig.3, there are few pulp samples in between these two groups, all of them representing intermediate cooking times. As earlier mentioned the relevance of each variable in the model can be

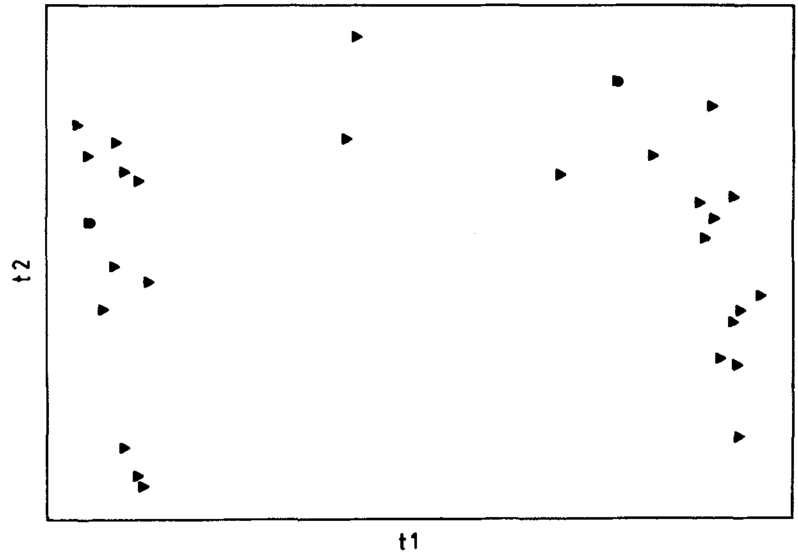


FIGURE 3 Plot of the first two principal components (score t_1 vs score t_2) from the PC analysis. Samples from initial cooking period to the left, and from the final cooking period to the right.

illustrated by a loading (b_{ia}) plot. Since the t_1 component approximately parallels the delignification process, one would expect a plot of the loading values for the first component vs. shift to show strong similarities to a ^{13}C NMR spectrum of lignin. As seen in Fig.4 it is possible to assign the main peaks in the "subspectrum" to lignin resonances[16].

In this way it would be possible to construct subspectra corresponding to each relevant PC component. The whole class analysis could be subdivided into, in this case, three separate classes based on the grouping pattern. New b_{ia}-plots will now approximate composition/ morphology changes within each

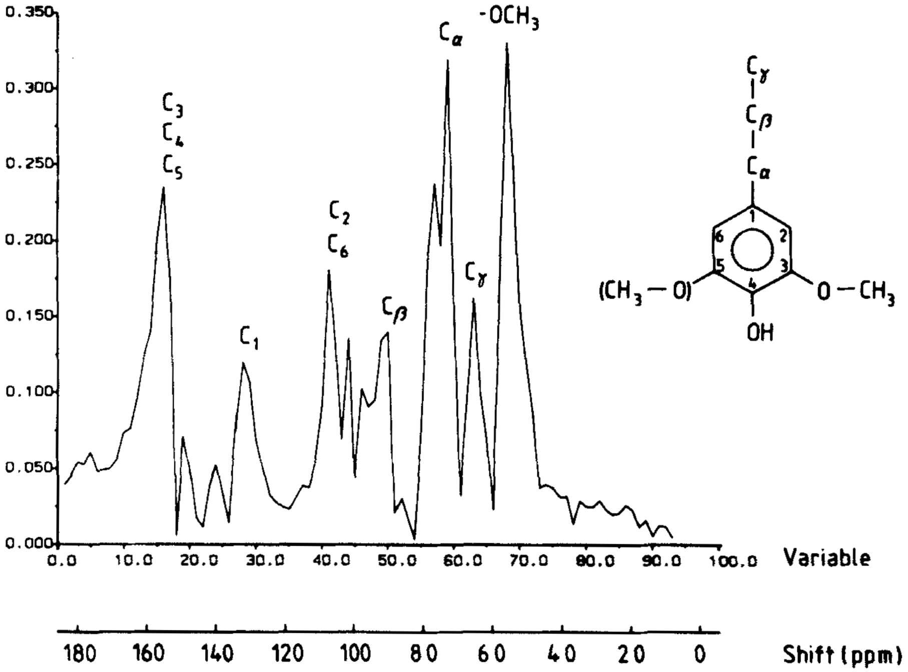


FIGURE 4 Plot of chemical shift values vs loadings (b_{ia}) for the first component in the PC analysis.

group. Thus by examining the loadings that result from a subclass analysis, it will be possible to sort out the composition changes that occur during a specific pulping period. This approach will be especially useful in the two-block analysis step, when we have access to other pulp descriptors than the lignin content. This approach will be considered in detail in a future paper.

Partial Least Squares Data Analysis (PLS).

Since the PC analysis showed the expected grouping due to the delignification process, it could be worthwhile to treat the

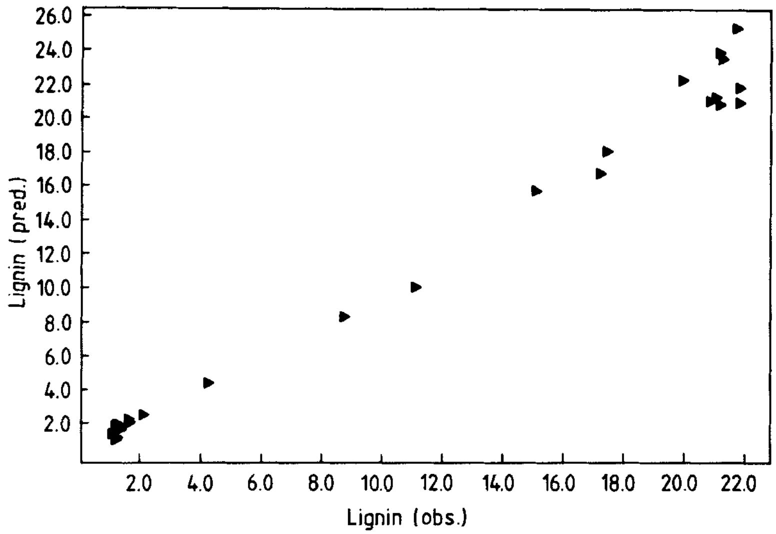


FIGURE 5 Plot of measured lignin content (obs) vs PLS predicted lignin content (pred) for "unknown" samples (not included to calculate the model).

problem as a two-block approach, where the intention is to approximate the information content in the NMR matrix and the calibration set (lignin content) simultaneously. The result of such an analysis showed that 99.7 % of the variance in the lignin content can be described by a three-component PLS model (26 of the 51 objects where used to calculate the model and the remaining samples were predicted). The first of these three components explained as much as 98 % of the variance in lignin data. Based on this learning set, we have a most efficient model to predict the lignin content of new samples, as can be seen in Fig.5 .

Hence, based on the three-component PLS model, a CP/MAS ^{13}C NMR spectrum is completely sufficient to predict the lignin content of an "unknown" pulp sample (which has not been included

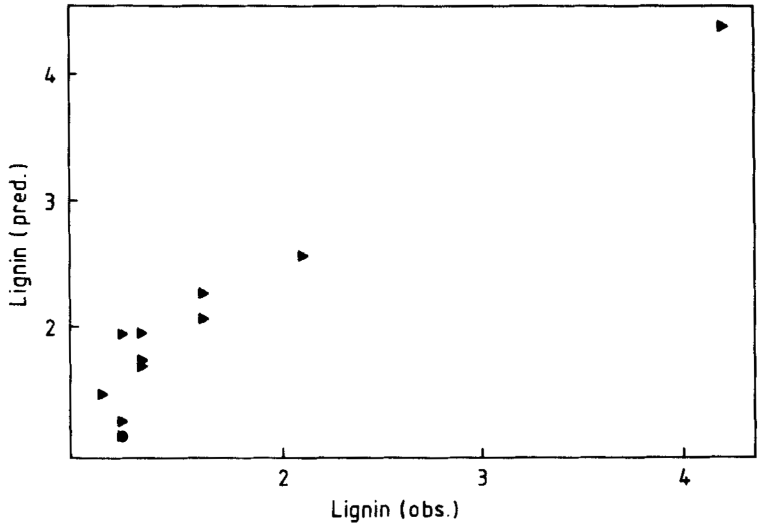


FIGURE 6 Expansion of figure 5, showing samples with low lignin content.

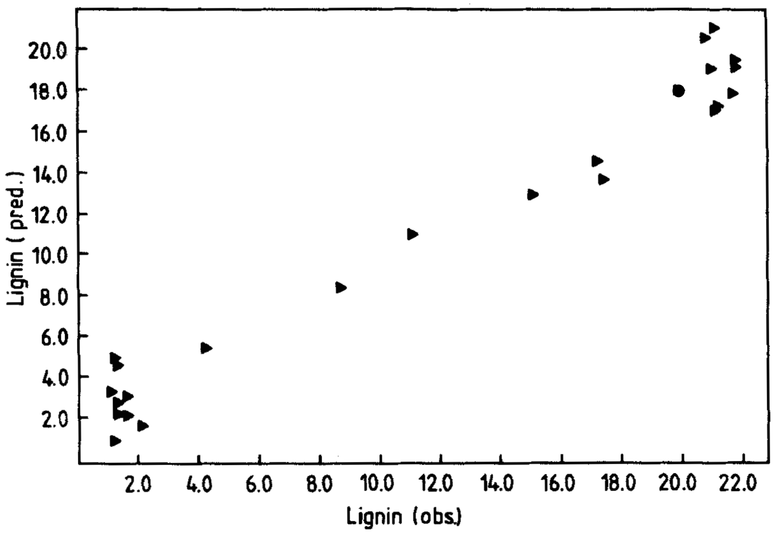


FIGURE 7 Plot of measured lignin content (obs) vs PLS predicted lignin content (pred) for "unknown" samples. The model is calculated based on resonances from the carbohydrate spectral region (60-115 ppm).

to calculate the model). It is especially interesting to consider the predictive ability using samples of low lignin content, where traditional data analytic methods based on the intensity of resolved resonances have failed. As seen in Fig. 6, the stability of the model is quite impressive i.e. samples having lignin concentration less than 5 % are all well predicted. As expected the b_{ia}-plot for the first PLS component is identical with the b_{ia}-plot from the PC analysis presented in Fig.4.

To further demonstrate the power of this technique, the analysis was repeated using only variables in the overlapped carbohydrate spectral region (60-115 ppm). As can be seen from Fig.7, the prediction of the lignin content is still acceptable. It is obvious that even when the resolved peaks for lignin (methoxy and aromatic) are excluded, there is still enough information in the nonresolved region which could be used to predict lignin content using the PLS method.

CONCLUSIONS

In situ pulp kinetics can be probed by ^{13}C CP/MAS NMR and the information content can easily be extracted using multivariate data analysis methods. The chemical interpretation is substantially facilitated if NMR data are coupled to other data representing chemical/physical descriptors of the pulp, as exemplified here with lignin content. The interrelations between such data, using partial least squares data analysis, open new probes to monitor changes in composition and morphology during pulping and to compare relative reaction rates in localized areas of wood polymers. An extension of this methodology is to add data from other spectroscopic techniques, such as IR and NIR. This will be considered in detail in a future paper.

ACKNOWLEDGEMENTS

We are grateful to MoDo AB for providing pulp samples and to Nils and Dorthii Troedsson Foundation for financial support.

REFERENCES

1. Jacqueline A. Hemmingson and Roger H. Newman, J. Wood Chem. Tech. 5 (2), 159-188 (1985).

2. David L. VanderHart and Rajai H. Atalla, Macromolecules, 17, 1465-1472 (1984).

3. F. Horii, A. Hirai, and R. Kitamaru, J. Carbohydrate Chem.,3 (4), 641-662 (1984).

4. Roger H. Newman, Fourth International Symposium on Wood and Pulping Chemistry, vol 1., pp. 195-199, Paris, 1987.

5. James F. Haw, Gary E. Maciel and Herbert A. Schroeder, Anal. Chem.,56, 1323 (1984).

6. C. Albano, G. Blomqvist, W.J. Dunn III, U. Edlund, B. Eliasson, E. Johansson, B. Nordén, M. Sjöström, B. Söderström and S. Wold, Proc. of 27th IUPAC Congr. of Pure and Appl. Chem., p 475, Ed. A. Varmavouri, Pergamon Press, Oxford 1980.

7. B. Nordén and C. Albano, "Principal Components Analysis of complex NMR spectra from heterogeneous material", Fuel, 1989, In press.

8. M. Sjöström, S. Wold, W. Lindberg, J. Persson and H. Martens, Analytica Chimica Acta, 150, 61-70 (1983).

9. A. Holmgren and B. Nordén, Applied Spectroscopy, 42, 255-262 (1988).

10. Bruker DISMSL Software, Magnitude calculation ($R_{new} = (R^2 + I^2)^{1/2}$).

11. The SIMCA 3B BASIC program package is available from Sepanova AB, Östrandsv. 14, S-122 45 Enskede, Sweden or from Principal Data Components, Columbia, Missouri, USA.

12. I.T. Jolliffe, Principal Component Analysis., Springer Verlag, New York, 1986.

13. S. Wold, C. Albano, W.J. Dunn III, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg and M. Sjöström, Multivariate data analysis in chemistry. In Chemometrics Mathematics and statistics in chemistry, B.R. Kowalski, Ed. D. Reidel Publ. Co., Dordrecht, Holland, 1984, p. 17.

14. M. A. Sharaf, D. L. Illman and B. R. Kowalski, Chemometrics, Wiley-Interscience, New York, 1986.

15. S. Wold, Technometrics 20, 397 (1978).

16. D. Kwoh, S. Bhattacharjee, J. Cael and S. Patt, Tappi Proceedings, Research and Development Division Conference, 1982, p. 113.