

**PRINCIPAL COMPONENT ANALYSIS OF THE  
HYDROXYMETHYLATION OF SUGARCANE LIGNIN:  
A TIME-DEPENDING STUDY BY FTIR**

Priscila Benar and Adilson R. Gonçalves  
Depto. Biotecnologia - Faculdade de Engenharia Química de Lorena  
Cx. Postal 116, 12.600-000 Lorena/SP/Brazil

Dalmo Mandelli, Márcia M. C. Ferreira, and Ulf Schuchardt  
Instituto de Química - Universidade Estadual de Campinas  
Cx. Postal 6154, 13.083-970 Campinas/SP/Brazil

**ABSTRACT**

The hydroxymethylation of Acetosolv lignin from sugarcane bagasse was studied at 40°C between 0.25 and 8 h by Fourier-transformed infrared spectroscopy (FTIR). Principal Component Analysis (PCA) of the spectra has determined that addition of formaldehyde to the lignin fragments was completed within 4 h. After that time, only condensation of the fragments occurs, leading to a resol-type crosslinked resin. PCA performed in selected regions of the FTIR spectra showed that the same results are also obtained for the stretching region of O-H and C-H within the range from 2800 to 3800 cm<sup>-1</sup>, which is not usually considered in the spectroscopic evaluations of modified lignins.

**INTRODUCTION**

Lignins have been used as a component for phenol-formaldehyde resins (resols).<sup>1,2</sup> The partial substitution of phenol by lignin gives rise to a

copolymer where lignin fragments are linked by methylene bonds (figure 1). The first step of the process consists in the addition of formaldehyde to the lignin fragment at free *ortho* positions of the aromatic ring (with respect to OH group) to obtain a 3- or 5-hydroxymethylphenol derivative (figure 1, reaction A). This step is followed by a condensation reaction between the OH groups of the hydroxymethylated lignin and another lignin fragment (figure 1, reaction B).<sup>3,4</sup>

We studied the hydroxymethylation of sugarcane bagasse (*Saccharum officinarum*) lignin using Fourier-transformed infrared spectroscopy (FTIR) followed by Principal Component Analysis (PCA). PCA is a mathematical procedure for data treatment.<sup>5</sup> The use of such chemometric methods is becoming more frequent in lignocellulosic chemistry, as can be seen by the focus of the recent ACS congress<sup>6</sup> and examples from the literature.<sup>7-9</sup>

## EXPERIMENTAL

### Hydroxymethylation and FTIR Spectroscopy

Acetosolv pulping conditions for the obtention of sugarcane bagasse lignin were described elsewhere.<sup>10</sup> Hydroxymethylation was performed with 10 g of lignin in 90 g of water by addition of 2.2 g of NaOH and 3.9 g of formaldehyde (10 mL of a 37%-aqueous solution) in a round bottom flask, which was kept at 40±2°C. Samples of the reaction mixture were taken at hydroxymethylation times between 0.25 and 8 h. The amount of formaldehyde in solution was determined by the reaction with Na<sub>2</sub>SO<sub>3</sub> and HCl, and back-titration with NaOH.<sup>11</sup> At each hydroxymethylation time, a sample of the mixture was also added dropwise to water. The solid obtained (hydroxymethylated lignin) was

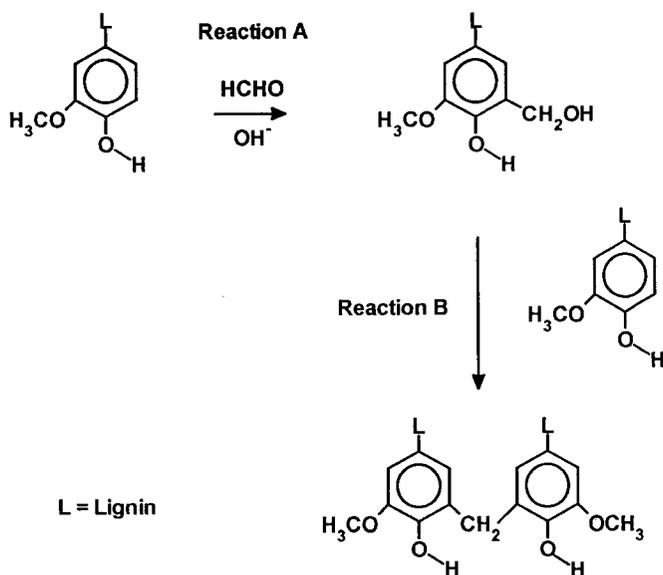


FIGURE 1: Reaction A) hydroxymethylation of lignin; Reaction B) condensation between the OH group of the hydroxymethylated lignin and another lignin fragment<sup>3,4</sup>

exhaustively washed with water, filtered, and dried over  $P_2O_5$ /vacuum. FTIR spectra were recorded in a BIO-RAD FTS 40 spectrometer using pellets containing 1.5 mg of lignin and 350 mg of KBr. All spectra were recorded from  $397\text{ cm}^{-1}$  to  $4002\text{ cm}^{-1}$  with a resolution of  $2\text{ cm}^{-1}$  (1870 points for each spectrum) and normalized for the absorption of the aromatic ring vibrations at  $1505$  to  $1515\text{ cm}^{-1}$ .

### Data Treatment

The MATLAB™ program<sup>12,13</sup> was used for PCA calculations. The PCA's goal is to transform the data into a new form, where the most relevant informations become more explicit. Each row in the matrix  $X$

corresponds to an FTIR spectrum of the hydroxymethylated lignin at a particular reaction time, while the columns consists of the relative absorbances at the respective wavenumbers. In the first place, all spectra were mean-centered, *i.e.* each row of each spectrum column was subtracted by the average value of that column. The 7 samples (0.25, 0.5, 1, 2, 4, 6, and 8 h) were distributed in a system with 1870 axes, corresponding to 1870 wavenumbers. This data matrix is decomposed in two matrices; the Scores  $\mathbf{T}$  and the Loadings  $\mathbf{P}$ , such that their product  $\mathbf{T}\mathbf{P}^T$  reproduces the original data. It is important to notice that this transformation does not change the relationship between the samples. In vector notation, the data matrix  $\mathbf{X}$  is represented by the sum of the vector-products  $\mathbf{t}_n$  and  $\mathbf{p}_n^T$ , as shown in figure 2, where  $\mathbf{p}_n$  are the eigenvectors of  $\mathbf{X}^T\mathbf{X}$ .

The first principal component ( $\mathbf{t}_1, \mathbf{p}_1^T$ ) describes the maximum of variation or spread in the samples.<sup>5</sup> A second principal component, orthogonal to the first, *i. e.* completely uncorrelated to it, describes the maximum of the variation not described by the first eigenvector, and so on. With this procedure, the most important features of the data set can be seen in a low dimensionality plot (hopefully 2D and 3D). The loadings  $\mathbf{p}$  are the weight of each original variable (wavenumbers) in the respective eigenvector, and correspond to the cosine of the angle between the eigenvector and the variable axes. The scores  $\mathbf{t}$  are the projection of each sample on the principal component axes, *i. e.* their coordinates in the new space.

After the PCA of the spectra, the regions where the modeling power<sup>14</sup> and the loadings had no significance, were excluded: 397-703  $\text{cm}^{-1}$  *i.e.*, in the beginning of the spectra, where there is no absorption of organic compounds, 1914-2392  $\text{cm}^{-1}$  where only the  $\text{CO}_2$  absorption is observed, and 3720-4002  $\text{cm}^{-1}$  a region without bands. This

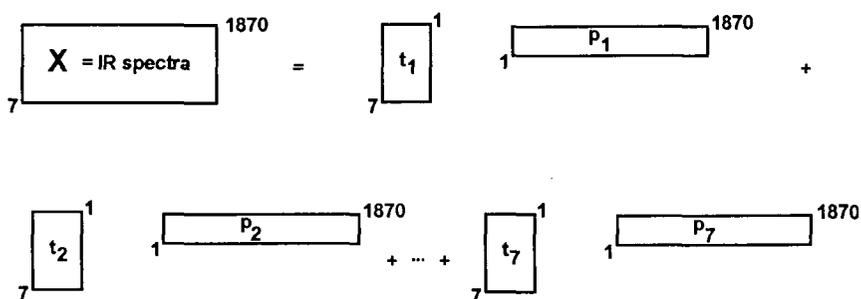


FIGURE 2: Schematic representation of the matrix X (FTIR data) as a vector product

whole procedure can be summarized in the flow chart shown in figure 3, where the indices indicate the matrix size. The PCA results were obtained from this pre-treated matrix 1310x7.

## RESULTS AND DISCUSSION

The FTIR spectra of the original and some hydroxymethylated lignins are shown in figure 4. The spectra were normalized for the  $1510 \text{ cm}^{-1}$  absorption<sup>15</sup> to avoid problems with different amounts of lignins in different samples.

A study of the reaction using these spectra needs a closer analysis due to the difficulty in choosing the bands which are significant for the process. For a univariate approximation we selected some wavenumbers, corresponding to maximum absorbances, to follow the reaction. The results obtained are shown in figure 5.

The determination of the end of the reaction is very difficult, because the absorbances of some bands stop increasing after 2 h, and others after 4 h. There is a random variation in the absorbances after this time.

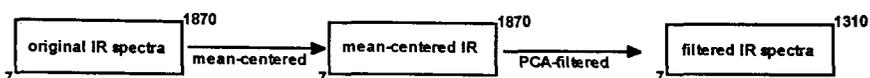


FIGURE 3: Schematic demonstration of the data pre-treatment

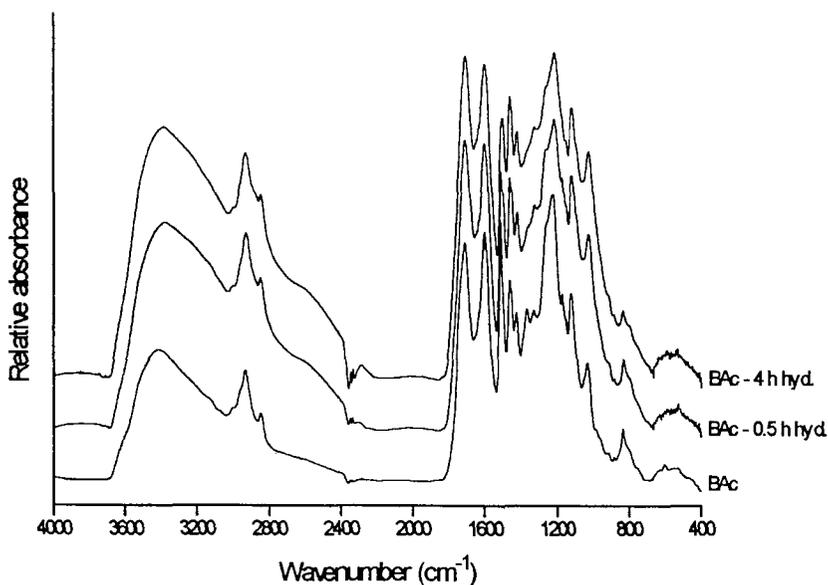


FIGURE 4: FTIR spectra of original and hydroxymethylated Acetosolv lignins from sugarcane bagasse (0.5 h and 4 h)

In such cases, due to the possibility of a higher signal-noise ratio, the use of a principal component analysis is very appropriate, allowing more precise conclusions.<sup>16</sup> Table 1 shows the results of the PCA of the pre-treated matrix of FTIR spectra.

The first principal component (PC 1) is related to the total reaction and describes almost 88% of the observed variance in the system. All bands have significant loadings values (figure 6a), more so at bands

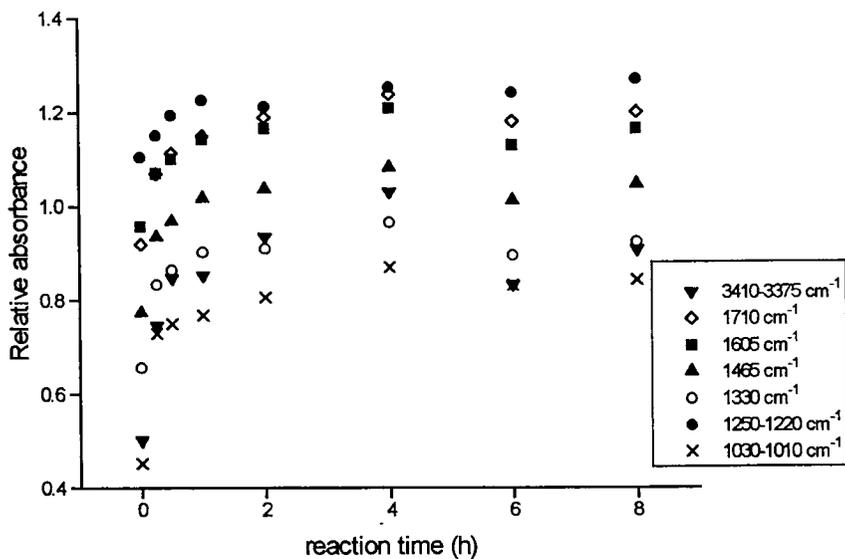


FIGURE 5: Univariate study of the hydroxymethylation of Acetosolv lignin from sugarcane bagasse by FTIR

TABLE 1

Percent Variance Captured by PCA Model - Total Spectra

Principal Component	Eigenvalue of Cov(X)	% Variance	Accumulated % Variance
1	2.44	87.7	87.7
2	0.24	8.7	96.4
3	0.07	2.6	99.0
4	0.01	0.5	99.5
5	0.01	0.3	99.8
6	0.00	0.2	100.0
7	0.00	0.0	100.0

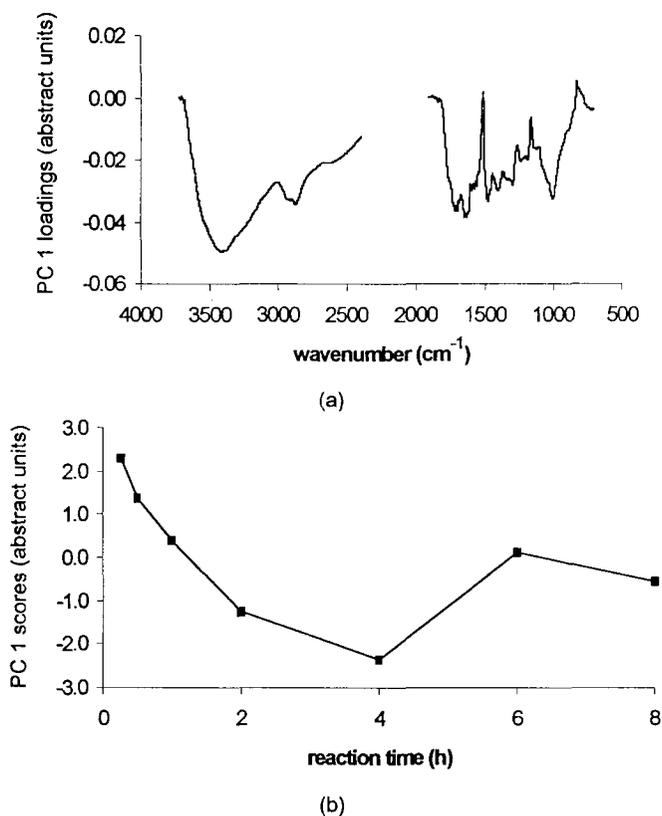


FIGURE 6: For whole spectrum: (a) PC 1 loadings vs wavenumber and (b) PC 1 scores vs reaction time

1010  $\text{cm}^{-1}$  (C-O stretching in primary alcohols), 1680  $\text{cm}^{-1}$  (C=O), and 3400  $\text{cm}^{-1}$  (O-H stretching). The hydroxymethylation (figure 1, reaction A) is completed in 4 h, as can be observed by the inflexion of the plot of scores PC 1 vs reaction time (figure 6b). After 4 h, a degradation or modification may occur in the lignin, probably due to the condensation between units (figure 1, reaction B), forming methylene bonds (crosslinking).

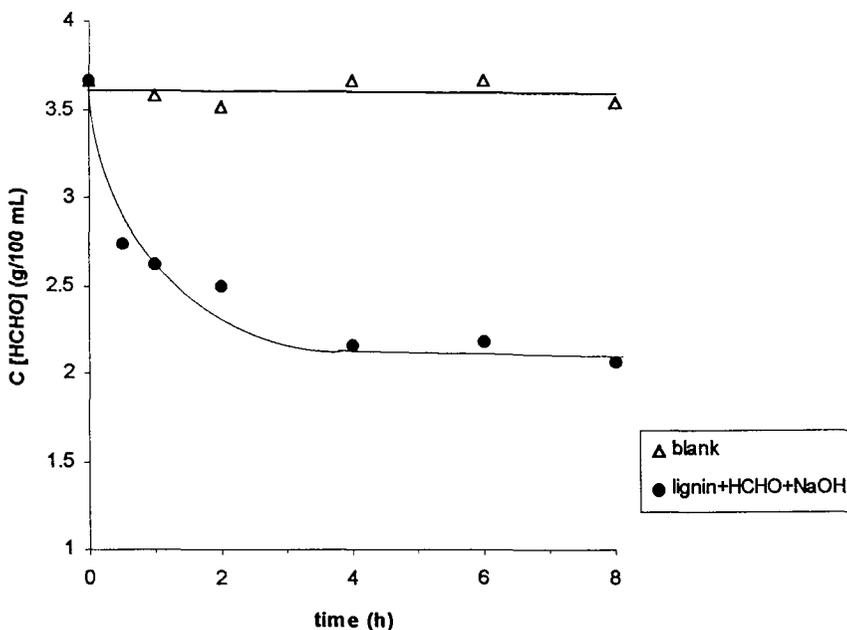


FIGURE 7: Monitoring of consumption of formaldehyde in solution during the hydroxymethylation of Acetosolv lignin from sugarcane bagasse

The end of the reaction was confirmed monitoring the consumption of formaldehyde in solution (figure 7), showing that the quantity of HCHO decreases until 4 h of hydroxymethylation at 40°C and then remains constant.

Only few wavenumbers show significant contribution to the loadings of the second and third PCs. These components are related to some specific chemical transformation that may occur on the lignin during the reaction. The second component describes almost 9% of the observed variance (figure 8a).

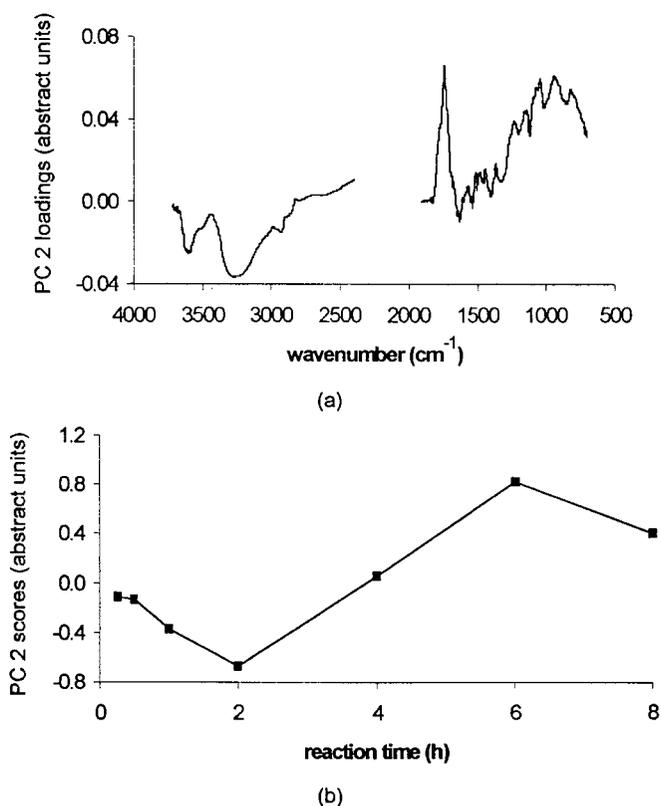


FIGURE 8: For whole spectrum: (a) PC 2 loadings vs wavenumber and (b) PC 2 scores vs reaction time

Score values with a given signal (+ or -) are mainly influenced by the loading values with the same signal. Comparing the scores (figure 8b) and loadings of PC 2, a correlation can be observed between loading values at 800-1100  $\text{cm}^{-1}$  and 1700  $\text{cm}^{-1}$  and scores of the samples with higher reaction time (mainly 6 h). Similarly, the region around 3250  $\text{cm}^{-1}$  corresponds to the scores of the samples with a smaller reaction time (mainly 2 h). The beginning of the reaction is dominated by the hydroxymethylation which gives rise to the formation of hydrogen bonds

between the OH of phenolic and the OH of alcoholic groups and decreases the energy of the OH stretching<sup>17</sup> ( $3400\text{ cm}^{-1}$  in the PC 1 to  $3250\text{ cm}^{-1}$  in the PC 2). Increasing the reaction time (above 4 h), condensation reactions take place, decreasing the amount of OH groups and, as a consequence, the weight of this variable. During the reaction, the aromatic ring substitutions also increase. This is shown by higher loadings at the region between  $800\text{-}1100\text{ cm}^{-1}$  related to the scores of the samples at later time. The high loading of the carbonyl band ( $1740\text{ cm}^{-1}$ ), mainly found for samples at the end of the reaction (4 to 8 h), has not a good explanation. This band is not related to C=O of residual formaldehyde, since the band of C-H stretching of aldehyde groups at  $2700\text{ cm}^{-1}$  was not observed.

The third PC describes almost 3% of the observed variance and completes the analysis of the reaction. Two main loading regions ( $1250$  and  $3400\text{ cm}^{-1}$ ) are related to the scores of the samples after 1 h and 8 h of reaction time (figures 9a and 9b). Probably, the absorption at  $3400\text{ cm}^{-1}$  influences the scores for 1 h, due to incorporation of OH groups. On the other hand, the absorption at  $1250\text{ cm}^{-1}$  influences both the scores for 1 h and 8 h, since carbon-carbon bonds are formed during the hydroxymethylation and the condensation.

The OH and CH stretching region ( $2394\text{-}3718\text{ cm}^{-1}$ ) may incorporate some error due to water contamination. Therefore, this region is normally not used for the analysis of modification of lignins. The PCA was performed using only the  $800\text{-}1950\text{ cm}^{-1}$  region and the results were very similar to that obtained using the whole spectrum.

Also, the PCA using only the OH and CH stretching region was performed. Table 2 shows the results of the PCA model and figure 10 shows the loadings and scores of the first PC in this case, which describes almost 96% of the variance in the system. As can be observed,

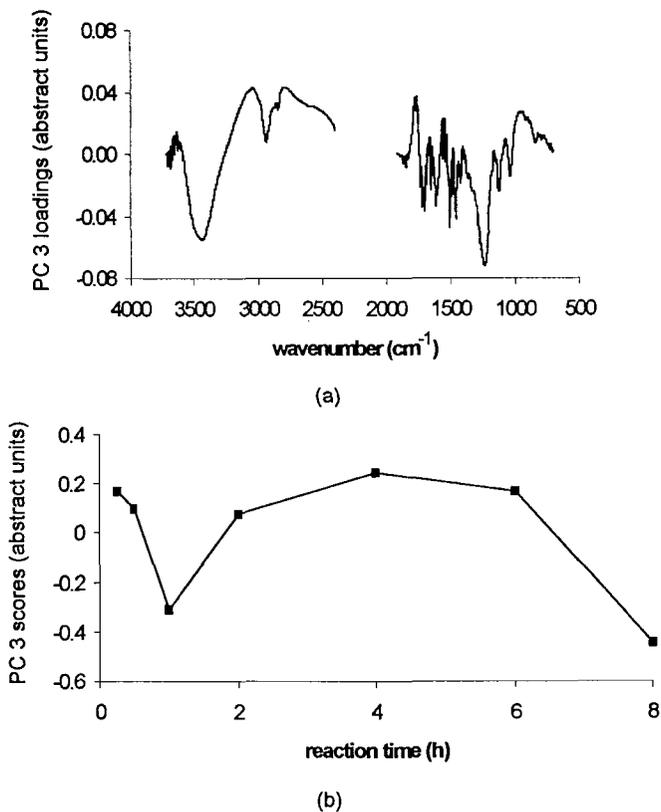


FIGURE 9: For whole spectrum: (a) PC 3 loadings vs wavenumber and (b) PC 3 scores vs reaction time

TABLE 2

Percent Variance Captured by PCA Model -OH and CH Stretching Region

Principal Component	Eigenvalue of Cov(X)	% Variance	Accumulated % Variance
1	1.74	95.8	95.8
2	0.05	2.8	98.6
3	0.02	1.2	99.8
4	0.00	0.2	100.0
5	0.00	0.0	100.0

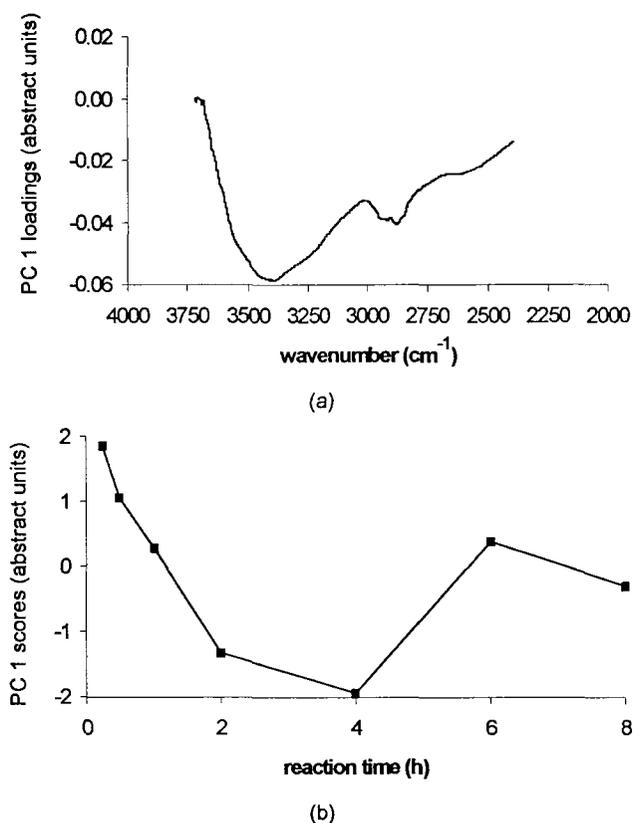


FIGURE 10: For O-H and C-H stretching region ( $2394\text{-}3718\text{ cm}^{-1}$ ): (a) PC 1 loadings vs wavenumber and (b) PC 1 scores vs reaction time

this region, which is not usually considered reliable, leads to the same conclusions obtained before.

### CONCLUSIONS

As evidenced by PCA of the FTIR spectra, the reaction time required for the hydroxymethylation of sugarcane bagasse lignin at  $40^{\circ}\text{C}$

is 4 h. This result is obtained from the scores of the first principal component, which explains 88% of the system, and is confirmed by the formaldehyde consumption during the reaction. After this time, condensation reactions occur. PCA of FTIR data allows the monitoring of the hydroxymethylation besides exhibiting the most significant bands for the process. Finally, a novel interpretation of these FTIR spectra can be given: the region comprising the OH and CH stretching, usually not considered in infrared investigations of lignins, can be also used for this analysis, giving very similar results.

### ACKNOWLEDGEMENTS

The authors acknowledge CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil) and FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo - Brazil) for fellowships.

### REFERENCES

1. D.J. Gardner and G.D. McGinnis, *J. Wood Chem. Technol.*, **8**, 261 (1988).
2. A. Pizzi, F-A. Cameron, and G.H. van der Klarshorst, In Adhesives from Renewable Resources, Chap. 7, R.W. Hemingway and A.H. Conner (ed.), ACS Symp. Series 385, Washington, 1989.
3. G. H. van der Klashorst, *J. Wood. Chem. Technol.*, **8**, 209 (1988).
4. P. Truter, A. Pizzi, and H. Vermaas, *J. Appl. Polym. Sci.*, **51**, 1319 (1994).
5. K.R. Beebe and B.R. Kowalski, *Anal.Chem.*, **59**, 1007 (1987).
6. ACS Symposium '96; Cellulose, Paper, and Textile Division; Chemometrics in Biomass Characterization and Biomass Conversion; Orlando/Florida, Aug. 28-29/1996.

7. L. Wallbäcks, U. Edlund, and T. Lindgren, *Nord. Pulp. Pap. Res. J.*, 10(2), 88 (1995).
8. L. Wallbäcks, U. Edlund, and B. Nordén, *Nord. Pulp. Pap. Res. J.*, 6(3), 104 (1991).
9. P. Benar, A.R. Gonçalves, D. Mandelli, and U. Schuchardt, *Biores. Technol.*, accepted for publication (1998).
10. P. Benar, U. Schuchardt, and H. Nimz, "Characterization of Acetosolv and Formacell Lignins from Sugarcane Bagasse", *Proc. 4<sup>th</sup> Brazilian Symp. Chem. Lignin and Wood Components*, p. 29-32, Recife/Brazil (1995).
11. W. Peng, B. Riedl, and A.O. Barry, *J. Appl. Polym. Sci.* 48, 1757 (1993).
12. B. M. Wise and N. B. Gallagher, *PLS\_Toolbox version 1.5 for use with MATLAB™* (1996).
13. *MATLAB User's Guide; The Math Works: South Natick, MA* (1994).
14. Modeling power, MP, is very useful for variable selection (points out the variables which can be excluded from the data set). It is defined as:  $MP=1-s_j/s_{0j}$  where  $s_j^2$  is the variable residual variance and  $s_{0j}^2$  is the total variance of that variable.
15. O. Faix, In Methods in Lignin Chemistry, Chap. 4, S.Y. Lin and C.W. Dence (ed.), Springer, Berlin, 1992.
16. H. Martens and T. Naes, Multivariate Calibration, John Wiley & Sons, New York, 1989.
17. R.M. Silverstein, G.C. Bassler, and T.S. Morrill, Spectrometric Identification of Organic Compounds, 5<sup>th</sup> edition, p. 158, John Wiley & Sons, New York, 1991.