Chemical Compositions of Hardwood and Softwood Pulps Employing Photoacoustic Fourier Transform Infrared Spectroscopy in Combination with Partial Least-Squares Analysis

Sofia Bjarnestad and Olof Dahlman*

Swedish Pulp and Paper Research Institute, P.O. Box 5604, SE,114 86 Stockholm, Sweden

In the present study, hardwood and softwood pulps were characterized by employing Fourier transform infrared photoacoustic spectroscopy (FT-IR-PAS). The pulp samples examined originated from Swedish sulfite and kraft pulp mills, which utilize different cooking processes and modern bleaching technologies. Partial least-squares (PLS) analysis was used to correlate the spectral data obtained with the kappa (K) numbers and carbohydrate compositions of the pulp samples determined by enzymatic hydrolysis and subsequent capillary zone electrophoresis. Using four principal components, the present PLS model based on photoacoustic FT-IR spectra could explain 85% of the variance in the X matrix and 81% of the variance in the Y matrix. The FT-IR-PAS technique in combination with PLS was found to accurately predict the contents of carbohydrates, i.e., xylose, glucose, mannose, arabinose, galactose, and hexenuronic acid residues, as well as the content of lignin measured in terms of K numbers and corrected K numbers of the pulps. From these predictions, the contents of xylan, glucomannan, and cellulose can also be predicted. The content of 4-*O*-methylglucuronic acid residues is, however, more difficult to predict accurately, using this approach.

The use of different types of wood and pulping procedures (mechanically or chemically) results in fibers with different properties, which will in turn influence the final properties of the paper products. Multivariate calibration models in combination with Fourier transform infrared spectroscopy (FT-IR) are ideally suited for analysis of the chemical composition of complex materials such as wood fibers. In a number of investigations partial least-squares (PLS) analysis has been used together with FT-IR for rapid characterization of lignin and carbohydrates in wood and pulps. Some of these studies are summarized in Table 1.

In a recent study by Jones et al.,[1] the lignin, hemicellulose, glucan, and extractive contents of a moving stream of wood chips were determined by employing thermal emission mid-infrared spectroscopy in combination with PLS regression analysis. This experimental setup was designed to mimic the movement of wood chips along a conveyor belt, and the PLS analysis was performed off-line as the spectra were gathered. On the basis of this study, these investigators suggested that PLS analysis can be performed on-line in real time once a regression model has been established.

Rosencwaig and Gersho[9] first described the photoacoustic effect in combination with spectroscopy in 1976. Since then, this approach has been developed considerably and several different procedures for processing data are now available. In the present study, digital signal processing (DSP) of step-scan FT-IR-photoacoustic spectroscopy (FT-IR-PAS) was employed. This software-based method has been described thoroughly by Manning and Griffiths[10] and by Drapcho et al.[11]

Even though the FT-IR-PAS technique is capable of providing spectra of good quality from minute quantities of fibers and, in addition, allows near-surface analysis, only a few applications of this technique to wood fibers and kraft pulps have been reported to date. Some papers concerning this photoacoustic approach are included in Table 1. Kuo and co-workers[14] demonstrated that FT-IR-PAS can be utilized to obtain high-quality spectra of different forms of wood samples. These investigators also concluded that, in many cases, the photoacoustic approach overcomes difficulties encountered with the KBr pellet, diffuse reflectance, and attenuated total reflectance methods.

* Corresponding author: (e-mail) olof.dahlman@stfi.se; (fax) +46 8 10 83 40.

(1) Jones, R. W.; Meglen, R. R.; Hames, B. R.; McClelland, J. F. *Anal. Chem.* **2002,** *74,* 453−457.

(2) Schultz, T. P.; Burns, D. A. *Tappi J.* **1990,** *73,* 209−212.

(3) Meder, R.; Gallagher, S.; Mackie, K. L.; Böhler, H.; Meglen, R. R. *Holzforschung* **1999,** *53,* 261−266.

(4) Rodrigues, J.; Puls, J.; Faix, O.; Pereira, H. *Holzforschung* **2001,** *55,* 265−269.

(5) Backa, S.; Brolin, A. *Tappi J.* **1991,** *74,* 218−226.

(6) Gellerstedt, G.; Josefsson, T., *Third European Workshop on Lignucellulosics and Pulp,* Stockholm, Sweden, August 28−31, 1994; pp 179−182.

(7) Coimbra, M. A.; Barros, A.; Barros, M.; Rutledge, D. N.; Delgadillo, I. *Carbohydr. Polym.* **1998,** *37,* 241−248.

(8) Coimbra, M. A.; Barros, A.; Rutledge, D. N.; Delgadillo, I. *Carbohydr. Res.* **1999,** *317,* 145−154.

(9) Rosencwaig, A.; Gersho, A. *J. Appl. Phys.* **1976,** *47,* 64−69.

(10) Manning, C. J.; Griffiths, P. R. *Appl. Spectrosc.* **1993,** *47,* 1345−1349.

(11) Drapcho, D. L.; Curbelo, R.; Jiang, E. Y.; Crocombe, R. A.; McCarthy, W. J. *Appl. Spectrosc.* **1997,** *51,* 453−460.

(12) Balderas-López, J. A.; Tomás, S. A.; Vargas, H.; Olalde-Portugal, V.; Baquero, R.; Delgadillo, I.; Yánez-Limón, J. M.; Alvarado-Gil, J. J.; Frías-Hernández, J.; Scheinvar, L.; Falconi, C.; Silva, M. D.; Miranda, L. C. Mà, *Forest Prod. J.* **1996,** *46,* 84−89.

(13) Harbour, J. R.; Hopper, M. A.; Marchessault, R. H.; Dobbin, C. J.; Anczurowski, E. *J. Pulp Pap. Sci.* **1985,** *11,* J42-J47.

(14) Kuo, M.-l.; McClelland, J. F.; Luo, S.; Chien, P.-L.; Walker, R. D.; Hse, C.-Y. *Wood Fiber Sci.* **1988,** *20,* 132−145.

Table 1. Summary of Papers Concerning Chemical Analysis Employing the FT-IR Approach

analytical approach	sample material	comments	ref
NIR and FT-IR	wood	lignin, hemicellulose and cellulose contents	2
PLS modeling of transmission and diffuse reflectance FT-IR spectra	*Pinus radiata* wood	chemical composition and the density	3
FT-IR spectroscopy together with PLS	*Eucalyptus globulus* wood	monosaccharide content	4
diffuse reflectance FTIR in combination with PLS	pulp	lignin and carbohydrates	5
FT-IR-PAS and NIR in combination with PLS	Kraft pulps	carbohydrate composition, methoxyl and phenolic groups	6
FT-IR spectroscopy in combination with PLS	fractions of cell wall material from olive and orange pulps	uronic acid and neutral sugars	7
several multivariate procedures, including PLS, were applied to FT-IR spectra	olive pulp cell wall	cell wall polysaccharides and quantitation of xylose	8
open PAS cell	wood	thermophysical properties	12
FT-IR-PAS	cellulose-containing materials	characterization of different materials	13
FT-IR-PAS	starch-based coated papers	depth profiling	16
FT-IR-PAS	keratin fibers	depth profiling	17

After analyzing samples of chemically bleached wood pulps produced in the laboratory, as well as commercially available papers employing photoacoustic FTIR, Lima and collaborators[15] concluded that this technique can be useful for comparing, controlling, and evaluating the effects of various processing parameters in an industrial plant on the properties of the resulting pulp and paper.

Gurnagul et al.[18] employed FT-IR-PAS to obtain spectra from papers prepared from a single batch of bleached kraft pulp using various degrees of beating and concluded that quantitative comparison of the spectra from different sheets requires consideration of sheet structure. Furthermore, in a comparison of different IR spectroscopic procedures for rapid analysis of wood constituents, Pandey and Theagarajan[19] concluded that the PAS spectra obtained from wood powder are similar to those obtained by transmission or diffuse reflectance spectroscopy (DRIFT) and are independent of particle size and the concentration of the wood powder.

We became interested in the use of FT-IR-PAS for analysis of wood samples for a number of reasons. In comparison to DRIFT, FT-IR-PAS is insensitive to the morphology of the sample; while compared to attenuated total reflectance spectroscopy (ATR), FT-IR-PAS provides more representative results since a larger sample is analyzed. Furthermore, with FT-IR-PAS, little or no preparation of the sample is required and depth-profiling analysis of pulp fibers is also possible.

Partial least-squares analysis can be used to correlate predictors, such as FT-IR data (X), to response parameters, i.e., data obtained by wet chemical analysis (Y), with the aim of being able to predict the latter from the former. To enhance the predictive power of multivariate calibration models, spectral data are often preprocessed prior to analysis, in attempt to minimize variation in these data that are unrelated to the response parameters of interest. In the present study, we have employed orthogonal signal correction[20] (OSC) for preprocessing of the spectral data.

OSC is a PLS-based approach that eliminates variation unrelated to the response parameters of interest, i.e., from the spectral data; what is eliminated from X should be mathematically independent of Y. OSC utilizes the response parameters to construct a filter for the spectral data and thereby create a new **X** matrix. The model employed here was constructed using the cross-validation approach.[20]

One important parameter for pulping and bleaching is the content of lignin in pulps. The lignin content can be estimated by determining the kappa (K) number, e.g., according to the procedure of the Scandinavian pulp, paper, and board Testing committee, SCAN-C 1:00.[39] In this test, the consumption of

(15) Lima, C. A. S.; Lima, M. B. S.; Miranda, L. C. M.; Baeza, J.; Freer, J.; Reyes, N.; Ruiz, J.; Silva, M. D. *Meas. Sci. Technol.* **2000**, *11*, 504−508.

(16) Halttunen, M.; Tenhunen, J.; Saarinen, T.; Stenius, P. *Vib. Spectrosc.* **1999**, *19*, 261−269.

(17) Jurdana, L. E.; Ghiggino, K. P.; Leaver, I. H.; Cole-Clarke, P. *Appl. Spectrosc.* **1995**, *49*, 361−366.

(18) Gurnagul, N.; St-Germain, F. G. T.; Gray, D. G. *J. Pulp Pap. Sci.* **1986**, *12*, J156-J159.

(19) Pandey, K. K.; Theagarajan, K. S. *Holz Roh-Werkstoff* **1997**, *55*, 383−390.

(20) Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. *Introduction to Multi- and Megavariate Data Analysis using Projection Methods (PCA & PLS)*; Umetrics: Umeå, Sweden, 1999.

(21) Dahlman, O.; Jacobs, A.; Liljenberg, A.; Olsson, A. *J. Chromatogr., A* **2000**, *891*, 157−174.

(22) Li, J.; Gellerstedt, G. *Carbohydr. Res.* **1997**, *302*, 213-218.

(23) Li, J.; Gellerstedt, G. *Nord. Pulp Pap. Res. J.* **1998**, *13*, 153−158.

(24) Bergnor-Gidnert, E.; Tomani, P. E.; Dahlman, O. *Nord. Pulp Pap. Res. J.* **1998**, *13*, 310−316.

(25) McClelland, J. F.; Jones, R. W.; Luo, S.; Seaverson, L. M. In *Practical Sampling Techniques for Infrared Analysis*; Coleman, P. B., Ed.; CRC Press: Boca Raton, FL, 1992.

(26) McClelland, J. F.; Jones, R. W.; Bajic, S. J. In *Handbook of Vibrational Spectroscopy*; Chalmers, J. M., Griffiths, P. R., Eds.; John Wiley: New York, 2002.

(27) McClelland, J. F.; Bajic, S. J.; Jones, R. W.; Seaverson, L. M. In *Modern Techniques of Applied Molecular Spectroscopy*; Mirabella, F. M., Jr., Ed.; John Wiley & Sons: New York, 1996.

(28) St-Germain, F. G. T.; Gray, D. G., *International Symposium Wood and Pulping Conference*, Vancouver, Canada, August 26−30, 1985; pp 261−262.

(29) Yang, C. Q.; Freeman, J. M. *Appl. Spectrosc.* **1991**, *45*, 1695−1698.

(30) Johansson, M. H.; Samuelsson, O. *Carbohydr. Res.* **1977**, *54*, 295−299.

(31) Kacurakova, M.; Capek, P.; Sasinkova, V.; Wellner, N.; Ebringerova, A. *Carbohydr. Polym.* **2000**, *43*, 195−203.

(32) Åkerholm, M. Dynamic FT-IR spectroscopy applied to studies on wood polymers. Licentiate Thesis, Royal Institute of Technology, Stockholm, Sweden, 2001.

potassiumpermanganate during treatment of the pulp under standardized conditions is measured, and the resulting value is used as an indirect measure of the lignin content.

An efficient procedure for determining the carbohydrate compositions of delignified wood and pulp samples free from extractives was developed previously in our laboratory.[21] This method involves enzymatic hydrolysis and subsequent capillary zone electrophoresis (CZE) and is somewhat time-consuming, since complete enzymatic hydrolysis of the wood fibers requires ~30 h of incubation. The time required for analysis would be shortened considerably if similar information could be obtained from an FT-IR spectrum of the pulp.

Therefore, in the present study we have employed multivariate data analysis by PLS to correlate results on pulps obtained by the CZE procedure with those obtained by FT-IR-PAS. As mentioned above, such an approach might be developed further to allow analysis of wood pulps on-line in a pulp mill,[1] which is, of course, also of great interest. Thus, we present here an analytical procedure based on FT-IR-PAS in combination with PLS for rapid determination of the chemical compositions of wood pulps.

EXPERIMENTAL SECTION

Materials. Hard- and softwood pulps originating from different manufacturing processes employing modern elementary chlorine-free (ECF) or totally chlorine-free (TCF) bleaching technologies were obtained from Swedish sulfite and kraft pulp mills.

Methods. *Analysis of Carbohydrate Composition.* The carbohydrate compositions of the pulps were determined by enzymatic hydrolysis of the polysaccharides present and subsequent analysis of the hydrolysates thus obtained by CZE, as described elsewhere.[21] The CZE system employed was a Beckman P/ACE MDQ capillary electrophoresis system (Beckman Coulter, Fullerton, CA) equipped with a diode-array UV detector. The sugars in the hydrolysates were derivatized with the UV chromophore 4-aminobenzoic acid ethyl ester (ABEE) and absorption at 306 nm then used for quantitation.

The contents of hemicellulose and cellulose polymers were calculated from the carbohydrate compositions of the pulps, by assuming that the glucose/mannose ratio was 1:4 for softwood glucomannan and 1:2 for hardwood glucomannan. Thus, the xylan, glucomannan, and cellulose contents of the pulps were calculated using the relationships depicted in Table 2.

Determination and Correction of the K *Number.* The K number of the pulps was determined by the standard of the Scandinavian pulp, paper, and board Testing committee (SCAN-C 1:00).[39] This standard procedure, traditionally utilized to estimate lignin content, is based on treatment of the pulp with potassium permanganate, which was originally assumed to result in oxidation of the double

(33) Dadashian, F.; Wilding, M. A. *J. Text. Inst.* **1999**, *90*, 275−287.
(34) Berben, S. A.; Rademacher, J. P.; Sell, L. O.; Easty, D. B. *Tappi J.* **1987**, 129−133.
(35) Derkachova, O. Y.; Sukhov, D. A.; Kazanskii, S. A.; Ryzov, A. I.; Osipova, N. N.; Osipov, B. P. *10th International Symposium on Wood and Pulping Chemistry,* Yokohama, Japan, 1999; pp 288−291.
(36) Wahls, M. W. C.; Kenttä, E.; Leyte, J. C. *Appl. Spectrosc.* **2000**, *54*, 214−220.
(37) Forsskåhl, I.; Kenttä, E.; Kyyrönen, P.; Sundström, O. *Appl. Spectrosc.* **1995**, *49*, 163−170.
(38) Kacurakova, M.; Wilson, R. H. *Carbohydr. Polym.* **2001**, *44*, 291−303.
(39) SCAN-C 1:00 Chemical pulp−Kappa number, Scandinavian pulp, paper, and board Testing committee, Stockholm, Sweden.

Table 2. Equations Employed To Calculate the Polysaccharide Composition of the Chemical Pulps

	calculation of content from the sugar composition
	Hardwood
glucuronoxylan	Xyl + 4-*O*-MeGlcA + HexA
glucomannan	Man + (1 mol of Glc/2 mol of Man)
cellulose	Glc − (1 mol of Glc/2 mol of Man)
	Softwood
arabinoglucuronoxylan	Xyl + Ara + 4-*O*-MeGlcA + HexA
galactoglucomannan	Gal + Man + (1 mol of Glc/4 mol of Man)
cellulose	Glc − (1 mol of Glc/4 mol of Man)

bonds and aromatic structures in the lignin. However, hexenuronic acid residues present in alkaline chemical pulps also contain double bonds which are oxidized by potassium permanganate and thereby contribute to the K number.[22,24] Accordingly, the K number was corrected to reflect the lignin content of the pulps more accurately, by subtracting the contribution by the hexenuronic acid residues present. This contribution was evaluated using the conversion factor suggested by Li and Gellerstedt,[22] i.e., 10 mmol of hexenuronic acid/kg of pulp corresponds to 0.85 K units.

FT-IR -PAS Analysis. The pulp samples were converted to their sodium form through the addition of 0.1 M aqueous sodium carbonate (pH of ~10), followed by washing with deionized water. Laboratory sheets (weighing ~20 g/m²) were subsequently prepared from the wet pulp samples, air-dried, and cut into pieces small enough to fit into the PAS sample holder. Spectral data in the range of 4000−400 cm⁻¹ were collected by employing a Bio-Rad FTS 6000 FT-IR spectrometer (Cambridge, MA) equipped with a MTEC 300 photoacoustic detector (Ames, IA). The detector was flushed with helium for ~5 min prior to the collection of data, to eliminate interference from water vapor. Win-IR Pro software was used to control the spectrometer, as well as to collect and process the data.

Measurements were performed using the following spectrometer settings: step rate, 1.0 Hz; phase modulation frequency, 100 Hz; phase modulation amplitude, $2\lambda_{\text{He−Ne}}$; spectral resolution, 8 cm⁻¹; A/D sampling rate, 4 kHz; and number of data samples per step, 1900. Triplicate measurements were carried out on each pulp sample. The digital signal processing utilized allows simultaneous demodulation of data collected at the fundamental of the phase modulation frequency and at its odd harmonics up to the ninth harmonic, thus probing five different depths within a single sample. However, since no investigation of the depth profile was carried out in the present study, only the data obtained with the fundamental of the phase modulation frequency (100 Hz) were evaluated.

Prior to performing multivariate analysis, the FT-IR-PAS spectra were normalized. Each spectral intensity, y_i, of a given pulp sample was divided by the square root of the sum of the squares of all intensities in the spectrum to give a new y_i' as follows:

$$y_i' = y_i / \sqrt{\sum_i y_i^2}$$

Multivariate Data Analysis. Multivariate data analysis employing PLS was performed with the SIMCA-P 9.0 software from
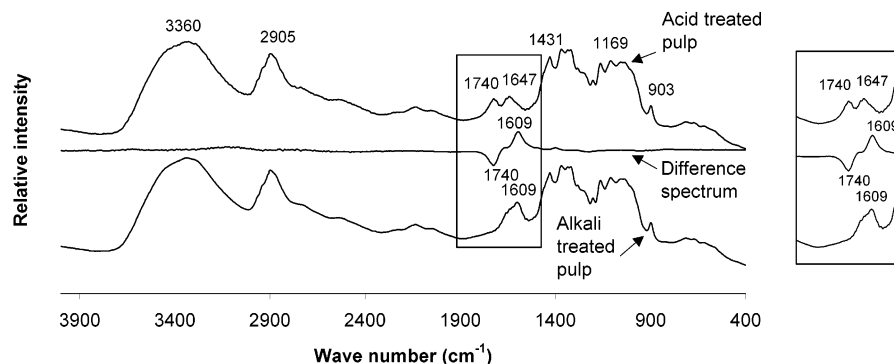
Figure 1. FT-IR-PAS spectra of a pulp sample treated with alkali (lower spectrum) or acid (upper spectrum) prior to analysis, together with the corresponding difference spectrum (in the middle). The peak present at ~1740 cm^{-1} after acidic treatment is eliminated by alkaline treatment, and the peak at ~1610 cm^{-1} is increased in intensity.

Umetrics (Umeà, Sweden). The PLS model was developed by utilizing the FT-IR-PAS spectra of the pulp samples (i.e., the spectral data from 4000 to 400 cm^{-1} divided into 935 variables) as the **X** matrix and the chemical determinations of the contents of various sugar residues (i.e., xylose, glucose, mannose, galactose, arabinose, hexenuronic acid, and 4-*O*-methylglucuronic acid), the K number, and the corrected K number as the **Y** matrix. The data of the **X** matrix were mean-centered, while the **Y** data were both mean-centered and variance-scaled prior to the performance of PLS calculations.

The spectral data were divided into a calibration set (consisting of two measurements on each sample), for construction of the PLS model and a test set (containing the third set of spectral data for each sample) for validation of this model. This validation involved predicting the carbohydrate composition of the pulp sample in the test set and calculating new values. These values were then inserted into the equations shown in Table 2 and the contents of xylan, glucomannan, and cellulose calculated. These predicted values were subsequently compared to the values obtained by CZE analysis.

RESULTS AND DISCUSSION

Sample Preparation for FT-IR-PAS Analysis. For comparison of the signal responses, the laboratory pulp sheets were prepared for analysis using two different procedures. In the first approach, a slurry of pulp in water was vacuum filtered through a nylon sieve to produce a sheet weighing ~20 g/m², which was then air-dried and later cut into pieces designed to fit the sample holder of the FT-IR-PAS instrument. The second preparation was performed by the standard procedure according to ISO 5269/1-1979. The pulp sheets thus obtained were subsequently analyzed without further treatment.

When these spectra were compared, several dissimilarities were detected, especially in the region of 900−1900 cm^{-1}, where the sheet prepared by the standard procedure seems to demonstrate somewhat higher signal intensity. As discussed above in the introduction, it was proposed earlier that the signal intensity could be influenced by the morphology of the sample;[18,19,25] but in contrast, a recent report reveals that photoacoustic spectroscopy is relatively insensitive to the condition of the surface.[26] Instead, the intensity differences observed might reflect differences in the gas volume in the sample cell, since the photoacoustic signal is inversely proportional to the gas volume in the sample compart-

ment.[25,27] Although, the difference between the two preparation procedures tested here was small, we decided that preparation of all the pulp sheets using the same preparation method (i.e., the slurry method) is more convenient and also minimizes any influence of different sheet structure.

Earlier investigations revealed the manner in which the spectra of different pulp samples may be changed by alkaline or acidic treatment.[28,29] Thus, the alkalinity or acidity of the sample might make it somewhat more difficult to evaluate its spectrum. The signal at ~1735 cm^{-2} is influenced by the carbonyl stretching in 4-*O*-methylglucuronic acid residues of xylan; a protonated carbonyl group is giving rise to a doublet in the spectrum.[28,29] Therefore, the 4-*O*-methylglucuronic acid and hexenuronic acid moieties in the pulp samples examined here were deprotonated and converted to the sodium form by alkaline treatment prior to analysis.

In Figure 1, the spectra obtained from the same pulp sample following acidic or alkaline treatment are shown, together with the corresponding difference spectrum. It can be clearly seen that the peak present at ~1740 cm^{-1} after acidic treatment is eliminated by alkaline treatment and the peak at ~1610 cm^{-1} is increased in intensity. The difference spectrum reveals that there are no other differences between these two spectra.

After employing alkaline pulp treatment using these conditions, spectra were obtained from 29 different pulp samples and the carbohydrate compositions, K numbers, and corrected K numbers calculated from these spectra using a single PLS model.

Internal Validation of the PLS Model Employed. To determine the appropriate number of components to include in our PLS model, the approach referred to as cross-validation (CV) was used.[20] Here 6/7 of the original spectral data were utilized to construct the model, from which the remaining 1/7 of the data were predicted. Finally, these predicted values were compared with the actual ones.

For this purpose, the squares of the differences between predicted and observed values were added together to obtain the predictive residual sum of squares (PRESS), which is a measure of the predictive power of the model being tested. Usually, PRESS is re-expressed as Q^2, which is (1-PRESS/SS), where SS is the sum of squares of *Y*, after mean-centering. A quantitative measure of the quality of the fit is provided by the parameter R^2 (i.e., the explained variation), and the ability to predict is indicated by the prediction parameter Q^2 (i.e., the predicted variation). Using four

Table 3. Statistical Data Obtained in Connection with Calibration and Validation of the PLS Model

component	range	mean	calibration				validation	
			R^2Y	Q^2	R^a	RMSEE[b]	R^a	RMSEP[c]
xylose	1.9-32[d]	20.2[d]	0.94	0.92	0.94	2.72	0.93	2.78
glucose	65-95[d]	76.1[d]	0.87	0.85	0.87	2.79	0.80	3.27
mannose	0-12[d]	3.05[d]	0.84	0.78	0.84	1.54	0.88	1.28
arabinose	0-1.2[d]	0.27[d]	0.83	0.76	0.84	0.18	0.80	0.19
galactose	0-0.65[d]	0.12[d]	0.91	0.88	0.91	0.05	0.89	0.06
hexenuronic acid	0-1.8[d]	0.85[d]	0.71	0.68	0.71	0.33	0.74	0.30
4-O-Me-glucuronic acid	0-1.02[d]	0.39[d]	0.58	0.50	0.58	0.20	0.42	0.22
K number	1-90[e]	14.7[e]	0.95	0.93	0.95	4.43	0.86	7.54
corrected K number	0-86[e]	10.7[e]	0.97	0.95	0.97	3.56	0.87	7.18
xylan	2-34[d]	21.7[d]			0.95	2.50	0.94	2.78
glucomannan	0-15[d]	3.96[d]			0.81	2.02	0.86	1.75
cellulose	64-94[d]	75.3[d]			0.81	2.94	0.74	3.45

[a] The correlation coefficient for the observed value as a function of the predicted value. [b] The root-mean-square error of estimation. [c] The root-mean-square error of prediction. [d] The weight percent. [e] Arbitrary units.

principal components, the PLS model developed here on the basis of our photoacoustic FT-IR spectra could explain 85% of the variance in the **X** matrix and 81% of the variance in the **Y** matrix.

One limitation of this cross-validation procedure is that although it assesses the predictive power, it does not evaluate the statistical significance of the estimated predictive power.[20] However, a validate diagram can be constructed in order to test whether the model obtained by cross-validation is unique or overfitted, i.e., is modeling noise. In the case of the calibration set, the spectral data are maintained intact, whereas the responses are permutated randomly to appear in a different order. This permutation is repeated several times and PLS models are then fitted to the permutated responses and, using cross-validation, both R^2Y and Q^2 values for the derived models are calculated. Comparison of these permutated R^2Y and Q^2 values using a validate plot to the estimated values of R^2Y and Q^2 for the "real" model then provides a first indication of the statistical significance of the model.

In the validate diagram, R^2Y and Q^2 values are plotted on the Y axis; while the X axis designates the coefficients of correlation between the original and permuted response variables. The resulting models should be as far as possible from the original model, i.e., as close as possible to the origin, to render the original model unique. If R^2Y and Q^2 are close to the corresponding values for the original model, this model is of poor quality and almost any variable can then be modeled. For valid models, the R^2Y intercept should not be greater than 0.3−0.4 and the Q^2 intercept should be ≤0.05. After 20 permutations of our four-dimensional PLS model, this model fulfills these criteria for validity.

Evaluation and External Validation of the PLS Model. The statistical data extracted from calibration and validation of the model are documented in Table 3. By investigating the R^2Y and Q^2 values of the individual responses, it was possible to determine for which responses the model worked well and less well. R^2 and Q^2 should be as large as possible, and in general, a Q^2 value of 0.5 may be regarded as good and a Q^2 value of 0.9 as excellent, depending on the particular application. It is important that the difference between R^2Y and Q^2 not become larger than 0.2−0.3.

Table 3 reveals that all of the responses, with the exception of 4-O-methylglucuronic acid, are described well by the model. The xylan, glucomannan, and cellulose contents of the test set were

calculated using the PLS model and these predicted values correlated well with the measurements, with correlation coefficients R close to 0.9 in most cases for the observed values plotted against the predicted values.

RMSEE is the root-mean-square error of estimation for the calibration set. The predictive ability of the model is assessed in terms of the root-mean-square error of prediction (RMSEP) for the test set, which is defined as follows:

$$\text{RMSEP} = \sqrt{\sum_{i=1}^{n_t} \frac{(\hat{y}_i - y_i)^2}{n_t}}$$

where n_t is the number of objects in the test set, y_i the known value of the parameter of interest for sample i, and $\hat{y}_i$ the value of the parameter of interest predicted by the model for sample i. The RMSEE and RMSEP values can be viewed as a measure of the standard deviation and should therefore be as small as possible. In our case, the K number and the corrected K number exhibit the largest RMSEP, but these values also span the greatest ranges. It is important to note that there are a number of extreme values within these ranges, especially in the case of the K number, which affect the mean values. The root-mean-square errors for determination of the major components of pulp in this manner vary from 0.06 to 7.54 wt %.

Altogether, these findings demonstrate that the FT-IR-PAS technique can be reliably used to predict the contents of xylose, glucose, mannose, arabinose, galactose, and hexenuronic acid residues, as well as the content of lignin measured in terms of K number and corrected K number for the pulp samples investigated here. Furthermore, the contents of xylan, glucomannan, and cellulose in these pulps can also be predicted accurately from these primary predictions. The content of 4-O-methylglucuronic acid residues was, however, more difficult to predict accurately, possibly reflecting the low abundance of this sugar in the pulps examined.

Characteristic Bands in the IR Absorption Spectrum. In the present study, we have utilized a tool for model interpretation referred to as the variable influence on projection parameter[20] (VIP), which facilitates interpretation of the PLS model consider-
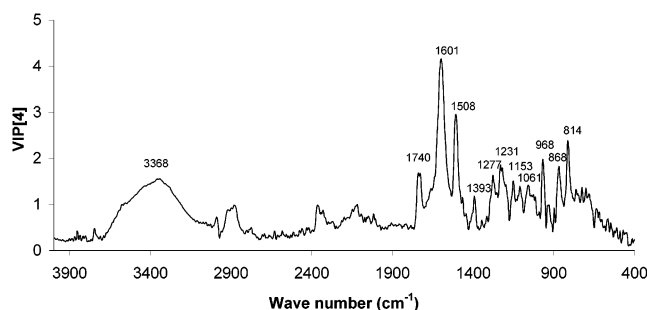
Figure 2. VIP plot for our four-component PLS model. This plot reveals which wavenumbers (variables) influence the model most strongly.

Table 4. Assignment of Bands in the Infrared Spectra of Wood and Pulp Fibers

wavenumber (cm⁻¹)	assignment	ref
814	glucomannan	31
870	mannose	32
872	glucomannan	31
898	cellulose, glucomannan	31
967	C—O stretch	33
1230	lignin, C=O in xylan	14
1400	cellulose	32
1505	lignin	14
1510	lignin	34−37
1595	lignin	37
1600	lignin	14
1600	xylan (O-groups)	32
1605	carboxylate groups	38
1605	C=C	37
1740	C=O stretch in xylan	36, 37
2900	cellulose	36
3350	cellulose	36

ably. VIP is the weighted sum of the squares of the PLS weights, w^*, taking into account the amount of Y variance in each dimension that can be explained by the model. A single VIP vector summarizes all of the components and responses.

The VIP plot for the present data set depicted in Figure 2 reveals which wavenumbers (variables) influence the model most strongly. Another approach to evaluating the influence of the spectral data on the responses is to evaluate the coefficients. The magnitude and sign of the scaled and centered regression coefficient indicate the influence of this term in the model.

By investigating the VIP plot and the coefficients of the model, characteristic IR bands were confirmed for lignin, glucomannan, xylan, and cellulose (see Table 4). Comparison of the coefficients of the model with one another reveals which parameters are correlated and which are not. The xylose content is correlated to the content of hexenuronic acid but not to the glucose content. The K number and corrected K number are strongly correlated to one another, as are the contents of glucose, mannose, arabinose, and galactose.

These correlations are reasonable, since a pulp containing a relatively high content of xylose will also contain a larger number of hexenuronic acid residues and smaller amounts of glucose and glucomannan. The coefficients of the K number and corrected K number are, of course, very much the same. Furthermore, the contents of arabinose and galactose residues are correlated to one
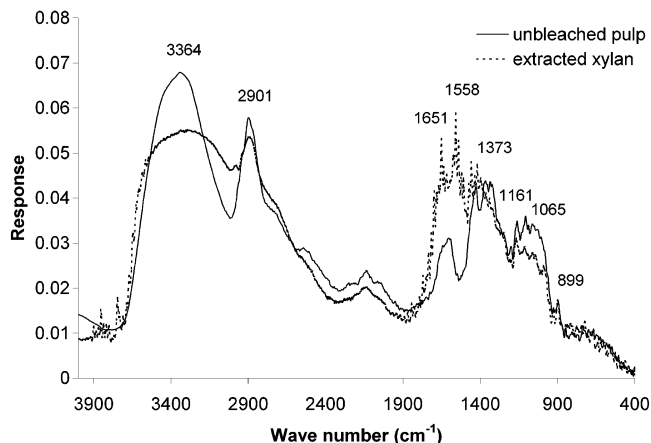

Figure 3. FT-IR-PAS spectra for an unbleached pulp sample (solid line) and for the xylan extracted with alkali from this same pulp (dotted line). The spectrum of the extracted xylan is noisy, probably due to the presence of water vapor in the extracted powder, as also indicated by the broader signal at ∼3500 cm⁻¹. Nevertheless, it can be seen that the important infrared absorption region for xylan is around 1600 cm⁻¹.

another, probably because these sugars are both present in softwood pulps.

In a separate experiment, xylans were extracted from samples of unbleached and oxygen-delignified pulps with potassium hydroxide. These extracts were then analyzed by employing FT-IR-PAS and the spectra thus obtained compared to the spectra of the original pulps (see Figure 3, where the spectra for unbleached pulp and its xylan extract are shown). The spectrum of the extracted xylan is very "noisy", probably due to the presence of water vapor in the extracted powder, as also indicated by the broader signal at ∼3500 cm⁻¹. Nevertheless, it can be seen that the important infrared absorption region for xylan is around 1600 cm⁻¹.

To elucidate the differences between a pulp containing a relatively large amount of hexenuronic acid and another pulp with a high content of 4-*O*-methylglucuronic acid, a difference spectrum was generated (Figure 4). Since the pulp that contains a higher level of 4-*O*-methylglucuronic acid was fully bleached, this sample also contained more cellulose. Characteristic signals for cellulose are thus observed at 2900 and 3350 cm⁻¹. The difference between hexenuronic acid and 4-*O*-methylglucuronic acid appears as a small signal at 1601 cm⁻¹, which is in agreement with the coefficient plot, where hexenuronic acid exhibit a characteristic IR absorption band at 1600 cm⁻¹.

Classification of Softwood and Hardwood Pulps. Score plots and loading plots were constructed, and on the basis of these plots, classification of the pulps could be achieved. Hardwood and softwood pulps are separated by the first score plot (Figure 5A). When this score plot is compared to the corresponding loading plot (Figure 5B), it can be seen that pulps in the left half of the figure demonstrate high K numbers, i.e., contain large amounts of lignin. These pulps are unbleached liner paper pulps and contain so much lignin that they are, in fact, "outliers".

The upper portion of the score plot contains the softwood pulps with large amounts of glucose and mannose, i.e., higher contents of cellulose and glucomannan. The hardwood pulps are located in the lower region of this score plot and contain larger amounts
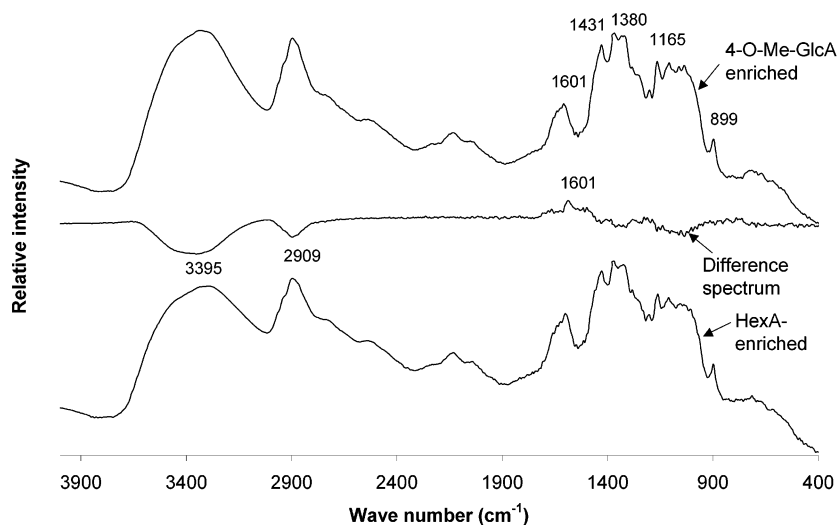
Figure 4. Spectra and the difference spectrum of pulps containing a relatively large amount of 4-*O*-methylglucuronic acid or a high content of hexenuronic acid residues. Since the pulp that contains a higher amount of 4-*O*-methylglucuronic acid was fully bleached, this sample also contained more cellulose. Characteristic signals for cellulose are thus obtained at 2900 and 3350 cm^{-1}. The difference between hexenuronic acid and 4-*O*-methylglucuronic acid appears as a small signal at 1600 cm^{-1}, which is in agreement with the coefficient plot, where hexenuronic acid exhibit a characteristic IR absorption band at 1600 cm^{-1}.
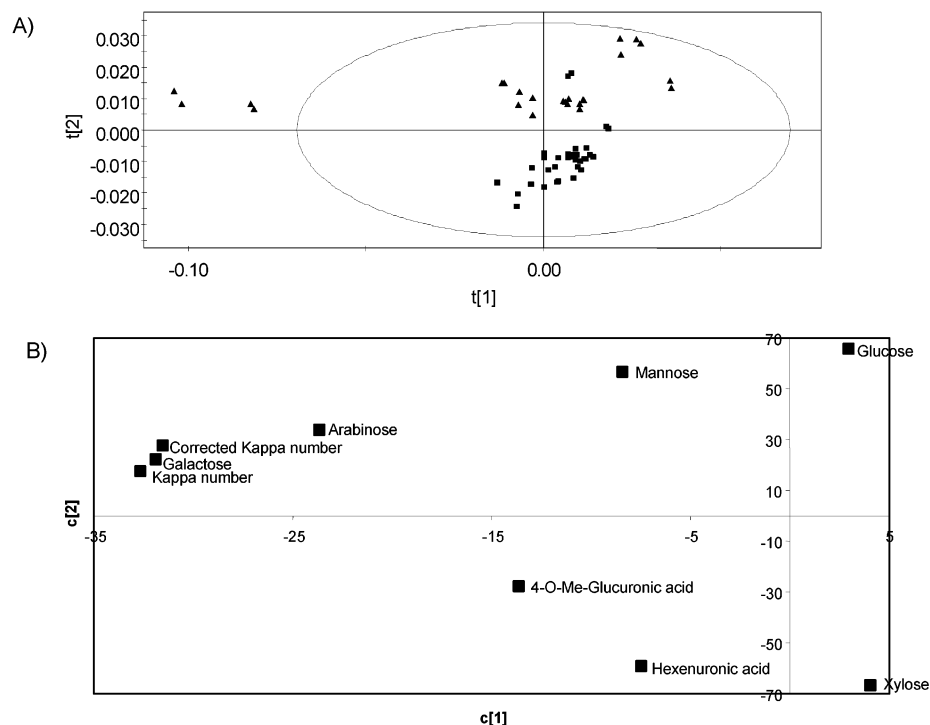

Figure 5. Score plot (A) for hardwood (■, solid gray box) and softwood pulps (▲) and (B) the loading plot for the responses. It can be seen that pulps in the left half of the score plot demonstrate high K numbers, i.e., contain large amounts of lignin. These pulps are unbleached and contain so much lignin that they are, in fact, "outliers". The upper portion of the score plot contains the softwood pulps with large amounts of glucose and mannose, i.e., higher contents of cellulose and glucomannan. The hardwood pulps are located in the lower region of this score plot and contain larger amounts of xylose, hexenuronic acid, and 4-*O*-methylglucuronic acid (i.e., xylan), than do softwood pulps.

of xylose, hexenuronic acid, and 4-*O*-methylglucuronic acid (i.e., xylan) than do the softwood pulps. This same pattern is observed when the test set is predicted by employing the PLS model. The hardwood pulp sample present in the region containing softwood pulps (gray dots in the score plot) is a sulfite pulp, which is similar to the softwood pulps, i.e., contains a relatively low amount of xylose and higher contents of glucose and mannose, as determined by CZE analysis.

CONCLUSIONS

A rapid procedure based on photoacoustic FT-IR and PLS has been developed for the analysis of carbohydrates and lignin in chemical pulps. The total time required for analysis is 30−40 min per sample (in addition to the time required for sample preparation).

It is shown here that, in combination with PLS, this FT-IR-PAS approach can be used to predict the contents of

xylose, glucose, mannose, arabinose, galactose, and hexen-uronic acid residues, as well as the K numbers and corrected K numbers of pulp samples. From these predictions, the contents of xylan, glucomannan, and cellulose in the pulps can also be predicted.

The analytical method developed here exhibits several advantages over other procedures. The only sample preparation required is conversion of the pulp to its sodium form by washing with alkali and deionized water and then air-drying. The morphology of the sample has little or no impact on the results obtained.

Since a reliable calibration has been performed, the analytical procedure developed here can be used on a routine basis to quantify pulp constituents with considerably less effort and in shorter time than is possible using chemical analysis.

SUPPORTING INFORMATION AVAILABLE

FT-IR-PAS spectra obtained from fiber sheets perpared from a pulp sample by employing two different procedures, i.e., the standard and the slurry procedures; validate diagram for the K numbers obtained with our PLS model (it can be seen from this validate plot that, after 20 permutations, our four-dimensional PLS model fulfills the criteria for validity); coefficient plots for the present PLS model. This material is available free of charge via the Internet at http://pubs.acs.org.