

Nondestructive Analysis and Dating of Historical Paper Based on IR Spectroscopy and Chemometric Data Evaluation

Tanja Trafela,[†] Matija Strlič,^{*,†} Jana Kolar,[‡] Dirk A. Lichtblau,[§] Manfred Anders,[§] Danijela Pucko Mencigar,[†] and Boris Pihlar[†]

University of Ljubljana, Faculty of Chemistry and Chemical Technology, Aškerčeva 5, SI-1000 Ljubljana, Slovenia, National and University Library, Turjaška 1, SI-1000 Ljubljana, Slovenia, and Zentrum für Bucherhaltung, Mommsenstrasse 7, 4329 Leipzig, Germany

Sampling restrictions in analysis of cultural heritage materials narrow the choice of appropriate analytical methods considerably. In this work, near- and mid-FT-IR reflectance data were related to paper properties determined with classical analytical methods using partial least-squares. Nondestructive determination of properties, which are of importance for evaluation of the long-term stability of historical paper, i.e., ash content, lignin content, degree of polymerization of cellulose, pH, and aluminum content, is possible. With the use of a considerable sample set, satisfactory reliability was achieved for all properties but aluminum content. Considering that with age, chemical properties of paper change, dating of historical documents was attempted for the first time, also with success.

Rightfully so, sampling of historical artifacts for analysis is rarely permitted. More and more often, only nondestructive or at best microdestructive analytical techniques are allowed. In this journal, we could follow the discussion of the Vinland Map, possibly a priceless document. It is evident, how cautious the approach to its analysis has been and how the value of information gained is carefully weighted against the loss of historical material destroyed in the course of analysis.^{1–6}

The analysis of cultural heritage objects has become an increasingly complex line of research in analytical chemistry. The application of many advanced analytical tools is attempted not long after introduction, synchrotron analysis being a good example.^{7–10}

However, the material, of which a work of art consists, is complex not only because it is usually very heterogeneous, but also because it underwent various degradation processes through time in unknown storage conditions.

In museums, archives, and libraries, the application of analytical techniques in the absence of chemicals or solvents is preferred. The time of analysis should be short, as occasionally a considerable number of items are surveyed to establish the state of preservation of a collection. If possible, the instrument should be easy to use by nonprofessionals.

For such a purpose, near-IR (NIR) spectroscopy is appropriate.¹¹ The use of mid-IR spectroscopy in material analysis is widespread, and NIR spectroscopy is gaining in importance.^{12–14} However, real (aged) paper is often too complex to allow for an exact analytical interpretation of spectra, a typical example being the determination of lignin in paper, where many IR spectroscopic methods exist based on different interpretations of spectra.^{15–17}

The basic feature of NIR spectra is that they exhibit low structural selectivity as opposed to mid-IR and Raman techniques. The spectra are dominated by overtones and combination vibrations, especially of NH, CH, and OH functionalities. In order to extract complex spectral information, chemometric analysis of data is a widespread approach instead of band assignment,¹⁸ which can be successful if care is taken to avoid misleading or overly optimistic results.¹⁹ Partial least-squares (PLS) allows for correlation of spectral and chemical information.^{20–22} With the use of PLS,

* To whom correspondence should be addressed. E-mail: matija.strlic@fkkt.uni-lj.si.

[†] University of Ljubljana.

[‡] National and University Library.

[§] Zentrum für Bucherhaltung.

- (1) McCrone, W. C. *Anal. Chem.* **1976**, *48* (8), 676A–679A.
- (2) Brown, K. L.; Clark, H. J. R. *Anal. Chem.* **2002**, *74* (15), 3658–3661.
- (3) Olin, J. S. *Anal. Chem.* **2003**, *75* (23), 6745–6747.
- (4) Clark, H. J. R. *Anal. Chem.* **2004**, *76* (8), 2423.
- (5) Henchman, M. *Anal. Chem.* **2004**, *76* (9), 2674.
- (6) Graham, R. *Anal. Chem.* **2004**, *76* (21), 407A–412A.
- (7) Harbottle, G.; Gordon, B. M. *Nucl. Instrum. Methods Phys. Res., Sect. B* **1986**, *14* (1), 116–122.
- (8) Zappala, A.; Bajt, S.; Gigante, E. G.; Hanson, L. A. *Nucl. Instrum. Methods Phys. Res., Sect. B* **1996**, *117* (1–2), 145–150.
- (9) Wilkinson, J. T.; Perry, L. D.; Martin, C. M.; McKinney, R. W.; Cantu, A. A. *Appl. Spectrosc.* **2002**, *56* (6), 800–803.

- (10) Arčon, I.; Kolar, J.; Kodre, A.; Hanžel, D.; Strlič, M. *X-Ray Spectrom.* **2007**, *36* (3), 199–205.
- (11) Anders, M.; Behnsen, P.; Plew, P. *Preprints of the 10th IADA Congress*, International Association of Book and Paper Conservators: Göttingen, Germany, September 22–26, 2003; p 11.
- (12) Günzler, H.; Gremlich, H. U. *IR Spectroscopy. An Introduction*; Wiley-VCH: Weinheim, Germany, 2002.
- (13) Siesler, H. W.; Ozaki, Y.; Kawata, S.; Heise, H. M. *Near-Infrared Spectroscopy. Principles, Instruments, Applications*; Wiley-VCH: Weinheim, Germany, 2002.
- (14) Ali, M.; Emsley, M. A.; Herman, H.; Heywood, J. R. *Polymer* **2001**, *42*, 2893–2900.
- (15) Fardim, P.; Ferreira, C. M. M.; Duran, N. J. *Wood Chem. Technol.* **2002**, *22* (1), 67–81.
- (16) Vives, G. J.; Monmany, D. J.; Guerra, A. R. *Restaurator* **2004**, 47–67.
- (17) Antti, H.; Sjöström, M. J. *Chemom.* **1996**, *10*, 591–603.
- (18) Blanco, M.; Villarroya, I. *TrAC, Trends Anal. Chem.* **2002**, *21*, 240–250.
- (19) Small, G. W. *TrAC, Trends Anal. Chem.* **2006**, *25*, 1057–1066.
- (20) Brereton, G. R. *Chemometrics, Data Analysis for the Laboratory and Chemical Plant*; Wiley & Sons: Chichester, U.K., 2003.

more than one variable (e.g., content, concentration) can be used simultaneously avoiding the problem of collinearity.¹³ However, the quality of these correlations depends on a number of factors, among which the quality of spectra and the quality of chemical analytical data play a decisive role.

We report on the development of a multianalytical nondestructive approach to historical paper analysis using FT-IR and NIR with chemometric data evaluation. Based on the chemical and spectroscopic analysis of more than 170 historical samples from AD 1650 onward, we propose a method that enables us to characterize historical paper in view of ash content, aluminum content, lignin content, pH, and degree of polymerization (number of monomer units in the macromolecule) of cellulose. As paper aging is characterized by chemical processes taking place during long-term storage (e.g., decrease in degree of polymerization, building-up of degradation products), the approach may even enable us to date historical paper.

MATERIALS AND METHODS

Samples. A large number of historical paper samples were made available for destructive testing through acquisition or donations. Paper representative of the vast majority of library and archival material was sampled, while fungi-infested, burnt, excessively soiled, and densely printed samples or laminated, transparent, and other specialty papers were not considered.

The sample set included more than 170 randomly selected rag and wood pulp containing (13% bleached and 87% nonbleached) single sheets and books (62% sheets, 38% books; production, 25% handmade, 75% industrial; sizing, 12% gelatin, 88% rosin), dating from the 17th century till present time. The year of publication was considered as the year of sample production.

Not all samples were analyzed for all properties, e.g., degree of polymerization (DP) was not determined in lignin-containing papers, as they cannot be dissolved in the standard solvent. For each chemical property, each paper sample was analyzed two or three times, target uncertainty being <15%, as uncertainty is limited by the inherent inhomogeneity of the historical paper sample itself. The average of determinations was used for chemometric data evaluation.

To reduce the uncertainty, the reference analytical methods were performed using strips sampled from nonprinted page area near the book spine. Spectra were taken using the same strips. In the case of single sheets, printed or written areas and areas close to the border were avoided.

Degree of Polymerization. The standard viscometric method was used (ISO 5351/1) to determine DP of the cellulosic fraction of paper. The method is not suitable for groundwood-containing papers, as the samples cannot be dissolved in the solvent used for viscometry, i.e., bis(ethylenediamine)cupric hydroxide. DP was calculated from intrinsic viscosity using the Mark–Houwink–Sakurada equation:²³

$$DP^{0.85} = 1.1[\eta]$$

Aluminum Content. Aluminum content (used in rosin or gelatin sizing) was determined using atomic absorption spectroscopy after extraction of 30–50 mg of sample in 15% HNO₃ for 1 h using sonication at 65 °C.

Ash Content. Ash content was determined gravimetrically according to a modified (sample mass 0.1 g) standard method (ISO 2144:2000).

Lignin Content. Due to its complex structure, quantitative determination of lignin is problematic. We used the acetyl bromide colorimetric method.²⁴ After sample dissolution in 0.5 mL of a mixture of 2.5 mL of acetyl bromide, 10 mL of glacial acetic acid, and 0.5 mL of 70% perchloric acid and heating at 70 °C, 1 mL of 2 mol L⁻¹ NaOH solution in 2.5 mL of glacial acetic acid was added, and the absorbance was measured at 280 nm. For external standard calibration alkali lignin (Sigma-Aldrich, Steinheim, Germany) was used.

pH. Many arbitrary methods for determination of paper pH are in use.²⁵ Cold extraction of microsamples was performed in the following way: to 20–50 µg of sample, 5 µL of deionized water was added and left overnight. pH was determined in the extract using a micro-combined-glass electrode (MI 4152, Microelectrodes, Bedford, NH).

Spectroscopic Analysis. A Perkin-Elmer Spectrum GX (Waltham, MA) equipped with a 76 mm Labsphere RSA-PE-200-ID (North Sutton, NH) integration sphere coated with Infragold, with a DTGS detector, was used. The reflectance spectra were collected in the interval of 6500–500 cm⁻¹, 128 scans per sample.

The number of layers needed to obtain a spectrum without the influence of the supporting material was investigated. We determined spectra of one, two, three, and four layers of purified cellulose sheets and compared each spectrum with the one obtained with five sheets. The comparison was done by correlating absorbance values corresponding to the same wavenumbers. The regression coefficients for these linear correlations increase with the number of layers used (Table 1). The correlation of spectra using four and five layers was considered satisfactory; hence, four layers were used in subsequent work.

Chemometric Data Analysis. Independent sample sets were used for calibration and for validation. For calibration, two spectra were taken on two different positions of the same sample. Spectrum Quant+ software (Perkin-Elmer, Waltham, MA), partial least-squares analysis (PLS) was used to model the chemical properties. No analysis of the influence of absorption at particular wavenumbers on the quality of correlations was attempted. Optimal correlations were sought for only by using different data pretreatments and different wavenumber intervals.

To evaluate the quality of calibration, two standard errors were calculated. The standard error of estimate (SEE) is the average error between the measured property values and the estimated values of the calibrated standards from a given model. SEE can be described as the square root of the residual variance divided by the number of degrees of freedom. The standard error of prediction (SEP) is the average error between the measured and

(21) Garthwaite, H. P. *J. Am. Stat. Assoc.* **1994**, *89* (425), 112–127.

(22) Höskuldsson, A. *Chemom. Intell. Lab. Syst.* **2001**, *55*, 23–38.

(23) Evans, R.; Wallis, A. F. A. Cellulose molecular weight determined by viscometry. Proceedings of the Fourth International Symposium on Wood Pulping Chemistry, Paris, April 23–24, 1987.

(24) Iiyama, K.; Wallis, A. F. A. *Wood Sci. Technol.* **1988**, *22*, 271–280.

(25) Strlič, M.; Kolar, J.; Kočar, D.; Drnovšek, T.; Šelih, V. S.; Susič, R.; Pihlar, B. *e-Preserv. Sci.* **2004**, *1*, 35–47.

Table 1. Regression Coefficients (*R*) for Linear Correlations between Spectra Taken with One, Two, Three, or Four Layers of Purified Cellulose Sheets and the Spectrum Obtained with Five Sheets of the Same Paper

	<i>R</i>
1 layer	0.9943
2 layers	0.9973
3 layers	0.9987
4 layers	0.9993

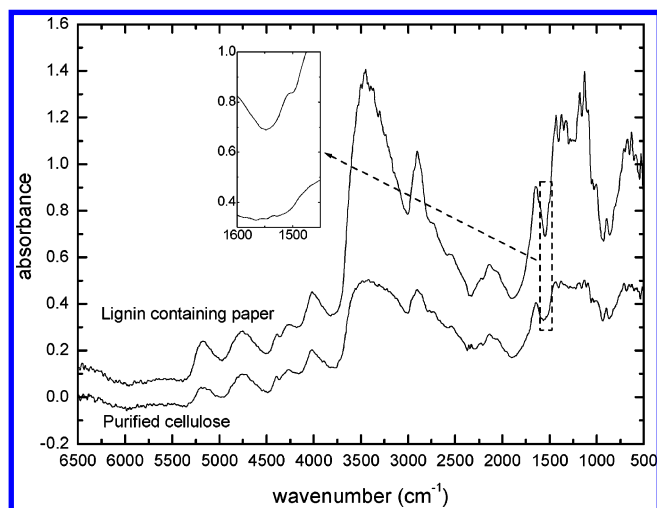


Figure 1. FT-IR reflection spectra of groundwood and purified cellulose sheets. The inset indicates a region of interest occasionally used for quantitative estimation of lignin content.

estimated property values of samples not used in the calibration. SEP gives an estimate of the magnitude of the error expected when independent samples are predicted using the developed model.

RESULTS AND DISCUSSION

IR spectroscopy is of major importance in paper characterization.²⁶ Spectra of two very different paper samples, however, may have very similar features, e.g., the lignin-containing groundwood pulp and the purified cellulose sample with no fillers or sizing (Figure 1). Often, the small absorption band at 1510 cm⁻¹ is used as a marker of the presence of lignin in paper and even used semiquantitatively.²⁷ However, given the big difference in lignin content of the two samples and the small difference in intensity of the lignin absorption band, it is evident that the analytical approach has to be complemented with chemometric spectral data evaluation, if we are to extract the necessary quantitative information from spectra.

Therefore, we chose to model the paper properties of interest, i.e., DP, pH, lignin, aluminum, and ash content. We used PLS to produce different models for each property (Figure 2).

Different spectral preprocessing strategies were tested and optimized with regard to the quality of calibration: derivation, baseline correction, smoothing, normalization and use of different wavenumber intervals. Calibrations of best quality for all properties

were obtained using untreated spectra in the wavenumber interval 6500–800 cm⁻¹. The parameters of the optimal PLS models are summarized in Table 2.

The content of ash is closely related to the content of inorganic fillers, the identity of which changed with technology of paper production, e.g., kaolin, CaCO₃, alum. To model this paper property, spectral data in the NIR interval gave optimal results.

Lignin content is of major interest in preservation of historical paper-based materials. It is a highly unstable natural organic polymer, and its limited photostability is important for exhibited objects. The influence of lignin on stability of cellulose in paper is not well understood; it may have a limited antioxidant effect.^{28,29} However, through its rapid oxidation, it may lead to accumulation of acids in the material, which have a destabilizing effect. To build the PLS calibration for lignin estimation, the interval of 6500–4000 cm⁻¹ led to best results. It is of interest that the fingerprint mid-IR region does not improve the model, although lignin has strong absorption bands in the interval of 100–1720 cm⁻¹.²⁹ The absorption due to the stretching–bending combination of OH near 5000 cm⁻¹, of CH near 4400 and 4300 cm⁻¹, and of aromatic CH overtones near 5950 cm⁻¹ must have a more decisive influence on the calibrations. Although many of these vibrations coincide with overtones and combination bands of cellulose OH and CH functionalities, PLS proves to be a viable approach to calibration.

Of all paper properties, acidity has the most important influence on the durability of this material. It is estimated that 70–80% of Western library collections are degrading rapidly due to the low pH of paper. While papers before ca. 1850 were produced using the traditional sizing process, which resulted in approximately neutral material, those produced between 1850 and 1990 frequently have pH of aqueous extracts less than 6, due to the use of alum–rosin size. This decreases the lifetime expectancy of paper from several thousand years to only several hundred years.³⁰ From 1990, the production of paper changed to alkaline sizing due to environmental reasons, and contemporary paper is again more stable. To build the method, we collected spectra of widely different samples with pH of aqueous extracts between 3.7 and 9.0. The obtained calibration is very satisfactory with the SEP equal to the uncertainty according to the standard procedure for determination of paper pH using cold extraction in water (ASTM D 778).

We used lignin-free samples to build a calibration for estimation of DP of cellulose, hence the smaller calibration and validation sets. Cellulose is the most important structural component in paper, and it has been shown that in bleached pulp sheets, DP correlates with mechanical properties,³¹ which is of high importance, as mechanical properties are relevant for the end user.

The calibration for estimation of aluminum content is less satisfactory. Aluminum was added to paper as a precipitating agent both during gelatin sizing and during rosin sizing. Alum is the major source of acidity in paper, although its content does not correlate with pH of aqueous paper extracts.

(26) Workman, J. J. *Appl. Spectrosc. Rev.* **2001**, *36* (2–3), 139–169.

(27) Calvini, P.; Gorassini, A. *Restaurator* **2002**, 48–66.

(28) Schmidt, A. J.; Rye, S. C.; Gurnagul, N. *Polym. Degrad. Stab.* **1995**, *49*, 291–297.

(29) Boeriu, C. G.; Bravo, D.; Gosselink, R. J. A.; van Dam, J. E. G. *Ind. Crops Prod.* **2004**, *20*, 205–218.

(30) Strlič, M.; Kolar, J. *Ageing and Stabilization of Paper*; National and University Library: Ljubljana, Slovenia, 2005.

(31) Zou, X.; Uesaka, T.; Gurnagul, N. *Polym. Degrad. Stab.* **1994**, *43* (3), 393–402.

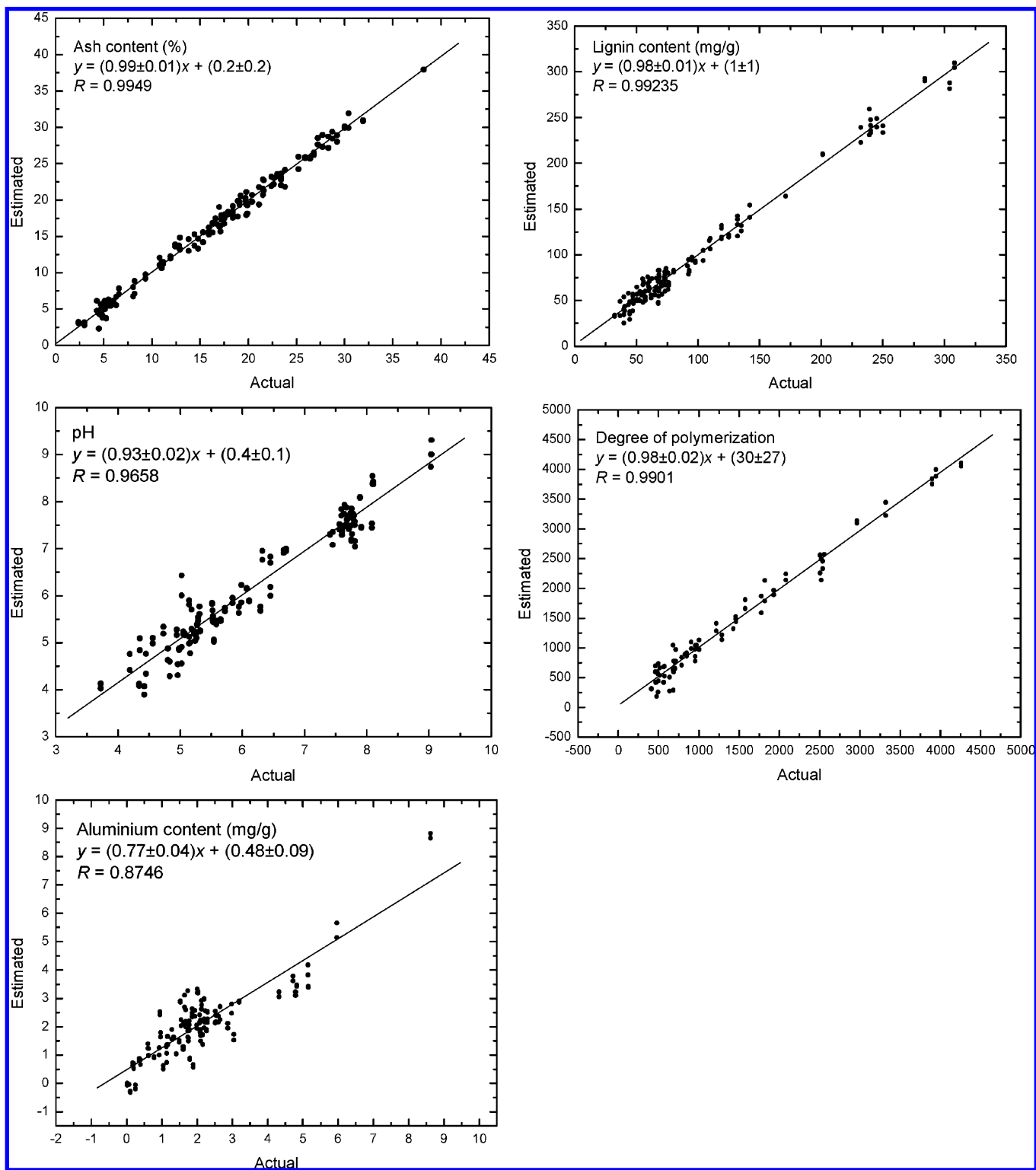


Figure 2. PLS calibrations for estimation of ash content, lignin content, pH, DP, and aluminum content of historical paper.

During history, paper production was a stable technological process with major changes introduced only in mid-19th century. Natural degradation proceeds at a very low rate and is influenced by intensive use, biodeteriogens, storage conditions, pollution, and inherent material instability. Under moderate care (e.g., absence of fungi, pests, excessive heat and humidity), as exercised in most libraries and archives, pH and possibly fiber source and production are among the most important parameters defining paper life-

time.³⁰ The aging of paper is thus reflected in changes in chemical composition proceeding with some kinetics: buildup of degradation products, decrease of DP, increase of crystallinity of cellulosic fibers, decrease of pH, etc.

These complex and interrelated changes are reflected in the IR spectra, although there has been no systematic study so far on how the natural aging process influences the chemical composition of paper. We could use these phenomena to our

Table 2. Summary of the Parameters of PLS Calibrations and Validations for Different Paper Properties (k —Slope, n —Intercept, N —Number of Samples, R —Correlation Coefficient)

paper property	wavenumber interval (cm ⁻¹)	calibration			validation			
		N	SEE	SEP	N	k	n	R
ash content	6500–4000	144	0.76	0.81	30	0.97 ± 0.04	-0.2 ± 0.9	0.9741
lignin content	6500–4000	144	8.9	9.3	30	1.01 ± 0.04	3 ± 6	0.9782
pH	6500–800	144	0.29	0.30	30	0.75 ± 0.04	1.1 ± 0.2	0.9573
DP	6500–800	86	161	172	21	1.06 ± 0.06	22 ± 87	0.9658
Al content	6500–800	144	0.72	0.73	30	0.6 ± 0.1	0.8 ± 0.2	0.7989
dating: pre-1850	6500–1000	24	6.6	8.6	4	2.0 ± 0.7	-1804 ± 1344	0.8850
dating: post-1850	6500–1000	174	8.4	8.6	30	1.08 ± 0.07	-150 ± 140	0.9418

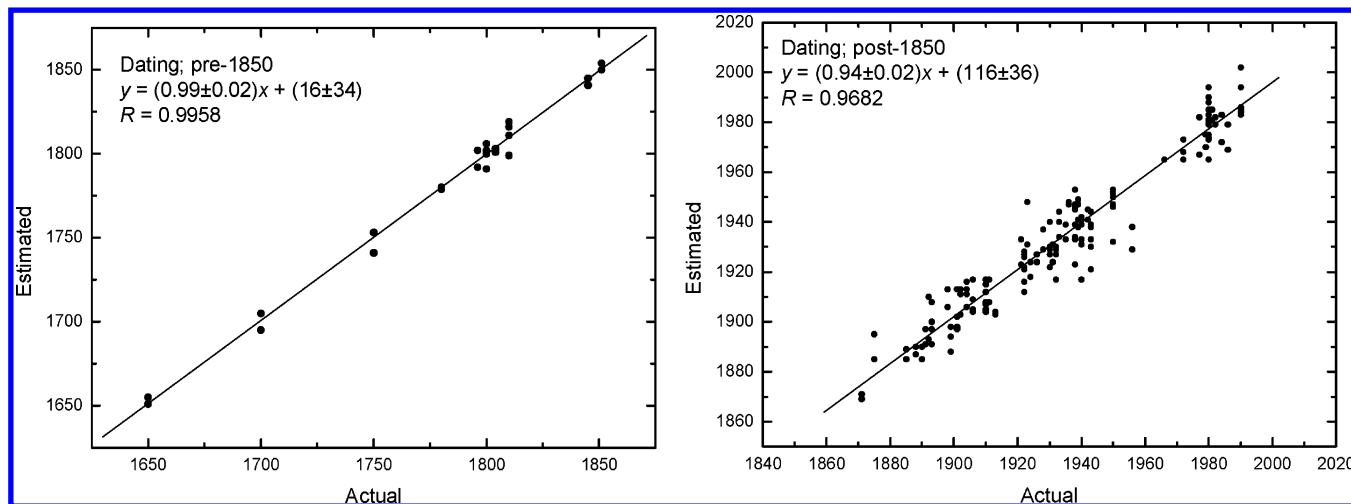


Figure 3. PLS calibrations for dating of paper: left, for pre-1850 papers; right, for post-1850 papers.

advantage, and in Figure 3 we show calibrations enabling us to date historical paper, and in Table 2 the statistical parameters are given. To decrease the uncertainty, we divided the calibration set in two subsets: pre-1850 and post-1850. This division is justified due to the change in paper production technology which occurred at approximately 1850. As the morphology of papers changed considerably at this date, selection of the appropriate calibration is not difficult for the operator. Due to the smaller size of calibration (24) and validation (4) sets for pre-1850 papers, the quality of validation for this parameter is less satisfactory (Table 2).

CONCLUSIONS

Using near- and mid-FT-IR in the reflectance mode, we collected spectra of a considerable number of historical papers. The selection of samples was random and spanned from 1650 to 2005.

All the samples were analyzed for their ash content, lignin content, pH, DP, and aluminum content using routine analytical methods. The results of these analyses and FT-IR spectra were used to produce PLS models. The correlations between predicted and measured properties were satisfactory.

IR spectroscopy can easily be performed in a museum, library, or archival environment. If combined with chemometric evaluation

of spectral data, we obtain a powerful tool for recognition, analysis, and perhaps even dating of paper. This approach will undoubtedly find practical application for a number of historical materials.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the European Community, Sixth Framework Energy, Environment and Sustainable Development Programme, Contract No. SSPI-006594 (SurveNIR) and Contract No. SSPI-006584 (PaperTreat). The work is the sole responsibility of the authors and does not represent the opinion of the Community. The Community is not responsible for any use that might be made of the data appearing herein. Further financial support has been obtained from the Ministry of Higher Education, Science and Technology of Slovenia, Program No. P1-0153. For donations of some historical samples, the authors are grateful to project partners and to Henk Porck, Doris Müller-Hess, Nina Hasselberg-Wang, Jozef Hanus, Tom Koch Svenessen, Istvan Kecskemeti, and Birgit Reissland.

Received for review February 25, 2007. Accepted June 11, 2007.

AC070392T