PEER REVIEWED

Detection of kappa number distributions in kraft pulps using NIR spectroscopy and multivariate calibration

HENRIK ANTTI, DANIEL ALEXANDERSSON, MICHAEL SJÖSTRÖM, AND LARS WALLBÄCKS

ABSTRACT

Chemical pulp is characterized by its average lignin content, commonly expressed as the pulp's kappa number. However, this average kappa number provides no information about the distribution of kappa number within the pulp. This study proposes a new method of pulp characterization using near-infrared reflectance (NIR) spectroscopy to measure distributions of kappa numbers within pulp samples. Pure pulps with different kappa numbers were mixed to create blended samples with a known nonuniformity of kappa number distribution. NIR spectroscopy—combined with multivariate calibration methods—was used to detect distributions of kappa numbers in the pulps. Models calculated from these data gave good predictions of the average kappa number as well as the standard deviation around the average. The results imply that NIR spectroscopy can provide information about the average kappa number as well as the distribution of kappa number within the pulp.

Application:

A proposed test method that measures average kappa number while also estimating a standard deviation around the average.

PULP QUALITY IS LARGELY A FUNCTION of the properties of the wood raw material. Chemical pulping of wood chips is a lignin-removal process, and longer cooking times remove greater amounts of lignin. However, variation in the size and thickness of wood chips affects the extent of chemical diffusion (and thus lignin removal) within the chips *(1)*, resulting in the production of inhomogeneous pulps. Ideally, mills could minimize variation in lignin content (commonly expressed as the pulp kappa number) by using very thin wood chips of similar size *(2)*.

Inhomogeneous raw material is a problem facing many industries. However, the real problem may not be so much the inhomogeneity as it is the lack of methods to control this variation. If inhomogeneity is to be controlled, the first step will be to develop methods to detect it.

BACKGROUND

Raw materials are commonly characterized using measurement techniques that provide an average value for a given property. However, while the average value is a useful measure of central tendency, it provides no information about the dispersion of values within the data set. In the absence of information about variations in raw material properties, variations in product quality are unavoidable.

A desire to reduce variation in product quality has moved the pulp and paper industry to search for more effective methods of characterizing the pulp. Kappa number, which is a measure of the lignin content in the pulp, is a widely used variable for pulp characterization. However, kappa number measurements provide only average values for an inhomogeneous pulp. It would be useful to identify a method that could also quantify inhomogeneous distributions of kappa number within the pulp.

Pulp inhomogeneity can be seen as a distribution of bonding capacity within the pulp. As the cook proceeds, the relation between different types of chemical bonds in the cellulose structure varies *(3, 4)*. Thus near-infrared reflectance (NIR) spectroscopy—which is sensitive to overtones of different vibration modes, such as stretching and bending of functional groups in organic molecules *(5)*—shows promise as a method for detecting the inhomogeneity in pulp.

The possibility of using the distribution of pulp kappa number as a descriptor for the pulp inhomogeneity introduces a new dimension to the utilization of kappa number. Moreover, measuring variations related to the kappa number distribution might open new possibilities for monitoring and controlling the production of pulp with specific properties.

The need for a fast and process-related method also makes application of NIR spectroscopy an interesting possibility *(6)*. In previous studies, NIR spectroscopy—in combination with multivariate calibration methods—has proved to be a reliable predictor of pulp kappa number *(7)* and has demonstrated its utility in other cellulose applications *(8-10)*. The present study evaluates the use of NIR spectroscopy and multivariate calibration to detect distributions of kappa numbers within a range of pulp samples with known levels of nonuniformity.

MATERIALS AND METHODS

Selection of wood chips

Wood chips were carefully selected to reduce kappa number variation in the pulp samples *(2)*. Selected wood chips were of equal size, with a maximum thickness of 3 mm.

Pulping

A total of 13 separate kraft pulps were prepared using a standardized autoclave cooking technique. The pulps were cooked to specified kappa numbers ranging from 30 to 90 in steps of five kappa number units.

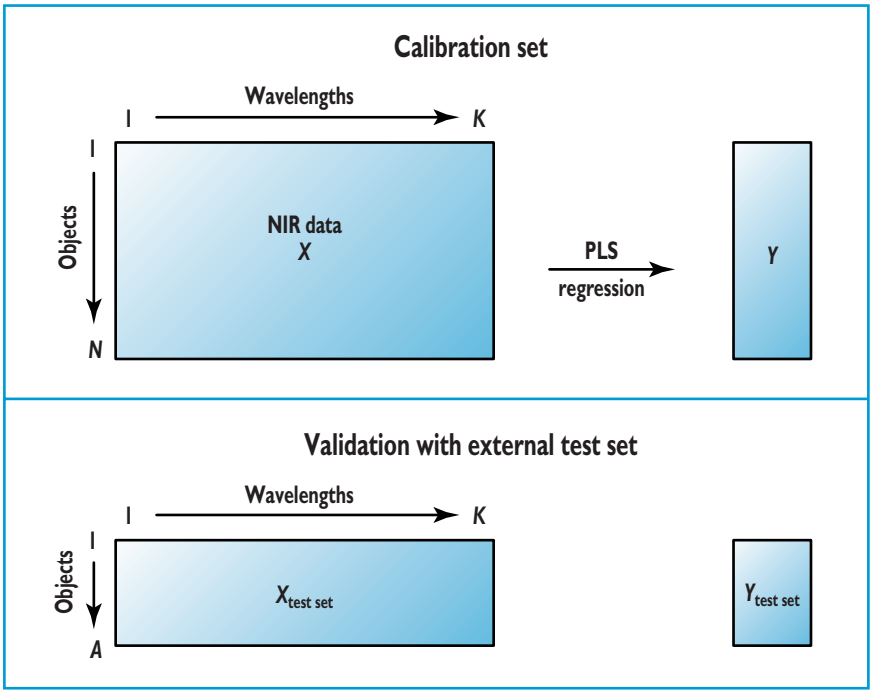
Wood chips (330 g) were dried and placed in each autoclave. Before the actual cooking, the chips were treated with 110°C steam for about 10 min. The liquor-to-wood ratio in the autoclave was 3:1, with an effective alkali content of 15% and a sulfidity of 32%. Cooking time for each batch varied, depending on the kappa number target.

At the end of each cook, the autoclave was removed from the reactor and cooled for about one minute with cold water before relieving the pressure. Cooked wood chips were removed from the autoclave and washed gently in water at 50°C for two hours and then left lying in water overnight prior to refining. After refining and beating, pulp kappa number was measured in accordance with SCAN C 1:77 *(11)*.

Pulp mixing

Kappa number for the 13 laboratory kraft pulps ranged from 30 to 90 in steps of five kappa number units. Given the careful selection of wood chips, the kappa number for each pulp was assumed to have a narrow distribution around the average.

To obtain a controlled inhomogeneity within the pulp samples, the 13 pure pulps were blended by weight in different proportions. This mixing produced a range of inhomogeneous pulp samples with known levels of nonuniformity in kappa distribution.



1.* Top:** ***predictor matrix consisting of* K *variables (wavelengths) and* N *samples is used to find correlation with some measured property* Y *via PLS regression;* bottom:** ***test matrix consisting of* K *variables (wavelengths) and* A *samples (not included among* N *calibration samples) is used to validate correlation

Two sets of samples were prepared. The first set consisted of samples with varying kappa number distributions but with the same average kappa number (kappa no. 60). This gave pulps that varied in homogeneity but, when analyzed by the traditional kappa titration method, gave the same response of kappa no. 60. The second set consisted of samples with varying kappa number distributions and different average kappa numbers. This approach provides a more realistic process simulation, since the average kappa number often varies with time during the pulping process.

Mixing was performed according to a mixture design *(12, 13)* with the constraint that the average kappa number for all mixtures had to lie between 40 and 80. A total of 20 pulp mixtures with varying kappa distribution and average kappa number were obtained by mixing 6 of the 13 pure pulps (kappa nos. 35, 45, 55, 65, 75, and 85). Pulp handsheets were formed from each mixture and then subjected to NIR analysis.

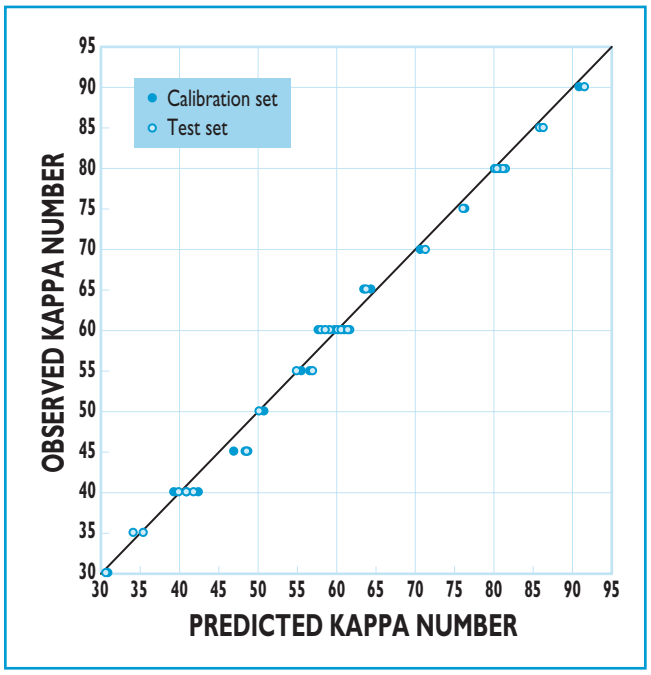
NIR measurements

NIR measurements were obtained using a scanning spectrophotometer (NIR Systems 6500) that measures 1200 wavelengths in the visible–NIR region (400–2500 nm). The sample cell is designed to make measurements over a large area (51 cm^2), which provides a better description of the sample than a standard sample cell (10 cm^2).

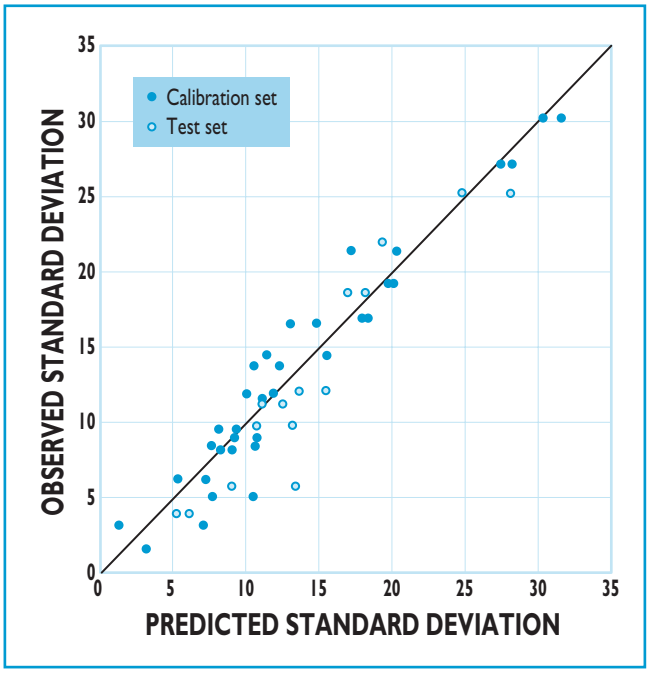
MULTIVARIATE MODELING

Calibration

When making calibrations from spectroscopic data, only one or a few wavelengths are typically used to estimate a given response variable. However, since a spectrum consists of many wavelengths (in this case 1200), all of which carry information about the measured sample, it makes sense to use all of the wavelengths as variables in the calibration. Multivariate calibration *(14, 15)* involves the use of the whole spectrum. This gives an $N \times K$ predictor matrix X, where N is the number of objects (samples) and K is the number of variables (wavelengths). The re-



2. Observed vs. predicted average kappa number for calibration and test sets (Model 1)



3. Observed vs. predicted standard deviation of pulp kappa number for calibration and test sets (Model 2)

sponse variable Y is often some property (or properties) measured on the samples using some reference method (or methods). The relationship between the predictor matrix and the response variable are depicted in **Fig. 1**.

The whole idea of creating calibration models from spectroscopic data is to find a method for predicting these measured properties that is simpler, faster, and more repeatable than the reference method or methods. Since wavelengths in spectra are highly correlated, the regression method used must be able to handle correlated X-variables. Partial least-squares (PLS) regression to latent structures *(16, 17)* is a regression method that solves the problem with correlated data by using the latent structures in X and Y.

Model validation

Model validation is crucial to verify the results of multivariate calibration. Cross validation *(18, 19)*—a commonly used method for validating multivariate calibration models—uses the samples already present in the model. This method provides information about the complexity of the model as well as its predictive ability, expressed as Q^2.

Since cross validation is based on the calibration data, an external test set should always be used to examine the model's predictive ability. An external test set consists of a number of A samples (objects) with known responses Y not included in the calibration model, as seen in Fig. 1. These results can thus be used to calculate the actual prediction error for "real" samples. This prediction error for the test-set objects, in turn, can be used to calculate an RMSEP (root mean square error of prediction) value *(14)*, which is a common parameter for describing prediction errors in multivariate calibration. Because samples from the external test set have not been used in the calibration, they provide an unbiased indicator of the model's predictive ability.

In this study, all models were validated using both cross validation and an external test set.

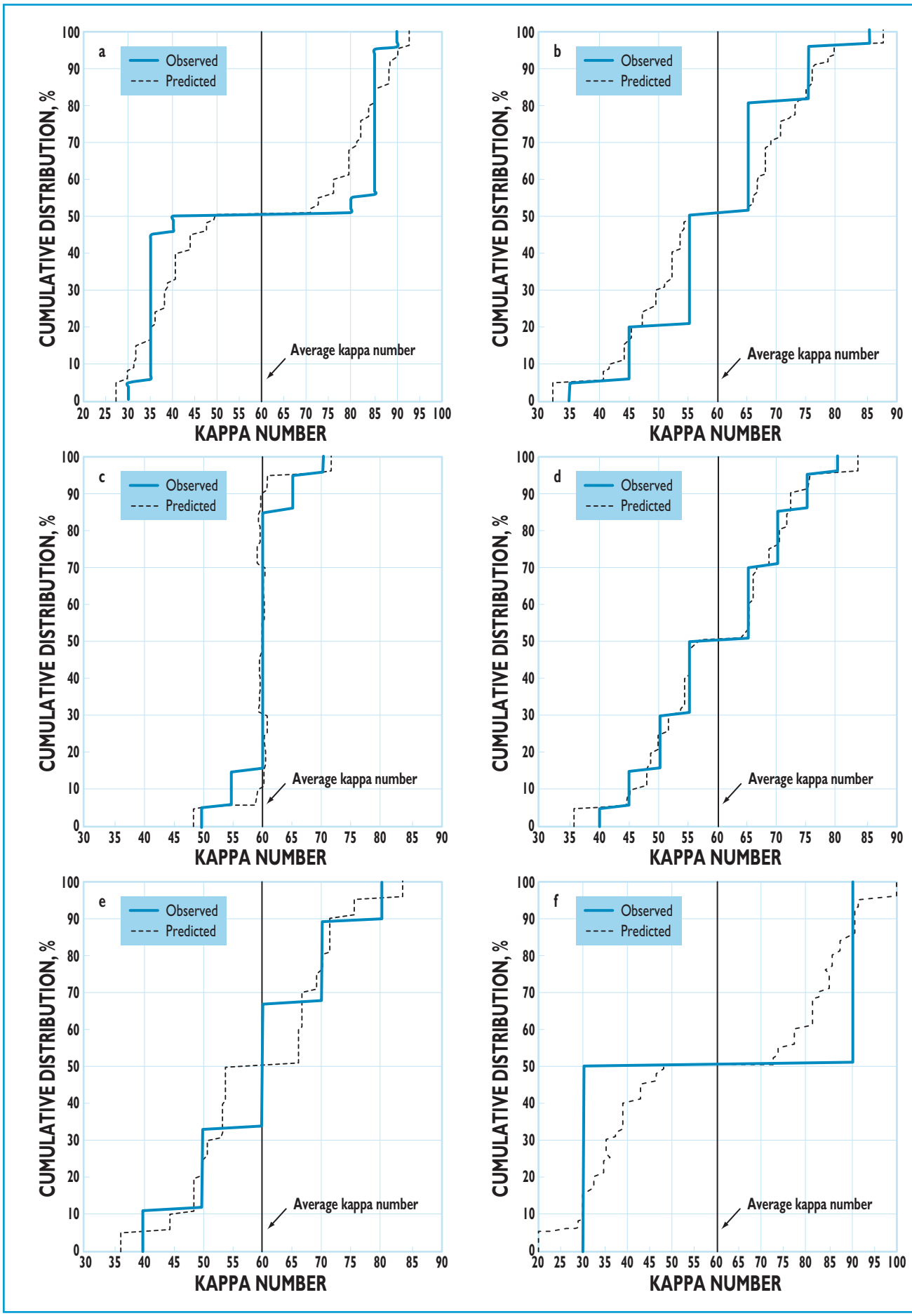
Response matrix Y

The response matrix Y was constructed to provide a clear description of the kappa distribution within the pulp samples. To accomplish this, each sample or pulp mixture was described by a row vector consisting of 100 elements. Assuming that the total content of kappa numbers in each sample or mixture reaches 100%, then each element in the vector provides an equal contribution of 1%. In this way, a sample constructed by mixing pulps with different kappa numbers will provide a row vector with 100 elements, with each constituent kappa number contributing the same number of elements as its percentage occurrence in the mixture. This approach provides a Y-matrix with 100 Y-variables, while the number of objects equals the number of pulp samples or mixtures included.

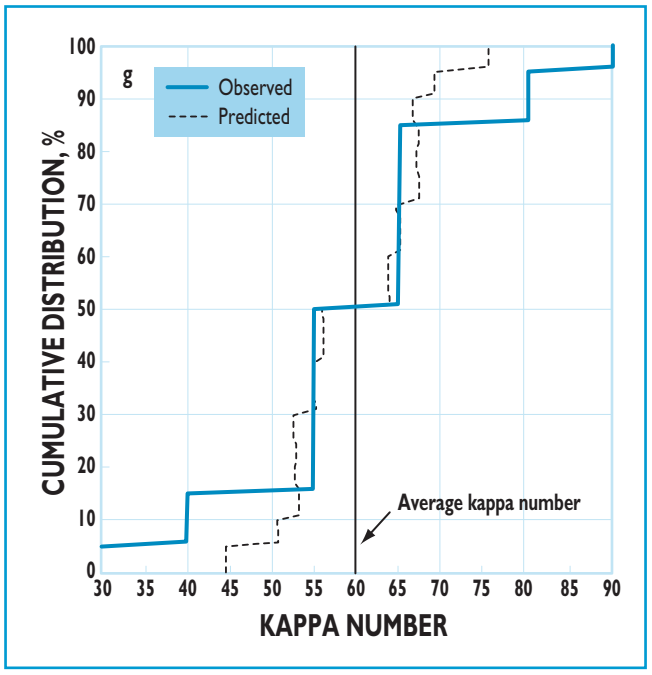
MODELS

Five different multivariate calibration models were calculated using PLS regression. For all models, the second derivative of the NIR spectra was used as predictor matrix X.

- Model 1 predicts the average kappa number in the pulp samples (kappa number range 30–90).
- Model 2 predicts the standard



4. Observed and predicted kappa number distributions for six of seven pulps with average kappa no. 60 (Model 3)



4. Observed and predicted kappa number distribution for the seventh of seven pulps with average kappa no. 60 (Model 3)

deviation around the average kappa number.

- Model 3 predicts the kappa number distribution (inhomogeneity) in blended pulps with the same average kappa number (kappa no. 60). In this case, the *Y*-matrix describes the distribution within each pulp sample by a row vector consisting of 100 elements.
- Model 4, an extension of Model 3, predicts the kappa number distribution in blended pulps with varying kappa number. As in Model 3, the *Y*-matrix describes the distribution within each pulp sample by a row vector consisting of 100 elements.
- Model 5 predicts the average kappa number and standard deviation for a "total" model that includes all the existing variation.

RESULTS AND DISCUSSION

Model 1: Prediction of average kappa number

The first step in the multivariate modeling was to calculate a PLS regression model for predicting the average kappa number from the NIR spectra of the pure pulps. The calculated PLS model gave two significant components describing 44.2% of the variation in X (r^2X = 0.442) and 98.6% in Y (r^2Y = 0.986). The predictive ability of the model (according to cross validation) was 98.2% (Q^2_{cv} = 0.982). A test set consisting of 36 samples excluded from the calibration model gave an RMSEP value of 1.68.

Figure 2 plots the observed vs. the predicted kappa numbers, showing that the model provides good predictions throughout the whole experimental range. This result is not new or surprising, since previous work has shown that the average pulp kappa number can be detected by using NIR spectroscopy in combination with multivariate calibration methods such as PLS regression *(7)*. The point in calculating this model was simply to demonstrate that the relationship existed in this specific case. The fact that it did is further proof that NIR spectroscopy is well suited for measuring kappa number, is as reliable as the traditional titration method, and is much faster.

Model 2: prediction of standard deviation

Standard deviation, in combination with the average, is a common way of summarizing a distribution. Model 2 was developed to predict the standard deviation around the average kappa number. The result was a two-component model describing 31% of the variation in X (r^2X = 0.31) and 93.2% of the variation in Y (r^2Y = 0.932). The model's cross-validated predictability of 76.4% (Q^2_{cv} = 0.764) implies that the model can predict the standard deviation for pulp kappa number. This was verified by using an external test set for validation. The calculated RMSEP value of 2.83 for the test set clearly demonstrates that the standard deviation can be predicted from the NIR spectra of the pulp samples. The results in **Fig. 3** indicate that the model provides good predictions throughout the whole experimental range.

Model 3: Prediction of kappa distribution for blended pulps with same kappa number

To investigate the possibility of extracting more information about pulp kappa number from the NIR spectra, a PLS model was calculated using the NIR data as *X*-matrix and the 100 elements (describing the kappa number distribution for each pulp) as *Y*-matrix. The idea was to find a method for detecting inhomogeneity in the pulp kappa number.

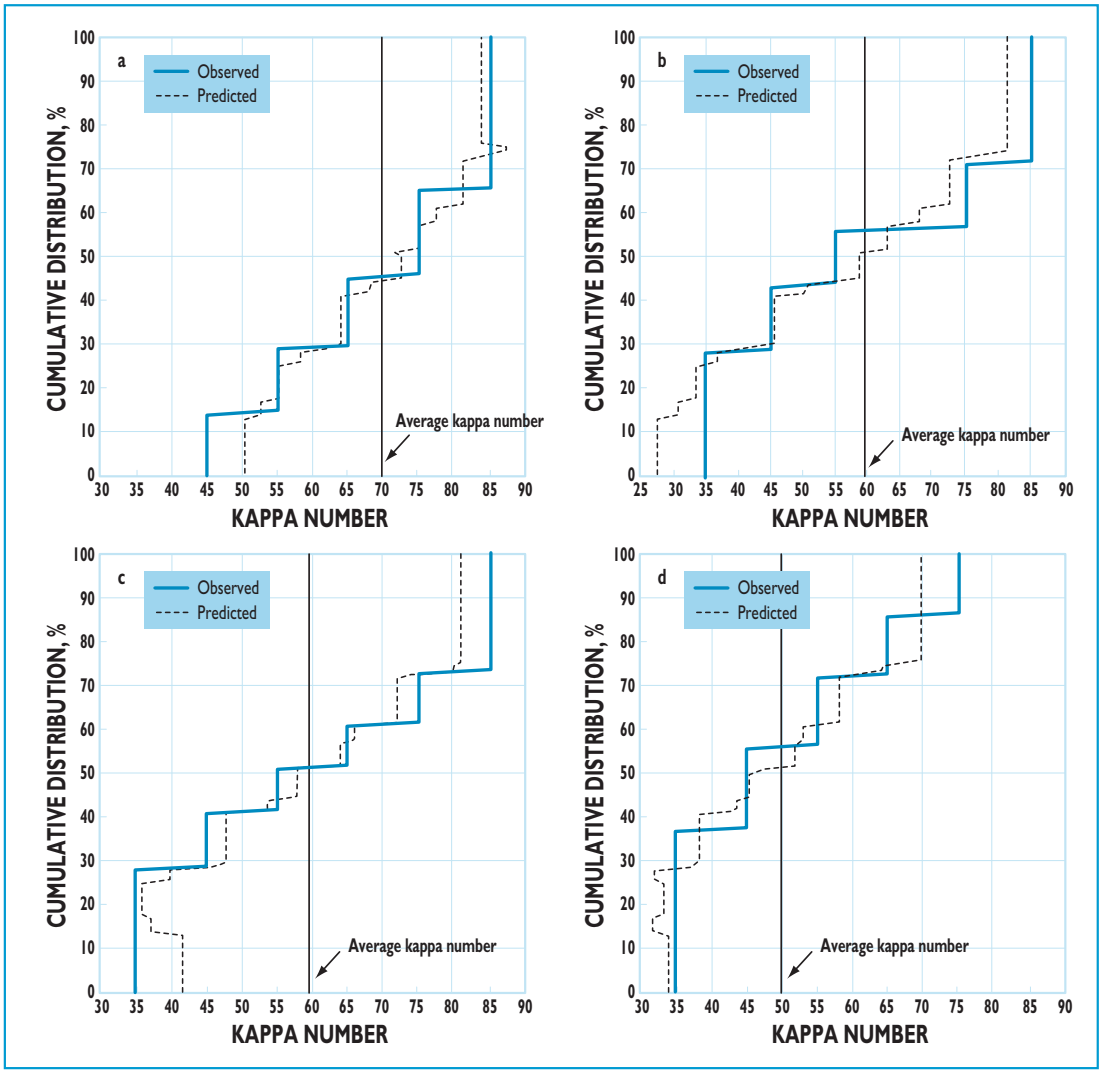
The first efforts were focused on nonuniform pulps with an average kappa no. 60. By keeping the average kappa number constant, we sought to demonstrate that NIR spectroscopy includes more information than just the average kappa number obtained by conventional titration. The calculated PLS model gave eight significant components describing 70.5% of the variation in the *X*-data (r^2X = 0.705), and 90.9% of the variation in Y (r^2Y = 0.909). Cross validation indicated that the model's predictive ability was 71.7% (Q^2_{cv} = 0.717). This result implies that the method can provide information about kappa number distributions within the pulp samples.

The calculated calibration model was validated by predicting the results for a test set consisting of seven samples that were not among those used in the original calculation of the model. **Figure 4** plots the measured values and the predicted val-

ues for all seven samples. The results in Figs. 4a–4g show that NIR data were generally able to predict trends in kappa number distribution within the seven pulps.

Model 4: Prediction of kappa distribution for blended pulps with varying kappa number

Model 3 showed that NIR data can be used to predict the average kappa number as well as the kappa number distribution for pulps with a constant average kappa number. Extending the model to include pulps over a range of average kappa number values provides a more realistic simulation of application in an industrial setting.



5. Observed and predicted kappa number distributions for four pulps with average kappa nos. 50–70 (Model 4)

A D-optimal statistical design *(12, 13)* was used to obtain a calibration set spanning the entire experimental region. A test set was also selected from the remaining samples. The calculated PLS model gave eight significant components describing 77.6% of the variation in X ($r^2X = 0.776$) and 95.5% of the variance in Y ($r^2Y = 0.955$). Cross validation indicated that the model's predictive ability was 85.1% ($Q^2_{cv} = 0.851$).

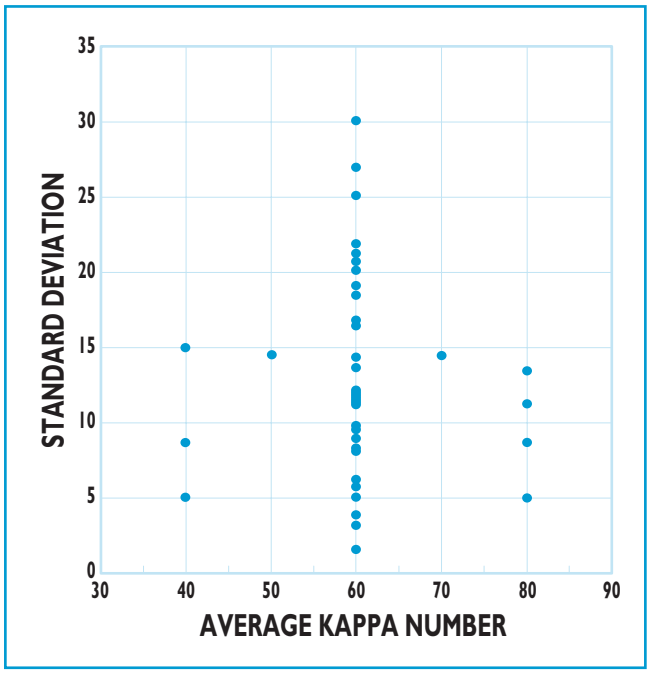
Observed and predicted results for the four samples in the test set are illustrated in **Fig. 5**. The results in Figs. 5a–5d show that the NIR data contain information on both the average kappa number and the kappa number distribution within each pulp sample.

Model 5: prediction of kappa number and standard deviation for a "total" model

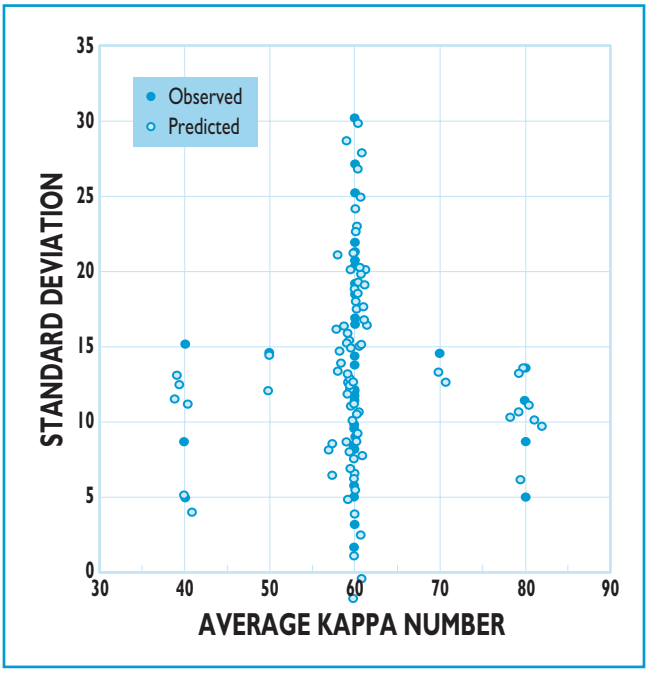
Calculation of a PLS model that predicted the two variables of interest—average kappa number and standard deviation—gave five significant components with a cross-validated predictive ability of 80.8% ($Q^2_{cv} = 0.808$). **Figure 6**, which plots the average kappa number vs. the standard deviation for the "total" model, provides a good visualization of the whole experimental design.

Figure 7 plots the predictions from the PLS model along with the results presented in Fig. 6. The combined results clearly demonstrate that the model's predictions are comparable with the observed values. The RMSEP values for the test-set objects were 0.84 for kappa number and 3.06 for the standard deviation.

Since the kappa number is easy to predict with great accuracy, the source of error in the models must be the standard deviation. One problem with predicting the standard deviation is that all of the models are based on the assumption that each pure pulp has zero variation around the average value. This assumption is probably not correct, since it is impossible to find a raw material so uniform that each wood fiber would receive precisely the same treatment during the cooking process. Unfor-



6. Standard deviation vs. average kappa number for the "total" model describing the distribution design



7. Observed and predicted standard deviation vs. average kappa number for the "total" model describing the distribution design

tunately, there is no way to determine the true standard deviations in the pure pulps.

CONCLUSIONS

The combination of near-infrared reflectance (NIR) spectroscopy and multivariate calibration methods proved to be an efficient tool for detecting nonuniformity in laboratory-cooked kraft pulps. This was demonstrated by testing pulp samples with a known nonuniform kappa number distribution. The results suggest that the proposed methodology could be used to monitor and control pulp inhomogeneity during the cooking process, leading to the production of more uniform, higher-quality pulps. **TJ**

Antti, Ph.D., Alexandersson, M.Ch.Eng., and Sjöström, professor, are affiliated with the Research Group for Chemometrics, Department of Organic Chemistry, Umeå University, S-901 87 Umeå, Sweden. Wallbäcks is a Ph.D. at AssiDomän Corporate Research, AssiDomän Tech Center, S-941 86 Piteå, Sweden.

Financial support from AssiDomän Corporate Research, Akzo Nobel Surface Chemistry, and the Center for Forest and Biotechnology and Chemistry is gratefully acknowledged.

Received for review Sept. 3, 1998.

Revised June 19, 1999.

Accepted July 3, 1999.

LITERATURE CITED

1. Gullichsen, J., Kolehmainen, H., and Sundqvist, H., *Paperi Puu* 74(6): 486(1992).
2. Agarwal, N., Gustafson, R., and Arasakesari, S., *Paperi Puu* 76(6-7): 410(1994).
3. Krässig, H. A., *Cellulose—Structure, Accessibility and Reactivity*, 2nd edn., Gordon and Breach Science Publishers, 1996.
4. Sjöström, E., *Wood Chemistry—Fundamentals and Applications*, 2nd edn., Academic Press, New York, 1993.
5. Bonanno, A. S., Olinger, J. M., and Griffiths, P. R., *Near Infra-red Spectroscopy—Bridging the Gap between Data Analysis and NIR Applications* (K. I. Hildrum, T. Isaksson, T. Naes, *et al.*, Eds.), Ellis Horwood, Chichester, UK, 1992, pp. 19–28.
6. K. I. Hildrum, T. Isaksson, T. Naes, *et al.*, Eds., *Near Infra-red Spectroscopy, Bridging the Gap between Data Analysis and NIR Applications*, Ellis Horwood, Chichester, UK, 1992, pp. 377–434.
7. Birkett, M. D. and Gambio, M. J. T., *Tappi J.* 72(9): 193(1989).
8. Schultz, T. P. and Burns, D. A., *Tappi J.* 73(5): 209(1990).
9. Wright, J. A., Birkett, M. D., and Gambio, M. J. T., *Tappi J.* 73(8): 164(1990).
10. Antti, H., Sjöström, M., and Wallbäcks, L., *J. Chemometrics* 10: 591(1996).
11. SCAN C 1:77, "Kappa Number," Scandinavian Pulp Paper and Board Testing Committee, Stockholm.
12. Cornell, J. A., *Experiments with Mixtures*, 2nd edn., Wiley, New York, 1992.
13. Kettaneh-Wold, N., *Chemometrics and Intelligent Laboratory Systems* 26(14): 57(1992).
14. Martens H. and Naes T., *Multivariate Calibration*, Wiley & Sons, New York, 1989.
15. Nortvedt, R., Brakstad, F., Kvalheim, O. M., et al., (Eds.), *Anvendelse av Kjemometri innen forskning og industri*, Tidsskriftforlaget Kjemi AS, 1996, pp. 169–177.
16. Höskuldsson, A., *J. Chemometrics* 7: 45(1993).
17. Wold, S., Johansson, E., and Cocchi, M., "PLS—Partial Least Squares projections to latent structures," in *3D-QSAR in Drug Design Theory Methods and Applications* (H. Kubinyi, Ed.), Escom Science Publishers.
18. Stone, M., *J. Royal Statist. Soc.* B36: 111(1974).
19. Wold, S., *Technometrics* 20: 4(1978).