Using near-infrared multivariate image regression to predict pulp properties

MANISH H. BHARATI, JOHN F. MacGREGOR, AND MARC CHAMPAGNE

ABSTRACT: Traditional laboratory testing of dissolving pulp is labor intensive and time consuming. Multivariate image regression (MVIR) offers a testing method that is much more rapid than the analytical chemistry approach. Pulp properties are predicted from multi-spectral, near-infrared (NIR) images of finished pulp. A single test sample can be used to predict four indicator variables for the quality of pulp—S10, S18, DCM resin, and intrinsic viscosity. The testing approach provides a framework for studying pulp heterogeneity through deriving spatial pulp property distribution across the imaged section of a pulp sample. Preliminary test results have been promising for both off-line and at-line studies.

Application: Product quality can be monitored efficiently by modeling data from NIR multi-spectral images of finished pulp.

Dissolving pulp is produced in various grades to be used in manufacturing several specialty products like rayon, pharmaceuticals, and photographic films. A number of quality control tests are performed on the finished, dried pulp product in analytical laboratories to monitor pulp quality for every grade manufactured.

Some of these laboratory tests involve complex and lengthy wet chemistry procedures. Furthermore, most tests are naturally destructive to the solid pulp samples since they entail dissolving the sample in alkaline or acidic solutions. As a result, multiple samples are required for a complete analysis of pulp quality.

A novel procedure for testing the finished product is based on multivariate image analysis (MIA) of near-infrared (NIR), multi-spectral images of pulp samples. Information extracted from the images is then statistically modeled to predict relevant pulp properties.

NIR SPECTROSCOPY

Traditional probe detectors

NIR spectroscopy measures the response of materials to light in the wavelength range of 700–2500 nm. The light is absorbed and converted to energy mainly in materials with characteristic functional groups such as C–H, O–H, N–H, and C=O. NIR is very sensitive to moisture, which has a signal that can sometimes be strong enough to hide the signals of other functional groups.

Traditional NIR spectrometers consist of a highly sensitive fiber-optic probe detector that provides precise NIR spectra by averaging over a small area of the sample being tested. This form of NIR spectroscopy has been extensively used to characterize samples in the pharmaceutical, food, and chemical industries for many years [1]. The pulp and paper industry has recently discovered its potential as a means to achieve the goals of rapid quality testing in a nondestructive manner [2, 3]. In the pulp and paper industry, NIR spectroscopy is used to predict quality indicators such as kappa number [4], pulp yield, and cellulose content [5–8]. NIR spectroscopy has been well tested and thoroughly documented [9, 10].

NIR imaging spectroscopy

A shortcoming of probe-based NIR spectrometers is that they cannot provide simultaneous multiple-point readings across a solid sample. NIR imaging spectrometers were introduced to compensate for this shortcoming. These imaging spectrometers work on the same principles as their traditional counterparts. The difference lies in the detector, which in our case is a digital camera with a light-diffracting spectrograph attached. The camera captures reflected light from the sample as an intensity image at unique NIR wavelengths. The scanned section of a sample is recorded as a three-dimensional, multi-spectral, digital image. In such images, the digitized NIR spectrum is defined in the third dimension (wavelength) for each pixel location across the sample.

The NIR imaging spectrometer consisted of a light-diffracting prism–grating–prism (PGP) spectrograph attached between the front objective lens and the back of an NIR camera. The unit was sensitive to light in the spectral range of 933–1787 nm. The back of the NIR camera consisted of an indium–gallium–arsenide (InGaAs) charge coupled device array capable of acquiring intensity images with a resolution of 128 × 128 pixels. Consisting of a narrow horizontal entrance slit, the spectrograph dispersed the entering line of light into 128 horizontal lines spanning the spectral range of 933–1787 nm.

The NIR camera recorded the 128 dispersed lines as a two-dimensional intensity image, where the horizontal axis (x) was the spatial dimension and the vertical axis was the spectral (l) dimension of the sample being imaged [11]. For this three-dimensional image, we obtained the second spacial dimension (y) by capturing multiple lines across a moving sample under the spectrograph–camera assembly. This setup is ideal for samples that are produced on moving webs, such as pulp and paper samples.

MULTIVARIATE IMAGE REGRESSION OF PULP NIR IMAGES

Overview of MIA

Data from imaging spectrometers can be viewed as *multivariate* images, where the NIR spectrum represents the variable dimension. A multivariate image is defined as a stack of congruent images [12], where each image is measured for a different wavelength, frequency, or energy.

Figure 1 shows an example of a 512 × 512-pixel multivariate image acquired at four different wavelengths arranged as a 512 × 512 × 4 cube ($\underline{\mathbf{X}}$). At each pixel location in $\underline{\mathbf{X}}$, there is high correlation in

pixel intensities because each point of the captured scene is defined by multiple pixels in the wavelength dimension.

Multivariate images usually contain an enormous amount of data (*e.g.*, $\underline{\mathbf{X}}$ contains 512 × 512 × 4 = 1,048,576 pixel intensities), making the computational analysis intensive. Multivariate statistical methods like multi-way principal component analysis (MPCA) have proven to be ideal tools for efficiently reducing data and analyzing multivariate images in a reduced dimensional latent variable subspace [12]. Figure 1 illustrates the MPCA decomposition of the three-dimensional image array $\underline{\mathbf{X}}$ (upon unfolding to a two-dimensional matrix $\mathbf{X}$) into a linear combination of two scores ($\mathbf{t}_1$, $\mathbf{t}_2$) and two loadings ($\mathbf{p}_1$, $\mathbf{p}_2$) plus noise ($\mathbf{E}$). Details about MPCA decomposition and MIA of $\underline{\mathbf{X}}$ are provided elsewhere [12, 13].

MIA for feature extraction of NIR multi-spectral images

Using NIR imaging spectrometry, we captured multi-spectral images of 60 dissolving pulp samples from two main pulp grades. The dissolving pulp grades were based on a bleached softwood mixture of pine and spruce. The pulp was made by the sulfite method. For confidentiality reasons, the pulp property values were scaled to a relative index to preserve the raw values, but the trends of the properties were retained in the results presented.

A pulp surface area of 93 mm (W) × 140 mm (L) was captured as intensity images measuring 102 pixels (W) × 448 pixels (L) in 108 unique wavelengths spanning the 933–1650-nm NIR range. Thus, for purposes of MIA the dimensions of each multivariate image were 102 × 448 × 108 pixels. We decomposed each sample image using MPCA to produce a one-dimensional MIA score and loading space.

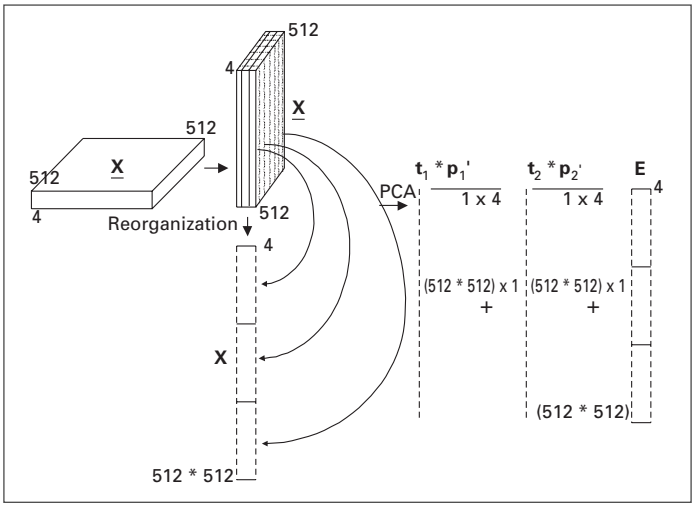
The MIA scores can be used if one is interested in the spatial variations of feature pixels throughout the imaged area of the pulp sample. However, our main aim was to derive *overall* quality data from NIR multi-spectral images of the pulp samples. We determined the pulp quality indicators through wet-chemistry laboratory testing.

The MIA loadings indicate the NIR spectral variations of pixels throughout the scanned area of the pulp sample. Thus, the loading vectors provide overall information about chemical features (related to signatures of various functional groups) in the imaged pulp sample. The first MIA loading vector $\mathbf{p}_1$ of each pulp sample image was used as its feature vector, which we analyzed further to infer pulp quality. No data scaling was applied to the multivariate images prior to MPCA decomposition for MIA [12].

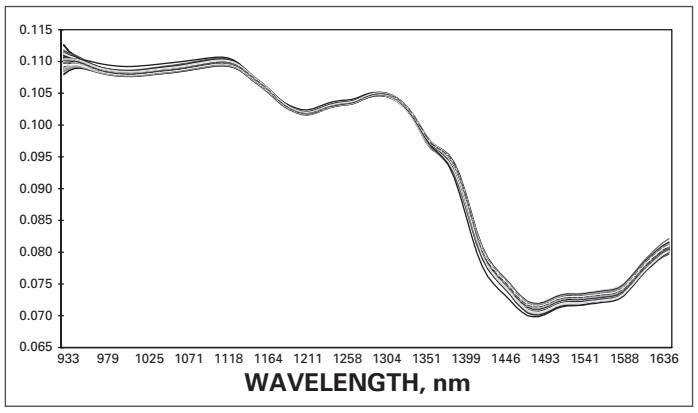
Hence, $\mathbf{p}_1$ was a normalized mean NIR spectrum of the 45,696 unique NIR spectra over the scanned area of the pulp sample (102 × 448 pixels). This feature vector is similar to a point reading of a traditional NIR probe. In both cases, the feature vector is an averaged NIR spectrum. The main difference between the two techniques lies in the spectral averaging. As opposed to point averaging in NIR probe instruments, the MIA loading vectors are averaged spectra from multiple points across the imaged area of the sample.

When MPCA is performed on the NIR multi-spectral images, weights of the $\mathbf{p}_1$ loading vector can be plotted against their respective variables (wavelengths). **Figure 2** shows the $\mathbf{p}_1$ feature vectors of the 60 spectral images of the 60 pulp samples. This new feature vector set represents a feature space ($\mathbf{X}_{feature}$), which can be used as regressor vectors to build predictive models for pulp quality.

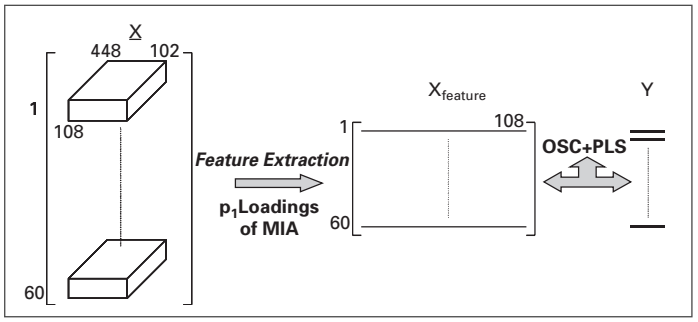
The $\mathbf{p}_1$ feature vectors in the figure indicate that the higher wavelengths (1400–1650 nm) generally exhibit low NIR reflectance, or high absorbance, whereas the lower wavelengths (933–1300 nm) exhibit high reflectance. The regions of highest varaition among the feature vectors are in the wavelength ranges from 933 to 1100 nm and from 1400 to 1650 nm.



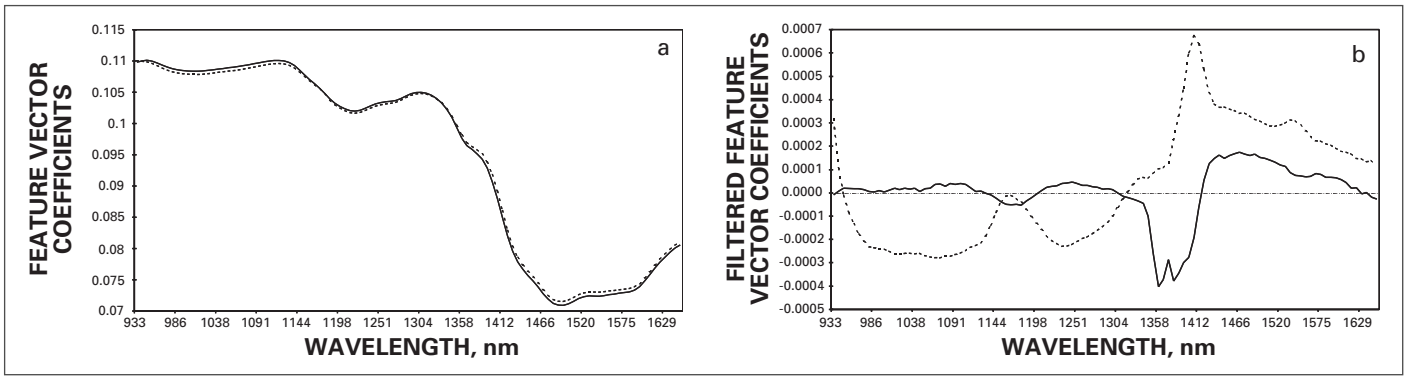
1. MPCA decomposition of a 512x512x4 multivariate image into a reduced dimensional latent variable subspace.



2. Weights of MIA p_1 loading vectors of 60 NIR multi-spectral images of pulp.



3. Scheme of the MVIR strategy for predicting laboratory data (Y) from NIR multi-spectral pulp images ($\underline{X}$).



4. Effect of OSC filtering on pulp image feature vectors. (a) Raw feature vectors of two sample images. (b) Filtered feature vectors upon removing two OSC components (Y = intrinsic viscosity).

MVIR modeling of MIA feature space to predict pulp quality

The feature vectors extracted from NIR multi-spectral pulp images can be further analyzed to predict overall pulp quality data. We performed this analysis via partial least squares (PLS) regression modeling between the feature space and corresponding pulp quality data from laboratory testing. PLS modeling has been popular in analytical chemistry for multivariate calibration of NIR spectra. Once trained, these models can be used to predict analyte concentrations from NIR spectra of new samples [14, 15].

The overall MVIR scheme is illustrated in **Fig. 3**. The MIA feature vectors ($\mathbf{X}_{feature}$) were individually regressed with four pulp quality variables ($\mathbf{Y}$ = S10, S18, DCM resin, and intrinsic viscosity). For dissolving pulp grades, these four parameters are the key values for customers who manufacture rayon, pharmaceuticals, photographic films, etc. These variables describe the solubility of hemicelluloses (S10 and S18), the measure of extractable materials using dichloromethane solvent (DCM resin), and the average length of the molecular chains of polymers making up pulp fibers (intrinsic viscosity).

Tembec Inc. previously performed a preliminary analysis to empirically predict pulp properties from NIR spectra obtained through a spectroscope with a probe detector [16]. PLS regression modeling was used to predict pulp properties from NIR spectra. The NIR spectra were orthogonal signal corrected (OSC) prior to PLS modeling. We also used OSC of the MIA feature vectors prior to PLS modeling for predicting pulp properties.

OSC is a spectral filtering technique [17, 18] for removing systematic variations from MIA feature vectors ($\mathbf{X}_{feature}$) that are unrelated to (*i.e.* orthogonal to) pulp properties ($\mathbf{Y}$). This type of filtering is particularly important because the NIR spectra are strongly affected by moisture in the pulp. Other systematic variations in $\mathbf{X}_{feature}$ that are unrelated to $\mathbf{Y}$ include surface effect variations in images of different samples and light variations between imaging runs. By subtracting the OSC components from $\mathbf{X}_{feature}$, we obtain a filtered predictor matrix, which contains the variations of interest that are related to the variations in $\mathbf{Y}$. (An excellent introduction to OSC of NIR spectra is offered by Wold *et al*. [18].)

We found it optimal to remove two OSC components, based on the predictive ability of the filtered model. The amount of information removed by OSC can be more fully appreciated by observing the raw and filtered feature vectors of the pulp NIR images.

The graph in **Fig. 4a** illustrates feature vectors extracted via MIA for the NIR multi-spectral images of two pulp samples from the "at-line" imaging study. **Figure 4b** illustrates the signal-corrected feature vectors after filtering with two OSC components using $\mathbf{Y}$ = intrinsic viscosity.

By comparing the magnitudes of the coefficients in the two feature vectors before and after filtering, we see that OSC removes most of the variation in $\mathbf{X}_{feature}$. Furthermore, OSC removes the trends in the feature vector coefficients with respect to the NIR wavelength spectrum. Subtle differences between the feature vectors are enhanced in certain wavelengths, indicating higher information content in specific regions of the NIR spectrum. Such information is related primarily to the intrinsic viscosity variations in the dataset.

To develop the empirical models, we then individually PLS regressed the OSC-filtered feature space with respect to the four different pulp properties with the respective laboratory-measured pulp property data

Regression modeling details and results

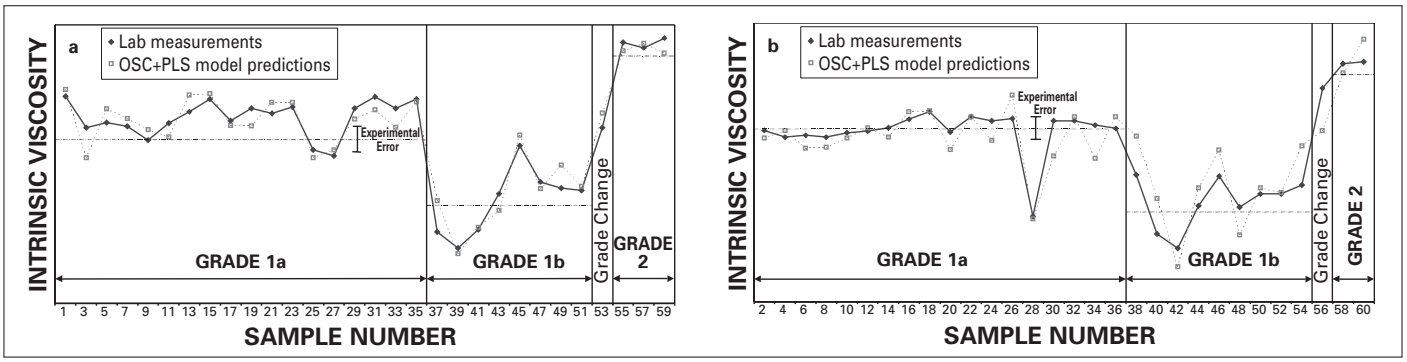
In an "at-line" imaging study conducted at the pulp mill, samples from the end of each roll were imaged immediately after production and prior to laboratory testing for pulp quality. The following is a description of the regression modeling and results from this "at-line" imaging study.

Sixty pulp samples were imaged in a time sequence over three days at the pulp mill. These samples belonged to two main grades, Grades 1 and 2. Grade 1 was further divided into two sub-grades, A and B, because of different specifications on some of the pulp properties. Laboratory tests for the four pulp properties were performed on the pulp samples after imaging. The NIR multi-spectral images of the pulp samples were decomposed via MPCA into $\mathbf{p}_1$ loading feature vectors and were assembled as observations (rows) of $\mathbf{X}_{feature}$ (Fig. 3).

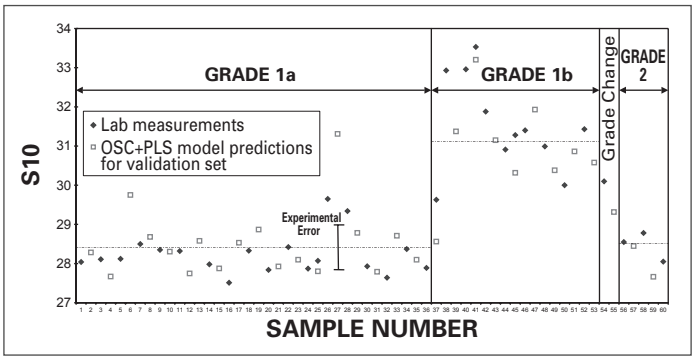
Two equal divisions of the complete dataset were made (30 samples each) for (a) training the multivariate regression model to fit the $\mathbf{Y}$ data (Observations 1, 3,…, 49 of $\mathbf{X}_{feature}$) and (b) using the developed model on a validation dataset (Observations 2, 4,…, 50) for predicting the pulp properties and calculating prediction errors. Alternate samples were divided into training and validation sets mainly to accommodate the analytical laboratory. The laboratory procedure for measuring intrinsic viscosity required only 20 min. As a result, this property was the only pulp property that was meas-

PULP PROPERTY	PLS COMPONENTS	R^2_{cum}	Q^2_{cum}	RMSEE	RMSEP	EXPERIMENT ERROR, 2 std. dev.
Intrinsic viscosity	3	0.929	0.894	0.289	0.451	0.223
S10	1	0.891	0.813	0.545	—	0.600
S18	2	0.974	0.964	0.068	—	0.300
ln(DCM resin)	2	0.76	0.715	0.596	—	0.126

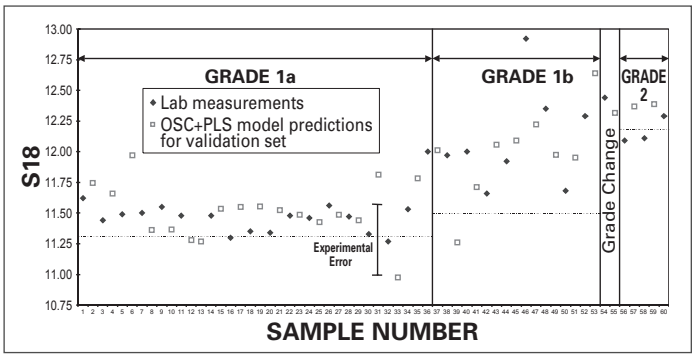
I. Multivariate regression model results for four pulp property variables.



5. Time series plot of intrinsic viscosity experimental data and OSC + PLS regression model (a) fit to the training dataset and (b) prediction of the validation dataset.



6. Time series plot of S10 pulp property experimental data and OSC + PLS regression model predictions.

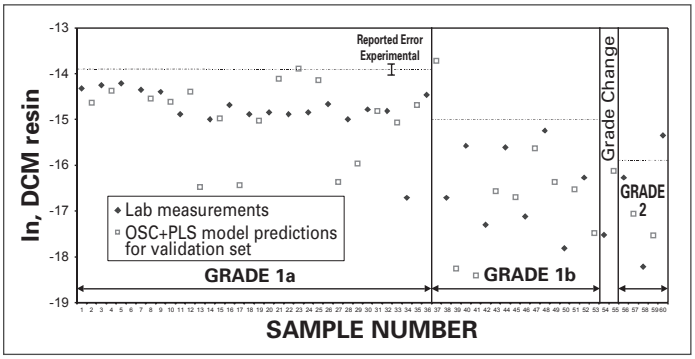


7. Time series plot of S18 pulp property experimental data and OSC + PLS regression model predictions.

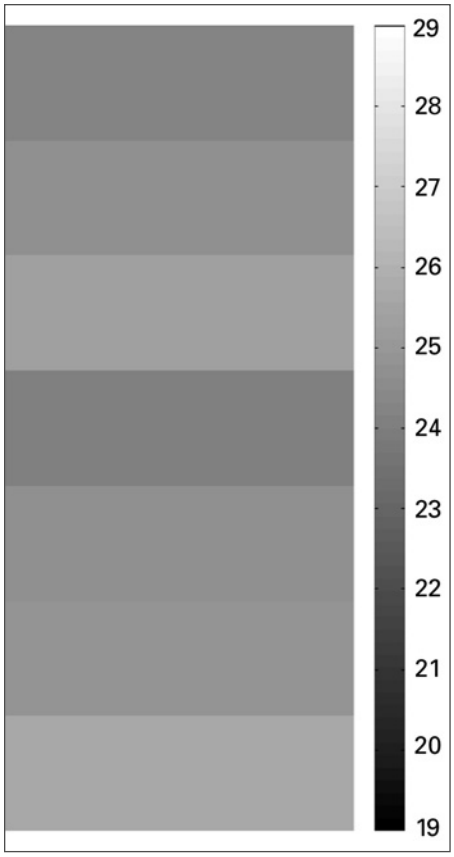
ured for all 60 pulp samples from the at-line study. The other three pulp properties (S10, S18, and DCM resin) were measured for only half of the dataset (*i.e.* only for the training dataset) since their wet-chemistry procedures were more tedious and took much longer. Because three property measurements were not available in the validation set, PLS model prediction errors are not available for S10, S18, and DCM resin.

The NIR spectral data in $\mathbf{X}_{feature}$ were mean-centered with respect to the 108 wavelengths (columns) prior to modeling. Each of the four pulp property variables in $\mathbf{Y}$ was mean-centered and auto-scaled to unit variance. DCM resin values were transformed by taking the natural logarithm, expressed as ln(DCM resin). Beebe *et al.* [19] and Eriksson *et al.* [15] provide insights to different types of variable scaling and variable transformations that can be performed to remove nonlinearity and improve the PLS model fit and predictions.

When the DCM resin values were transferred into ln(DCM resin), the corresponding relative experimental errors (coefficient of variation = ±6.3%) were also stabilized to absolute errors (2s experimental error = ±0.13). Separate PLS regression models were developed for each of the four pulp properties in the training dataset. The corresponding models were then used to predict pulp property values from feature vectors in the validation dataset. The errors between pulp property values from the analytical laboratory tests and PLS model predictions from the training and validation datasets (for intrinsic viscosity) were then calculated.



8. Time series plot of ln(DCM Resin) pulp property experimental data and OSC + PLS regression model predictions.



9. Intrinsic viscosity distribution across seven subsections of NIR multi-spectral image of Sample 16 from the at-line imaging study.

An overall measure of the quality of a regression model can be determined by comparing its root mean squared error of fit (RMSEE) and its root mean squared error of prediction (RMSEP) with the experimental errors from laboratory tests of pulp properties. If the RMSEE and RMSEP are similar to experimental errors, the regression models can fit and predict the pulp property at least as well as the laboratory testing can. The results of four OSC + PLS multivariate regression models used for the pulp properties of interest are summarized in **Table I**.

In Table I, R^2_{cum} is the cumulative sum of squares of **Y** explained by the fitted PLS regression model. Q^2_{cum} is the percent sum of squares that can be predicted as determined by cross-validation [20]. If Q^2_{cum} is close to R^2_{cum} and both are close to 1.0, then the model is considered to be highly predictive. RMSEE is defined as the root mean squared error of estimation for the fitted data using the PLS model:

$$RMSEE = \sqrt{\sum (\hat{y} - y)^2 / N} \qquad (1)$$

where $\hat{y}$ and y are the predicted and observed values of the response variable, respectively, and N is the number of fitted data points. RMSEP is calculated in the same manner as RMSEE, except that it is defined as the root mean squared error of prediction for the data *not* used in the model building stage (*i.e.*, on the validation dataset) [21].

Figure 5 shows the time series plots of the training and validation datasets of the intrinsic viscosity pulp property collected during the "at-line" study. As these plots show, the predictions closely follow the trends of the laboratory test data from Grade 1A to Grade 2. The error bar around the "target value" (dash-dot line) gives the 2s experimental error limit for the laboratory measurements of intrinsic viscosity. The model fit and predictions seem generally within the limits defined by the error bar. The OSC + PLS model predictions not only follow the general trends of intrinsic viscosity between grades, but the model can also detect any process upsets within a grade (*e.g.*, the temporary decrease in intrinsic viscosity around Sample 28 in Grade 1A).

Figure 6 illustrates a time series plot of the S10 pulp property. PLS model predictions for the validation set have been overlaid with the training set laboratory measurements as part of a *common* time series plot. Actual laboratory measurements for the validation set are not available for the S10 pulp property. As the figure shows, the OSC + PLS model performs reasonably well in following the S10 pulp property trends from Grade 1A to Grade 1B to Grade 2. Furthermore, this model can also detect the sudden increase in the S10 value for Grade 1A around Samples 27 and 28. Thus the model performs well both within and between the grades studied. The experimental error bar (±0.6) is included to show the model's ability to predict the pulp property values for the validation set.

Similar time series plots are shown in **Figs. 7 and 8** for S18 and ln(DCM resin) pulp properties, respectively. The chosen OSC + PLS regression model for S18 also seems to perform reasonably well in following the pulp property trends from Grade 1A to Grade 1B to Grade 2.

As Fig. 8 shows, the PLS model predictions for ln(DCM resin) are not as good as those for the other three pulp properties. This discrepancy could be attributed to the noisy laboratory measurements of the DCM resin pulp property for Grade 1B and Grade 2. As illustrated in the figure, the laboratory measurements (solid diamonds) for these grades exhibit a bouncing pattern between –15 and –18 for Samples 37–60. These inaccuracies may contribute to a bad training set for the PLS models, which is reflected in the poor predictions of ln(DCM resin) for the validation set.

EXTRACTING SPATIAL DISTRIBUTIONS USING MVIR MODELS

The main objective of the MVIR pulp property modeling study was to relate multi-spectral NIR images of finished pulp samples to their average properties. We did not use any spatial information from the pulp sample images in the MVIR calibration models. Instead, we tried to extract spatial information of a pulp sample through the application of the overall MVIR model on subsections of its multi-spectral NIR image [22]. Doing so would enable us predict the spatial variations in pulp properties throughout the sheet and would therefore provide a measure of the heterogeneity of the sample with respect to each property. This advantage is a potentially important one for the two-dimensional NIR imaging spectrometer over a traditional NIR instrument, which uses a single point as a source.

When the multi-spectral NIR image is segmented into various subsections and the overall MVIR model is applied to predict local pulp properties for each subsection, one can obtain a spatial distribution of predicted pulp properties throughout a pulp sample. Such a distribution can be plotted as a two-dimensional histogram of property predictions. The histogram gives an idea of the mean and variance of a pulp property as it is distributed through a particular sample image. The mean property prediction gives an overall measure (similar to the overall MVIR model predictions), whereas the variance provides a measure of pulp heterogeneity with respect to the property being predicted.

Since the NIR imaging spectrometer scans the pulp sample as a line-scan camera, if one were to segment the pulp image into subsections containing several line scans, the corresponding prediction histogram could be used to monitor changes in pulp properties in an on-line fashion. In a typical monitoring scheme, one could track both the mean and the variance of the pulp properties.

The idea is illustrated by calculating the intrinsic viscosity distribution across Sample 16 from the "at-line" imaging experiment. The intrinsic viscosity value (along with 2s experimental errors) for the sample as measured in the laboratory was 25.03±0.22, whereas the MVIR prediction using the OSC + PLS regression model was 25.14. The multi-spectral NIR image of the entire pulp sample was segmented into seven subsections, with each subsection covering the full line-scan width of the original pulp sample image.

Using the overall OSC + PLS MVIR model coefficients within every image subsection, we obtained intrinsic viscosity predictions for seven subsections of the scanned surface of the pulp sample. The size of image segments chosen will be a tradeoff between obtaining finer spatial detail with finer segmentation and increased variance of the predictions in each subsection as result of fewer image pixels and local variations in lighting [22]. (The distribution of pulp property predictions across the same pulp sample image segmented into 48 subsections is illustrated elsewhere [23].)

To investigate the variation of the properties throughout the sheet, the overall OSC + PLS MVIR model can be applied to any image subsection to obtain predictions for that subsection of the image. The mean intrinsic viscosity prediction over the seven subsections is 24.87, and the standard deviation of the predictions across the seven subsections is 0.549. The physical dimension of the pulp sample area scanned by the imaging spectrometer is 140 mm (L) × 93 mm (W). Compared to the width of an industrial pulp web, the imaging spectrometer covers less than 1% of its true horizontal dimension. As a result, it would be reasonable to assume very little variability in the spatial distribution of predicted intrinsic viscosity across the seven subsections of the multi-spectral NIR pulp image.

CONCLUSIONS

Figure 9 shows the distribution of intrinsic viscosity predictions obtained by applying the PLS model coefficients to seven subsections. A multivariate image regression technique was used to predict pulp quality data from NIR multi-spectral images of pulp samples. The MVIR technique can be used to relate feature information from multivariate images of finished products to corresponding data obtained from the quality control lab. Once trained, the MVIR model can then be used to infer quality data from product images in on-line industrial processes equipped with multivariate imaging sensors.

The proposed technique has been tested in a feasibility study at a pulp mill with promising results. The empirical modeling technique produces adequate predictions of three out of the four pulp quality indicators studied. With current trends in high-speed camera technology, the proposed image regression scheme has the potential for on-line monitoring of pulp quality through NIR multi-spectral imaging spectroscopes mounted on pulp machines. **TJ**

LITERATURE CITED

1. Stark, E., and Luchter, K., *Appl. Spectrosc.* 22(4): 335(1986).
2. Wallbäcks, L., Edlund, U., Lindgren, T., and Agnemo, R., *Nordic Pulp Paper Res. J.* 2: 88(1995).
3. Antti, H., Sjöström, M., and Wallbäcks, L., *J. Chemometrics* 10: 591(1996).
4. Birkett, M. D., and Gambino, M.J., *TAPPI J.* 72: 193(1989).
5. Schultz, T.P., and Burns, D. A., *TAPPI J.* 73: 209(1990).
6. Wright, J.A., Birkett, M.D., and Gambino, M.J., *TAPPI J.* 73: 164(1990).
7. Wallbäcks, L., Edlund, U., and Norden, B., *TAPPI J.* 74: 201(1991).
8. Woitkovich, C.P., McDonough, T.J., and Malcolm, E.W., *TAPPI 1994 Pulping Conference Proceedings*, TAPPI PRESS, Atlanta, p. 721.
9. Whetsel, K., *Appl. Spectrosc.* 2(1): 1(1968).
10. *Handbook of Near-Infrared Analysis* (D.A. Burns and E.W. Ciurczak, Eds), Marcel Dekker, New York, 1992.
11. Herrala, E., and Okkonen, J., *Int. J. Pattern Recognition and Artificial Intelligence* 10(1): 43(1996).
12. Geladi, P., and Grahn, H., *Multivariate Image Analysis*, John Wiley and Sons, Chichester, UK, 1996.
13. Bharati, M.H., and MacGregor, J.F., *Ind. Eng. Chem. Res.*, 37: 4715(1998).
14. Martens, H. and Næs, T., *Multivariate Calibration*, John Wiley and Sons, Chichester, UK, 1989.
15. Eriksson, L., Johansson, E., Kettaneh-Wold, N., and Wold, S., *Multi- and Megavariate Data Analysis*, Umetrics AB, Umeå, Sweden, 2001.
16. Champagne, M., Meglen, B., Wold, S., and Kettaneh-Wold, N., *Pulp Paper Can.* 41(2001).
17. Svensson, O., Kourti, T., and MacGregor, J.F., *J. Chemometrics* 16: 176(2002).
18. Wold, S., Antti, H., Lindgren, F., and Öhman, J., *Chemometrics Intell. Lab. Syst.* 44: 175(1998).
19. Beebe, K.R., Pell, R.J., and Seasholtz, M. B., *Chemometrics: A Practical Guide*, John Wiley and Sons, New York, 1998.
20. Wold, S., *Technometrics* 20: 397(1978).
21. Umetrics, *SIMCA-P 9.0*, Umetrics AB, Box 7960, S-907 19 Umeå, Sweden, 2001.
22. Yu, H., and MacGregor, J.F., "Multivariate image analysis and regression for prediction of coating content and distribution in the production of snack food", submitted to *Chemometrics Intell. Lab. Syst.*, 2002.
23. Bharati, M.H., *Multivariate image analysis and regression for industrial process monitoring and product quality control*, Ph.D. thesis, McMaster University, Hamilton, ON, Canada, 2002.

Received: September 17, 2002
Revised: July 21, 2003
Accepted: May 10, 2003

This paper is also published on TAPPI's web site <www.tappi.org> and summarized in the May *Solutions! for People, Processes and Paper* magazine (Vol. 87 No. 5).

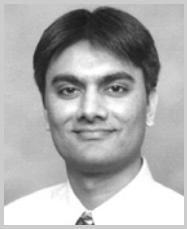
INSIGHTS FROM THE AUTHORS

For many years now, we have been doing research on process control and multivariate statistical methods for extracting information from sets of industrial data. Our main area of research is process control as it relates to sensors for monitoring quality. On-line digital imaging systems represent a powerful new kind of sensor for monitoring and controlling quality in solid products such as pulp and paper.

The tremendous potential of NIR imaging for evaluating quality was evident to us in this work because we were able to do better than we expected in determining the properties accurately. One problem we had was getting pulp samples in a timely enough manner to prevent the pulp from showing the effects of aging. The solution was obvious. We moved the camera from McMaster University to Tembec's mill site and took the measurements as the rolls came off the line.

Our next step in developing the method is to get more detailed spatial information on all the quality variables throughout the sheets, rather than just the average quality over the image. We also want to develop the imaging system for on-line use. In the meantime, the technique can give mills a new laboratory method for monitoring pulp quality that can replace many of the tedious methods now used.

Bharati and MacGregor are with the Dept. of Chemical Engineering, McMaster University, Hamilton, Ontario L8S 4L7, Canada. Champagne (photo not available) is with Tembec Inc., Temiscaming, Quebec J0Z 3R0, Canada. Email MacGregor at macgreg@mcmaster.ca.



Bharati

MacGregor