Analytical Methods

ROYAL SOCIETY OF CHEMISTRY

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Multivariate classification of pulp NIR spectra for end-product properties using discrete wavelet transform with orthogonal signal correction

N. Tavassoli,[a] **W. Tsai,**[b] **P. Bicho**[b] **and E. R. Grant**[*a]

Natural material variations uncorrelated with physical properties of fibre networks hinder the development of robust calibration models by which to predict paper properties from on-line near-infrared (NIR) spectra of production pulps. Such a simple process gauge of product quality would offer attractive advantages for optimized manufacturing. The present work explores the effectiveness of data processing strategies designed to remove uncorrelated variance from calibration models linking NIR spectra with standard measures of paper quality, including tensile, tear, burst strength, wet and dry zero span length, freeness, absorption and scattering coefficients. Post processing of spectra by discrete wavelet transform (DWT) is shown to suppress baseline and high-frequency noise, and orthogonal signal correction (OSC) substantially improves prediction accuracy by reducing the amplitude of uncorrelated (orthogonal) variations. We find that combined pretreatment by DWT and OSC yields a spectral data set that exhibits the best prediction accuracy.

1 Introduction

To secure market superiority, a pulp manufacturer must control the quality of materials in process to maximize product sheet strength. While some pulp properties can be tested at line, conventional determinations that best predict the physical properties of an end-use product require exacting measurements in a controlled laboratory environment. Such steps add cost and introduce delay that can give rise to process variability.

A need therefore exists for on-line methods to gauge pulp stream composition and morphology that can predict the structural and physical properties of the end-use paper it forms. To this end, the industry has sought to develop spectroscopic probes.[1] Chief among such methodologies is near-infrared (NIR) spectroscopy.[1–3]

NIR diffuse reflectance absorption spectroscopy has been applied with success for the analysis of lignocellulosic materials, particularly chipped wood feedstocks,[4,5] for physical properties, such as moisture content,[6–9] density,[10–12] surface roughness,[13,14] as well as levels of lignin,[5,15–17] cellulose,[17–21] hemicellulose[5,22] and other extractives.[17,22–24] Regression models based on NIR spectra of laboratory test sheets have shown promise as a means to predict conventionally mea-

sured pulp and paper properties, such as freeness, stretch and tensile strength.[25,26]

Two factors limit the full-scale application of NIR spectroscopy as an on-line process-control tool. Broad overtone absorption bands, owing to every substance in the sample, overlap to form a spectrum in which signature features are hard to discern. Variations in pulp composition unrelated to a target property modulate the spectrum, and this tends to mask determinate spectral variations.

The present work reports progress in an effort to overcome these limitations. Taking a large NIR data set of production pulps, standardized by conventional measurements of product sheet properties, we show that multi-resolution decomposition by discrete wavelet transform (DWT) effectively recasts spectra to facilitate the isolation of determinant variance. Preprocessing calibration spectra by orthogonal signal correction (OSC), we minimize uncorrelated variance, which enables the development of multivariate classification models that predict paper sheet properties from pulp NIR spectra with greatly improved accuracy.

2 Materials and methods

2.1 Physical, mechanical and optical properties of pulp samples

Industry defines the quality of a kraft pulp by measures of its strength and optical properties, as well as other properties per-

[a] *Department of Chemistry, University of British Columbia, Vancouver, BC V6T 1Z3, Canada; E-mail: edgrant@chem.ubc.ca*

[b] *Canfor Pulp Innovation, 8610 Glenlyon Parkway, Burnaby, BC V5J 0B6, Canada.*

tinent to processing and end use. These measures in part reflect characteristics of the pulp, and in part characteristics of paper sheets made and tested under specific conditions. Standard procedures call for mechanical treatment and specified degrees of refining to conform with particular categories of end use. The response of a pulp to refining under defined conditions constitutes an inherent property of the pulp subject to standard measurement.

Thus, the pulps used in this study were tested, prior to refining, for freeness according to TAPPI standard method T 227 om-99.[27] During the course of refining, we prepared seventy five test sheet samples from these production pulps in the Burnaby laboratories of Canfor Innovation, using a semi-automatic sheet maker following T 205 om-88, with reference to conditioning and testing standards, T 402 sp-03 and T 220 sp-01. We used handsheets made with unrefined pulp to test for wet and dry zero-span tensile strength (WZS and DZS, respectively) in accordance with T 231 cm-96.

We performed all other tests on sheets prepared from pulps taken at a series of freeness stages, expressed to the nearest 500 ml. These tests included tear index (T 414 om-98), tensile strength (T 494 om-01), burst strength (T 403 om-02), light scattering and absorption coefficients (T 425 om-01). Table 1 gives the range of each parameter as well as its measured uncertainty.

2.2 Collection of NIR spectra

We collected NIR spectra using a Nicolet 6700 FT-IR spectrometer (ThermoScientific) equipped with an NIR integrating sphere module and a 5 cm diameter sample cup spinner. Operated in the diffuse reflectance mode, this instrument illuminates samples with broadband near-infrared radiation from a tungsten halogen lamp, and collects interferograms using an InGaAs detector. Fourier transforms span a spectral range from 4000 cm^{-1} to 9900 cm^{-1} at a resolution of 2 cm^{-1}. Each acquisition represents the sum of 64 scans of 10 second exposure.

We cut five circular samples sized to fit the sample cup from different positions in each pulp sheet, scanning the top-side of each. Five spectra from each sample were averaged and used in subsequent analyses.

3 Multivariate analysis

3.1 Data pretreatment

NIR spectral data can benefit to a great degree from preprocessing to compensate the effects of varying baseline, remove high-frequency noise and amplify features that appear only as inflections in the raw spectrum. The present work tests several target-independent preprocessing methods, including derivatization, standard normal variant (SNV) correction, multiplicative scatter correction (MSC) and discrete wavelet transform (DWT). Each of these approaches refines the data set without reference to standardizing classification information.

We compare prediction errors following pretreatment with these methods with results obtained following orthogonal signal correction (OSC) alone, and OSC in combination with DWT. OSC is a target-directed preprocessing method that makes reference to a multivariate classification model to reduce the weight of irrelevant spectral information.

3.1.1 First-derivative transformation NIR absorption spectra convey sample composition information in the form of overlapping vibrational overtone features that vary smoothly on a scale of tens of wavenumbers. First derivative transformation provides a ready means of identifying such features, even when their appearance in the primary spectrum is subtle.[28,29] The first derivative also serves to highlight the degree to which high-frequency properties of the spectrum, unrelated to variations in the sample set, might affect its classification.[29,30]

3.1.2 Standard Normal Variant Correction Standard Normal Variate (SNV) transformation operates by row on spectra, x_i in the data matrix X, subtracting the individual mean (zeroth-order detrending) and scaling each spectrum by its standard deviation.[31]

$$\hat{x}_i = \frac{x_i - \bar{x}_i}{s_i} \tag{1}$$

When much of the amplitude fluctuation in a data set arises from noise, this pretreatment transformation can improve a model by reducing the amplitudes of its noisiest component spectra. However, if overall signal amplitudes of sample spectra vary significantly SNV can degrade calibration by introducing non-linearities.[32]

3.1.3 Multiplicative Scatter Correction Multiplicative scatter correction (MSC) removes uncorrelated background from measured spectra, X, arising from multiplicative factors, such as path-length variations, as well as offsets, owing, for example to stray light. The procedure uses an averaged spectrum, $\bar{x}_j$, formed by the mean of a selected calibration subset, and finds parameters, a_i, b_i and e_i, fitting x_i to $\bar{x}_j$ as closely as possible by least squares:

$$x_i = a_i + b_i\bar{x}_j + e_i \tag{2}$$

where e_i represents the residual spectrum, containing the chemical information in x_i. a_i defines the intercept, and b_i the slope, that yield the corrected spectrum, $x_{i,\text{MSC}}$:

$$x_{i,\text{MSC}} = \frac{x_i - a_i}{b_i} = \bar{x}_j + \frac{e_i}{b_i} \tag{3}$$

Table 1 The experimental range of measurement and observed reproducibility of different the physical and optical properties independently determined for the samples that served as calibration and validation standards for this study.

parameters	Minimum	Maximum	Mean	Repeatability	Reproducibility
Tensile (km)	2.5	4.9	3.8	5 %	10 %
Freeness (ml)	649.5	700.5	684.3	25.3 %	32.4 %
Burst (kPa m^2 g^{-1})	1.4	3.4	2.3	22 %	28 %
Dry Zero Span (km)	14.44	16.8	15.7	5 %	10 %
Wet Zero Span (km)	12.4	15.8	14.6	5 %	10 %
Tear (mN m^2 g^{-1})	19.9	30.3	26.1	4.2 %	12.5 %
SRE600	23.5	82.7	53.4	—	—
Absorption (m^2 g^{-1})	0.1669	0.2278	0.1901	—	—
Scattering (m^2 g^{-1})	32.6	39.8	35.6	—	—

This procedure presents a risk in that a_i and b_i might correlate with a target property, in which case MSC would remove chemical information from the data set.[31,33–35]

3.1.4 Discrete Wavelet Transform Discrete wavelet transform produces a multi-scale representation of a digitized signal by using a sequence of high- and low-pass cutoff filters to sort the signal in terms of the frequency with which it varies in the wavelength space of the spectrum. Filtering divides this information to resolve the signal into a set of subbands.[36–38]

Frequencies of importance, corresponding to peak widths in the wavelength- or λ-space of the original spectrum, appear with large amplitude in the DWT decomposition without loss of λ-space position information. Subsampling the result removes the unimportant information, including the slowly varying background and high-frequency noise. Errors of prediction can serve as a guide to choose a wavelet basis and constrain frequencies to best preserve classification information while removing uncorrelated variance.[36]

3.1.5 Orthogonal Signal Correction Even though multivariate classification models such as Partial Least Squares (PLS) regression serve to extract feature information from complex data sets, uncontrolled variations, irrelevant to the properties of interest can add substantially to the computational effort, reducing the accuracy and robustness of prediction. Sometimes, uncorrelated variation appears in obvious dimensions, such as low-frequency baseline oscillations, or high-frequency noise. In other cases, irrelevant variations intrinsic to the sample or the instrument occur at the bandwidth of the determinant variation.

In the latter event, preprocessing methods focused on the minimization of uninformative variation at all frequencies can serve to improve the accuracy of a multivariate calibration. Orthogonal Signal Correction (OSC) provides one approach to succeed in this respect.[39]

OSC finds features that affect the total variation in a spectral matrix, but occur in dimensions that extend in directions orthogonal to a target variance, and then removes them. Omitting these features reduces the complexity of the model and consequently improves the linearity of the relation between a target variance and an input data set. We use a piece-wise variant of the Wise orthogonal signal correction algorithm (WPOSC).[40]

Our implementation of WPOSC applies Principal Component Analysis (PCA) to a 75 percent subset of the spectral data in $\mathbf{X}$ in order to obtain an initial score vector, $\mathbf{t}$. We orthogonalize $\mathbf{t}$ with respect to the corresponding target property matrix, $\mathbf{Y}$. This orthogonalized $\mathbf{t}^*$ then serves as the first target in a subsequent cycle of supervised refinement. At each step in this cycle, we use five repetitive application of non-linear iterative partial least squares (NIPALS) regression, using random partitions of the data, 75% to calibration and 25% to validation, providing an independent error of prediction on the basis of which to select an optimum number of latent variables.

From optimized PLS regression, we find weights, $\mathbf{w}$, such that the vector of PLS scores $\mathbf{t} = \mathbf{Xw}$ conforms optimally with $\mathbf{t}^*$. We orthogonalize this $\mathbf{t}$ with $\mathbf{Y}$, and, with this as a target, return to PLS to obtain a new $\mathbf{t}$. Upon convergence, the vector of scores, $\mathbf{t}$, and loadings, $\mathbf{P}$, from this process describes the information in $\mathbf{X}$ that is orthogonal to $\mathbf{Y}$. Applying the transpose of $\mathbf{P}$ yields a feature-selected dataset, $\mathbf{X}_{OSC}$ by:

$$\mathbf{X}_{OSC} = \mathbf{X} - \mathbf{tP}' \qquad (4)$$

$\mathbf{X}_{OSC}$ then serves as spectral data input for a new cycle of OSC, generating new orthogonalized component, $\mathbf{t}_{new}$ and $\mathbf{P}_{new}$, with which to feature-select $\mathbf{X}_{new}$. From an examination of prediction errors obtained using independent validation data sets, we have determined that OSC performs best without overfitting when implemented with two such score and loading components.[41,42]

3.2 Multivariate calibration

We use the method of Partial Least Squares (PLS) regression to compare efficacy of these preprocessing methods for classifying physical properties of production pulps on the basis of NIR spectra. To build a linear regression model, PLS finds the covariance between a set of observed variables (the matrix of NIR spectra, **X**) and predicted variables (the matrix of values of a selected property, **Y**). It rotates **X** into a coordinate direction that best represents the variance in **Y**. [43,44]

This method enables modelling success in situations for which the number of observed variables determining **X** greatly exceeds the number of standardizing property measurements contained in **Y**. It also tolerates multicollinearity in **X**, as present when basing models on largely parallel spectra.

PLS reduces the dimension of data into latent structures in the **X** block according to a selection of principal components. The algorithm calculates scores of blocks to maximize the covariance between **X** and **Y**. The weight vector calculated for each PLS component gives the maximum covariance between the two blocks. [43]

Various criteria serve to describe prediction success. We refer to the correlation coefficient in the linear regression of predicted versus standard values to judge the precision of calibration. The root mean square error of prediction (RMSEP)

$$\text{RMSEP} = \left(\sum_{i=1}^{m} (\hat{\mathbf{y}}_\mathbf{i} - \mathbf{y}_\mathbf{i})/m \right)^{1/2} \qquad (5)$$

describes how well a model applied to an independent standard data set **X** predicts the corresponding values, **Y**. Here, $\hat{\mathbf{y}}_\mathbf{i}$ and $\mathbf{y}_\mathbf{i}$ are predicted and measured properties of the sample i and m is the number of samples. Examining the standard deviation about the mean RMSEP value derived from the independent validation of a large number of different calibration reflects the reliability with which a tested pretreatment method improves the accuracy of prediction.

Increasing the number of latent variables or PLS factors usually increases the correlation between known and predicted values. However, the use of too many factors causes overfitting, which degrades the generality of a model. The number of samples used to build a model, together with the number of determinant features in the spectrum combine to define an optimum number of latent variables. [43–45] To optimize the number of latent variables for the present models, we minimize RMSEP and maximize r^2 values determined for ten different combinations of training and independent validation data sets.

For the purposes of evaluating the pre-processing strategies described above, we constructed and tested sets of classification models using SIMPLS. [46] We began by building nine **Y** matrices, detailing the known pulp physical characteristics, tensile index, freeness, burst index, dry zero span, wet

zero span, tear index, SRE600, absorption and scattering. We then applied PLS to find latent components in **X** matrices of preprocessed spectra characterizing the spectrum-property covariance. For each property and preprocessing strategy tested, we used 75 percent of the available spectra for calibration and the remaining excluded 25 percent as a test set to validate the model.

PLS requires one to choose an optimal number factors minimizing prediction error without overfitting the data. Interestingly, we generally found that the optimal number of PLS factors declined after pre-processing. This reflects the fact that preprocessed spectra had less complexity, necessitating fewer PLS factors to optimally capture the covariance.

All of the calculations described above were performed on a desktop computer running MATLAB 8.0 with the PLS Toolbox (Eigenvector Research, Inc.)

4 Results

The following sections draw upon a data set formed by the near infrared spectra of 75 pulp test sheets, comparing the effects of pre-treatment by first derivative and standard normal variant (SNV) transformation, multiplicative scatter correction (MSC), discrete wavelet transform (DWT), orthogonal signal correction (OSC), and the combination of DWT with OSC. We explore the comparative utility of these pretreated data sets as multivariate predictors of various mechanical and physical properties of sheet paper, including tensile, dry and wet zero-span, tear and burst index, freeness, SRE600 and light scattering.

We have constructed these models and tested prediction accuracy by applying the method of partial least squares to independent pretreated data subsets. Here, we compare RMSEP values obtained by use of preprocessed spectral data with prediction errors found for PLS models based on the original spectra without pretreatment.

4.1 NIR spectra with target independent pretreatment

Figure 1 plots representative, mean-centred and normalized NIR spectra of all 75 pulp samples, together with this dataset pre-processed by non-targeted first-derivative transformation, SNV and MSC. In the untreated data, we see that absorption bands associated with many vibrational overtone transitions overlap to form smooth, reproducible and, on this scale, relatively undifferentiated spectra.

Less apparent in the raw NIR spectra is the presence of a high-frequency noise component. This contribution to the experimental signal plainly appears upon first-derivative transformation. No NIR spectral intensity varies on this scale of absorption frequency. Therefore, this evident variation in the
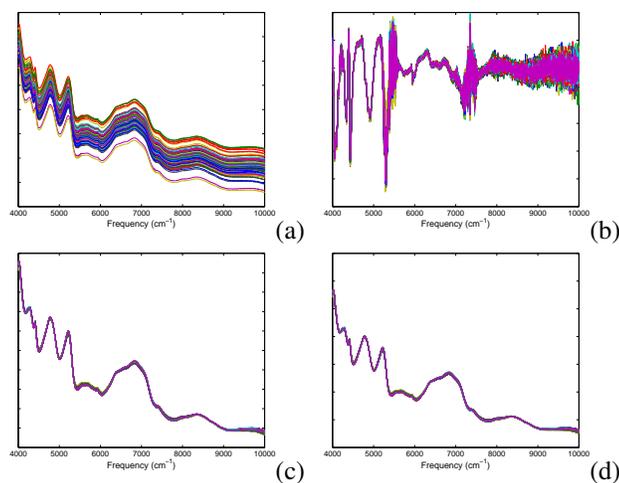
Fig. 1 (a) Representative NIR spectrum in cm^{-1} for each of the 75 pulp samples in the present dataset (c.f. Table 1) (b) Spectral data subjected to first-derivative transformation. Spectral data subjected to (c) standard normal variant (SNV) transformation and (d) multiplicative scatter correction (MSC).

Fig. 2 (a) Spectra following a discrete wavelet transform of the raw spectral data in Figure 1 (b) The same spectra following orthogonal signal correction (OSC), targeted to a prediction of the paper sheet mechanical property, tensile strength. (c) Raw spectral data pretreated first by discrete wavelet transform, as above, followed by two-component orthogonal signal correction targeted to tensile strength. (d) Pretreated spectral data reversing the order of transformation by OSC and DWT.

dataset cannot correlate with the variation in any property of the pulp samples.

Standard normal variant (SNV) transformation and multiplicative scatter correction (MSC) have similar effect on the NIR data set. Comparing plots in Figure 1, we find that both of these pretreatment methods highlight structural features while reducing baseline offsets and high-frequency noise.

Discrete wavelet transform enables the selective application of low and high-frequency filters that remove baseline offsets and noise with great effectiveness. Figure 2(a) plots the full dataset after a seven-level decomposition in a Symlet-5 basis followed by the removal of the three highest frequency components (*details*), and the lowest one (*approximations*).

Here we see an elimination of baseline slope and offset, together with a significant amplification of reproducible structure over the full range of the spectrum and the elimination of high-frequency noise.

4.2 Effects of target-directed pretreatment

All of the data preprocessing methods illustrated above can serve to isolate structure and reduce noise in the NIR spectra of cellulosic pulps. However, these global, non-targeted preprocessing strategies do not distinguish between a spectral feature that correlates with a target variation in a mechanical or physical property of the paper made from these pulps and one that does not.

Target-directed preprocessing techniques make reference to measured standards in order to select features that enhance the

determinate variations in a data set, while suppressing variations that are uncorrelated or orthogonal to a targeted property. We explore the effectiveness of such a feature selection strategy for the NIR classification of pulps by applying the method of orthogonal signal correction (OSC).

Figure 2(b) plots the the present NIR data set following OSC which has been targeted to a prediction of the paper sheet mechanical property, tensile strength.

5 Discussion

5.1 Summary overview of pretreatment effects

First-derivative preprocessing, as displayed in Figure 1(b), clearly increases the high-frequency noise in the spectral range above 7000 cm^{-1}. However, as discussed below, we find its elimination of the baseline improves the prediction models that we can construct for most target parameters. Background and noise both contribute to limit the accuracy of a prediction model, and for the present case, stabilizing the baseline outweighs the introduction of completely uncorrelated high-frequency noise.

SNV and MSC pretreatment (Figure 1(c) and (d)) yield spectra with a similar structure. Neither method enhances resolution or discrimination. They do decrease the high-frequency noise. But they leave spectra with background os-

cillations that do not correlate with target variance.

DWT operates in a more deterministic way to decompose the spectrum into subcomponents based on the bandwidth of its feature variations. Each such component represents different domain of information. The conformance of features in Figure 2(a) with vibrational overtone line shapes, shows how a choice of bandwidth scale can substantially reduce baseline variations and high-frequency noise while highlighting determinate information.

Pretreatment by OSC sharpens resolution compared with raw spectra, and serves to accentuate variation relevant to target properties. However, its implementation here introduces noise, particularly above $9,000\,\text{cm}^{-1}$, and fails to suppress the background.

Thus, we see that wavelet transform offers an effective means to subtract background and remove noise, while OSC suppresses irrelevant information. Seeking the discrimination enhancement afforded by OSC, for spectra filtered against background and noise, we have explored combinations of these prepossessing methods. Figure 2(c) shows that OSC applied to a spectral data set pretreated first by DWT yields a baseline-stabilized NIR signature with apparently enhanced signal variance, but also some increase in high-frequency noise.

Figure 2(d) reverses the strategy. Here we remove orthogonal variance to enhance intensity differences that correlate best with the target. Then we de-noise by applying DWT. This produces a sparser representation, with less high-frequency noise and better isolated variation within the bandwidth of the molecular NIR spectral response.

To assess the effectiveness with which preprocessing improves the accuracy of classification we have built multivariate models from subsets of pretreated spectra, and run independent validations to estimate residual mean errors of prediction.

5.2 Prediction of physical properties on the basis of NIR classification models

Our study examines the effectiveness of various methods for preprocessing NIR spectra of pulp on the success of multivariate prediction models for nine physical properties of paper, including tensile, freeness, burst, DZS and WZS, tear, SRE600, absorption, and scattering. We have constructed PLS models and calculated the accuracy of prediction for each property to evaluate the efficiency of each preprocessing method.

Ultimately, the accuracy that any of these assessment can have is limited by accuracy of standard measurements used to determine property values of our standards. As indicated in Table 1 the measurements that determine the properties of the samples used as standards in this study vary in their reproducibility. Bearing this limitation in mind, we proceed now to compare the relative effect of various methods of data pretreat-

ment on classification accuracy within each individual property.

We carry out this comparison by applying different analytical criteria, such as accuracy, precision, and linearity to evaluate PLS models. With reference to Tables 3 and 2, we evaluate the prediction models for tensile strength using untreated and first-derivative treated spectra.

We see that, using a sufficient number of PLS factors, untreated spectra can serve to yield a linear PLS calibration model. The application of a simple first-derivative pretreatment reduces the number of factors required and improves performance at the stage of validation. But, we find that the regression model after first-derivative preprocessing exhibits an intercept that differs more from zero. Evidently, the reduction in background with pretreatment decreases scatter, but the increase in high-frequency noise adds a systematic prediction offset.

Using spectral data sets preprocessed by SNV and by MSC. We find a moderate improvement in tabulated RMSEP values, and PLS regression models with (x, y) intercepts closer to $(0,0)$, suggesting that these pretreatment methods better succeed in removing random variance without adding a systematic offset to the model.

Of the remaining two pretreatment methods available to us in this study, DWT provides a means to filter slowly varying background and highly oscillatory noise, while OSC reduces the dimensionality of the data by suppressing variance in coordinates deemed orthogonal to the targeted analysis. Table 3 lists the errors of prediction obtained for paper properties by applying these pretreatments individually. Here we see comparable levels of improvement in RMSEP owing to the removal of background and noise (DWT) or the amplification of relevant signal components (OSC). OSC, applied alone, yields a slightly better prediction model for tensile strength.

Because pretreatment by DWT and OSC suppress uncorrelated variance in distinctively different ways, it seems reasonable to explore whether there is further accuracy to gain by preprocessing the spectral data both by DWT and OSC. Table 3 shows the results of such a strategy both for the case in which we first denoise the data by DWT and then apply OSC, and for the case in which we first suppress uncorrelated variance by OSC and then remove uninformative noise by DWT.

We find, by and large, that pretreatment in the sequence OSC-DWT yields calibration models with the lowest RMSEP and smallest intercept for almost all of the properties tested. Table 3 gives average RMSEP value found by the application of ten distinct PLS regression models (using ten randomly selected partitionings of spectra into calibration and validation datasets) to predict each physical property using NIR spectra preprocessed by each of the methods discussed above. We find for predicting tensile strength that the order, OSC followed by DWT, yields a significantly lower average RMSEP. This or-

Table 2 Optimized number of PLS factors, h, and coefficients of determination (correlation coefficients) obtained from PLS regression model fits to calibration data-set, following various methods of NIR spectral pretreatment.

Parameters	Non h	Non r^2	1D h	1D r^2	SNV h	SNV r^2	MSC h	MSC r^2	DWT h	DWT r^2	OSC h	OSC r^2	DWT-OSC h	DWT-OSC r^2	OSC-DWT h	OSC-DWT r^2
Tensile	10	0.95	3	0.94	10	0.94	10	0.94	7	0.94	7	0.96	7	0.99	7	0.96
Freeness	10	0.68	3	0.70	7	0.53	7	0.53	7	0.71	7	0.74	7	0.79	7	0.70
Burst	10	0.91	3	0.92	10	0.95	10	0.95	7	0.95	7	0.98	7	0.99	7	0.95
Dry zero span	10	0.82	3	0.77	10	0.83	10	0.83	7	0.89	7	0.91	9	0.95	10	0.91
Wet zero span	7	0.72	3	0.55	7	0.68	7	0.68	7	0.85	7	0.89	7	0.95	7	0.86
Tear	7	0.62	2	0.68	7	0.70	7	0.70	7	0.78	7	0.97	7	0.98	7	0.97
SRE600	10	0.75	3	0.79	10	0.78	10	0.78	7	0.87	7	0.94	7	0.96	7	0.94
Absorption	10	0.82	3	0.86	10	0.86	10	0.86	4	0.87	10	0.94	7	0.97	10	0.96
Scattering	10	0.90	3	0.83	7	0.82	7	0.82	7	0.91	9	0.96	7	0.98	7	0.96

der gives a slightly more accurate prediction of DZS. For all other properties the difference falls within the uncertainty with which we can specify the prediction error based on ten random models.

Table 3 also reports the standard deviation of the distribution of RMSEP in each case. This quantity reflects the precision to which we can specify the prediction improvement achieved by a given pretreatment strategy. The lower this standard deviation, the more consistently the pretreatment strategy operates to improve prediction accuracy for a given property, regardless of the conformance that might happen to exist between a particular choice of calibration and validation sets. A higher standard deviation indicates an inconsistency in prediction accuracy, reflecting an incomplete suppression of variance in the pretreated data.

Table 3 shows that OSC yields a relatively broad distribution of RMSEP results, characterized by a larger standard deviation in each case. However, combining OSC with DWT narrows the distribution of RMSEP for almost all parameters. As a targeted preprocessing method, OSC operates best when the validation spectra happen to vary over a data space that is well represented by the calibration spectra (including orthogonal variance). DWT is non-targeted, and preforms similarly in reducing error regardless of the degree of conformance that happens to exist between the calibration and validation data spaces.

Combining DWT with OSC in either order lowers RMSEP with greater reproducibility for all properties except freeness. Apparently, the application of a wavelet filter acts to reduce random variance, yielding spectra of greater uniformity from which to draw calibration and validation datasets.

The correlation coefficient of a calibration model describes how well the variance measured for a set of standards conforms to the variance known for those standards. A model for which the calibration measurements strongly correlated with the known standard quantities, r^2 approaches 1. The degree to which r^2 differs from 1 describes the unexplained variance in the calibration model.

As can be seen in Table 2, any application of OSC removes components of the measurement with variations orthogonal to a standardization, and this improves r^2 compared with simpler treatments. Also evident in Table 2, the application of DWT after OCS appears in some cases to reintroduce orthogonal variance, which can be seen in a slightly worsened r^2 for calibration compared with DWT-OSC. Nevertheless, on balance, values of RMSEP obtained in validation favours OSC-DWT in most cases, and we find that this order of pretreatment better isolates spectral features for chemical interpretation.

5.3 Preprocessing as a means of isolating spectroscopic features

NIR spectroscopy presents the inevitable problem of highly overlapped spectroscopic features. By suppressing uncorrelated variance, all forms of pretreatment can serve to accentuate the elements of a spectrum that correlate the variance of interest. First-derivative, SNV and MSC preprocessing methods achieve this by reducing noise and baseline variations. DWT flattens baseline, much like first-derivative preprocessing, and removes high-frequency noise. OSC serves to highlight features of determinate variance by subtracting orthogonal information from the spectrum.

Examining the spectroscopic outcome of pretreatment can provide a visual tool to understand the source of a target variance in terms of its spectroscopic manifestation. The physical properties of a pulp relate to each other by common correlations with fibre morphology and chemical composition. Feature-selected spectra, built to conform individually with target properties should therefore exhibit similarities signalling these underlying chemical and physical correlations.

Table 3 Average residual mean square errors of prediction following the application of various methods of NIR spectral data pretreatment. The RMSEP averages determined in each case by partial least squares calibration models using ten different randomly selected training and test data sets. Tabulated standard deviations reflect the reproducibility of RMSEP following different preprocessing methods.

Parameter	PLS	1st-PLS	SNV-PLS	MSC-PLS	DWT-PLS	OSC-PLS	DWT-OSC-PLS	OSC-DWT-PLS
Tensile (km) (2.5-4.9)	0.2742± 0.0096	0.2473± 0.0061	0.2214± 0.0033	0.2213± 0.0033	0.1985± 0.0019	0.1960± 0.0018	0.1958± 0.0018	0.1754± 0.0013
Freeness (ml) (649.5-700.5)	12.04± 1.603	11.92± 0.9726	10.18± 0.8147	10.17± 0.8165	9.327± 0.7857	10.02± 0.8009	10.03± 0.7741	10.10± 0.7437
Burst (kPa m^2 g^{-1}) (1.4-3.4)	0.2509± 0.0067	0.2455± 0.0048	0.2349± 0.0022	0.2348± 0.0024	0.2183± 0.0022	0.2229± 0.0027	0.2175± 0.0007	0.2171± 0.0016
Dry Zero Span (km) (14.44-16.8)	0.4556± 0.0020	0.4344± 0.0023	0.4109± 0.0007	0.4109± 0.0007	0.4081± 0.0009	0.4169± 0.0015	0.4062± 0.0006	0.4025± 0.0008
Wet Zero Span (km) (12.4-15.8)	0.5881± 0.0052	0.5677± 0.0045	0.5266± 0.0030	0.5264± 0.0029	0.5010± 0.0022	0.5145± 0.0032	0.5023± 0.0024	0.5018± 0.0022
Tear (mN m^2 g^{-1}) (19.9-30.3)	1.8163± 0.0046	1.8001± 0.0043	1.7421± 0.0075	1.7418± 0.0079	1.6994± 0.0086	1.7099± 0.0097	1.6898± 0.0038	1.7061± 0.0059
SRE600 (23.5-82.7)	12.7136± 0.0705	12.6064± 0.0661	12.1212± 0.0439	12.1215± 0.0440	12.3330± 0.0511	12.3978± 0.0713	12.2578± 0.0563	12.2095± 0.0471
Absorption (m^2 g^{-1}) (0.1669-0.2278)	0.0094± 0.0007	0.0093± 0.0007	0.0090± 0.0007	0.0090± 0.0007	0.0085± 0.0005	0.0087± 0.0005	0.0082± 0.0006	0.0084± 0.0005
Scattering (m^2 g^{-1}) (32.6-39.8)	1.0116± 0.0052	1.2047± 0.0052	0.8780± 0.0054	0.8777± 0.0053	0.8101± 0.0036	0.9006± 0.0050	0.8093± 0.0029	0.8033± 0.0055

The spectra obtained here, following pretreatment geared to properties known to be correlated, present evidence supporting this idea.

For example, Figures 3 to 5 compare NIR spectra after OSC-DTW pretreatment tied to Burst and Tensile Strength, Wet Zero-Span and Dry Zero-Span, and SRE600 and Tear. In the universe of standard samples for this study, Burst and Tensile Strength, which range over a factor of two, correlate linearly, with a least-squares correlation coefficient of 0.964. We see this reflected in the feature-selected spectra obtained following OSC-DWT, which exhibit a large variation in the region of from 4300 to 5800 cm^{-1} with distinct isosbestic points at 4700 and 4900 cm^{-1}.
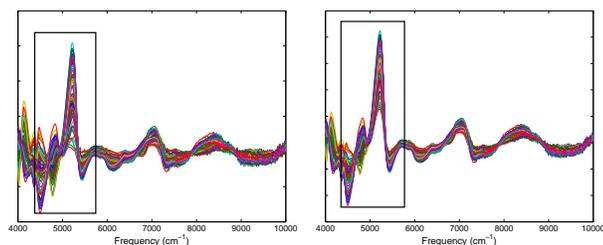
Fig. 3 Pulp NIR spectra after processing by combination of orthogonal signal correction with two components and discrete wavelet transform with selected wavelet range of (3-7) with reference to (left) Burst, and (right) Tensile Strength.

Wet Zero-Span and Dry Zero-Span show a similar degree of correlation in our standards, but far less variation. OSC-DWT preprocessing tied to either of these properties yields a similar set of resonances in the region of 5000 cm^{-1}, which are very different from the sub-spectra characteristic for Tensile and Burst. These features span the WZS-DZS range of our samples with a much smaller amplitude variation.
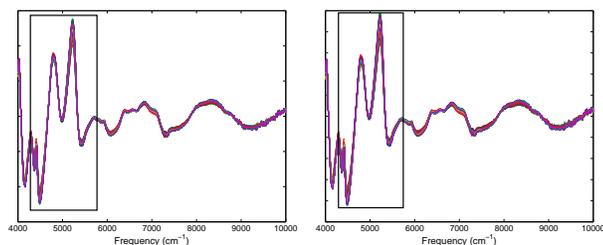
Fig. 4 Pulp NIR spectra after processing by combination of orthogonal signal correction with two components and discrete wavelet transform with selected wavelet range of (3-7) with reference to (left) Wet Zero Span, and (right) Dry zero span.

Finally, SRE600 refers to a pulp property that determines the amount of refining energy necessary to reach a Freeness

of 600. This property relates to fibre length much like that characteristic of the fibre network measured by Tear. In our samples, we find that SRE600 does correlate well with Tear, and spectrally we see that OSC-DWT selects for a very similar set of features.
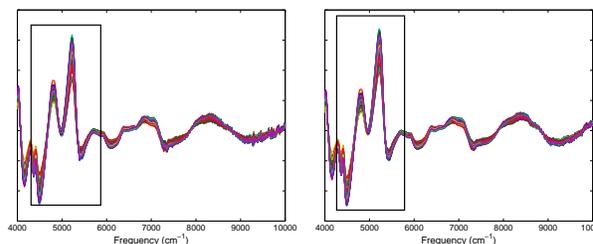
Fig. 5 Pulp NIR spectra after processing by combination of orthogonal signal correction with two components and discrete wavelet transform with selected wavelet range of (3-7) with reference to (left) SRE-600, and (right) Tear.

Among all the pulp properties investigated in this study, Scattering is unique in that it represents an optical property of the fibre network, which correlates poorly with indexes derived from measures relating to fibre network strength. As shown in Figure 6, amplitude of the features in the NIR spectra of pulp selected to isolate Scattering shifts decidedly to the red compared to strength-related elements of the spectrum, and variance concentrates in a resonance at 5100 cm^{-1}.

As also shown in Figure 6, we find that OSC-DWT selects a unique set of features to describe Freeness as well, despite the expectation that, in its morphological and chemical basis, freeness should correlate with other measures related to fibre length such as SRE600, Tensile Strength, etc. However, we see from Table 3 that Freeness represents the only parameter in this study for which prediction from the NIR spectrum was not improved by OSC-DWT preprocessing. We conclude from this that the OSC algorithm, applied with reference to Freeness, fails to capture the most correlated features. For this reason, we must not regard the OSC-DWT selected features as a trustable source for spectroscopic interpretation for Freeness.

5.4 Conclusion

We have explored the effectiveness with which a range of preprocessing methods improves the multivariate prediction of paper physical and morphological properties from the NIR spectra of pulp fibres. Simple data treatments, including first-derivative, SNV and MSC preprocessing do reduce prediction errors. However, DWT, which removes high-frequency noise and low-frequency baseline variations, and OSC, which suppresses variance orthogonal to a specified target property, im-
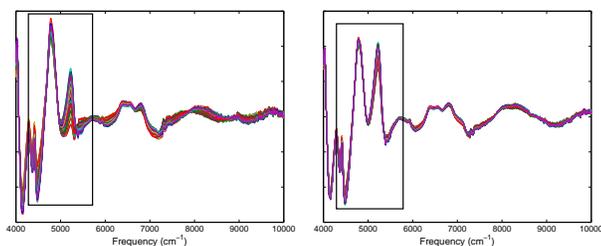
Fig. 6 Pulp NIR spectra after processing by combination of orthogonal signal correction with two components and discrete wavelet transform with selected wavelet range of (3-7) with reference to (left) Scattering, and (right) Freeness.

prove performance to a better degree. Interestingly, DWT, which preprocesses without reference to a target property, yields regression coefficients that fall shorter of one, indicating the presence of a systematic residual uncorrelated variance. Combining the pretreatment methods, OSC and DWT yields the smallest errors. Using Tensile Strength as a benchmark, the sequence, OSC then DWT produces a more robust prediction, featuring a lower RMSEP with a narrower distribution. Either order improves RMSEP and standard deviation of prediction error to similar degrees for other properties, For all cases the application of DWT followed by OSC gives a regression coefficient closer to one, suggesting that this approach best avoids the biasing of noise suppression with respect to magnitude of the target property.

NIR spectra corrected by OSC-DWT reveal interpretable features with respect to selected properties. Correlations in such spectra provide a means to logically connect certain properties with assigned features. These results underline the effectiveness with which OSC-DWT extracts essential information from spectra subject to uncorrelated variations. Thus, we can conclude that the application of OSC-DWT as a pretreatment can improve the utility of a NIR spectroscopy as a method for predicting the end-point properties of pulp and paper.

References

1   S. Tsuchikawa, *Appl Spectrosc Rev*, 2007, **42**, 43–71.
2   M. Birkett and M. Gambino, *TAPPI J*, 1989, **72**, 193–197.
3   A. Marklund, M. Paper, J. Hauksson and M. Sjostrom, *Nordic Pulp & Paper Res J*, 1999, **14**, 140–148.
4   G. Yu-mei and Z. Wei, *Spectrosc Spect Anal*, 2008, **28**, 1544–1548.
5   T. Schultz and D. Burnds, *TAPPI J*, 1990, **73**, 209–212.
6   L. G. Thygesen and S.-O. Lundqvist, *J Near Infared Spectrosc*, 2000, **8**, 191–199.
7   S. Tsuchikawa and S. Tsutsumi, *J Matl Sci Lett*, 1998, **17**, 661–663.
8   L. Thygesen and S. Lundqvist, *J Near Infared Spectrosc*, 1998, **8**, 191–199.
9   T. Lestander and C. Rhen, *Analyst*, 2005, **130**, 1182–1189.
10 W. Gindl and A. Teischinger, *Wood and Fiber Sci*, 2002, **34**, 651–656.
11 L. Schimleck and R. Evans, *Appita J*, 2003, **56**, 312–317.
12 B. Via, T. Shupe, L. Groom, M. Stine and C. So, *J Near Infared Spectrosc*, 2003, **11**, 365–378.
13 S. Tsuchikawa and S. Tsutsumi, *Appl Spectrosc*, 1999, **53**, 1033–1039.
14 W. Gindl, A. Teischinger, M. Schwanninger and B. Hinterstoisser, *J Near Infared Spectrosc*, 2001, **9**, 255–261.
15 F. Poke, J. Wright and C. Raymond, *J Wood Chem and Tech*, 2004, **24**, 55–67.
16 T. Lindgren and U. Edlund, *Nordic Pulp & Paper Res J*, 1998, **13**, 76–80.
17 F. Poke and C. Raymond, *J Wood Chem and Tech*, 2006, **26**, 187–199.
18 T. Lindgren, U. Edlund and T. Iversen, *Cellulose*, 1995, **2**, 273–288.
19 H. Henriksen, T. Naes, R. Rodbotten and A. Aastveit, *Chemom Intell Lab Syst*, 2005, **77**, 238–246.
20 J. Wright, M. Birkett and M. Gambino, *TAPPI J*, 1990, **73**, 164–166.
21 L. Chimleck, P. Kube and C. Raymond, *Can J Forest Res*, 2004, **34**, 2363–2370.
22 S. Tsuchikawa, Y. Hirashima, Y. Sasaki and K. a. Ando, *Appl Spectrosc*, 2005, **59**, 86–93.
23 M. Birkett, M. Gambino, J. Meyer and D. Egers, *Tappi J*, 1989, **72**, 193–197.
24 L. Schimleck and Y. Yazaki, *Holzforschung*, 2003, **57**, 520–526.
25 P. Fardim, M. Ferreira and N. Duran, *J Wood Chem and Tech*, 2002, **22**, 67–81.
26 P. Fardim, M. Ferreira and N. a. Duran, *J Wood Chem and Tech*, 2005, **25**, 267–279.
27 Technical Association of the Pulp and Paper Industry, `http://www.tappi.org/Bookstore/Standards-TIPs/Standards.aspx`.
28 A. T. Giese and C. S. French, *Appl Spectrosc*, 1955, **9**, 78–96.
29 R. N. Hager and J. R. C. a. Anderson, *JOSA*, 1970, **60**, 1444–1449.
30 G. Baldini, E. Grilli and M. Guzzi, *Appl Opt*, 1975, **14**, 2687–2690.
31 P. Geladi, D. MacDougall and H. Martens, *Appl Spectrosc*, 1985, **39**, 491–500.
32 R. J. Barnes, M. S. Dhanoa and J. S. Lister, *Appl Spectrosc*, 1989, **43**, 772–777.
33 H. Martens and E. Stark, *J Pharm Biomed Anal*, 1991, **9**, 625–635.
34 H. Martens, J. Nielsen and S. Engelsen, *Anal Chem*, 2003, **75**, 394–404.
35 A. A. C., *Chemom Intell Lab Syst*, 1999, **47**, 51–63.
36 D. Chen, Z. Chen and E. Grant, *Anal Bioanal Chem*, 2011, **400**, 625–634.
37 F. Ehrentreich, *Anal Bioanal Chem*, 2002, **372**, 115–121.
38 B. Walczak and D. Massart, *Trends Anal Chem*, 1997, **16**, 451–463.
39 S. Wold, H. Antti, F. Lindgren and J. ?hman, *Chemom Intell Lab Syst*, 1998, **44**, 175–185.
40 B. M. Wise, *Orthogonal Signal Correction*, `http://www.eigenvector.com/MATLAB/OSC.html`.
41 J. Sun, *J Chemom*, 1997, **11**, 525–532.

42 R. N. Feudale, H. Tan and S. D. Brown, *Appl Spectrosc*, 2003, **57**, 1201–1206.

43 S. Wold, M. Sjöström and L. a. Eriksson, *Chemom Intell Lab Syst*, 2001, **58**, 109–130.

44 P. Geladi and R. B. Kowalski, *Anal Chimica Acta*, 1986, **185**, 1–17.

45 A. Höskuldsson, *J Chemom*, 1988, **2**, 211–228.

46 S. de Jong, *Chemom Intell Lab Syst*, 1993, **18**, 251–263.