Near-Infrared Spectroscopic Separation of Green Chain Sub-Alpine Fir Lumber from a Spruce-Pine-Fir Mix

Amir Sohi,* Stavros Avramidis, and Shawn Mansfield

The intention of this exploratory study was to determine whether near-infrared spectroscopy, combined with multivariate statistical modeling, could become a swift and accurate tool for identifying sub-alpine fir within a typical spruce-pine-fir (SPF) lumber mix in the green chain of a sawmill. This need arises from the difficulty encountered in the drying sub-alpine fir. Its identification and removal from the SPF mix before kiln drying may be quite beneficial for producing high quality lumber. Near-infrared spectra were obtained from scanning of small specimens that were prepared from freshly cut trees. The results of the initial principal component analysis indicated that all four components could be used for species differentiation with the help of partial least squares discriminant analysis. All specimens in the training set were fitted into the correct sub-group of either fir or spruce-pine groups. The test set was validated and it revealed that all specimens were correctly classified. The outcome also confirmed that near-infrared spectroscopy combined with multivariate statistical modeling could be a suitable prediction model for separation of sub-alpine fir from the SPF mix.

Keywords: Multivariate statistical modeling; Near-infrared spectroscopy; Principal component analysis; Spectra; Sub-alpine fir; Spruce-pine-fir; Kiln drying; SPF

*Contact information: Department of Wood Science, University of British Columbia, Vancouver, BC, Canada, V6T 1Z4; *Corresponding author: amirsohii@yahoo.ca*

INTRODUCTION

Spruce, pine, and fir (SPF) is a species group comprised of *Picea glauca*, *Pinus contorta*, and *Abies lasiocarpa*, that have been processed and marketed worldwide from the interior British Columbia, Canada, for primarily construction applications. Although these three species are quite similar in a visual sense, they still display substantial differences in properties and processing. For instance, the sub-alpine fir, known as "F-part" of the SPF mix, is considered problematic in the kiln-drying process due to the presence and irregular distribution of large volumes of wet-wood. The wet-wood (a.k.a. "wet-pockets") is referred to areas in heartwood with abnormally high moisture content that are almost impossible to dry. Such features considerably reduce the overall quality and value of produced lumber and veneer (Alkan *et al.* 2007). Furthermore, sub-alpine fir has low permeability, which also causes long drying times, thus affecting kiln capacity problems for many SPF lumber producers (Zhang and Cai 2008). Consequently, wet-wood and low permeability result in high after-drying moisture contents within the F-part of the SPF mix, which increases the standard deviation and percent of under-dried (a.k.a. wets) lumber and is detrimental to production quality control and sawmill revenues (Elustondo and Avramidis 2005). Therefore, if sawmills had a quick and precise system to remove sub-alpine fir from the SPF mix while still in the green chain, production of high quality dried lumber could be feasible.

Near-infrared (NIR) spectroscopy is a swift and accurate method for material characterization. The method consists of illuminating a surface using an NIR source, collecting the infrared reflected spectrum (wavelength region of 350 to2500 nm), and thereafter, statistically analyzing the obtained spectrum. The feasibility of NIR as a non-destructive method for multiple wood attributes classification was assessed in various studies. The NIR has been proven capable for moisture-based sorting of green hem-fir timber (Watanabe *et al.* 2011), differentiating among wood species (Pastore *et al.* 2011), distinguishing between sapwood and heartwood (Haartveit and Flæte 2008; Sandberg and Sterley 2009), and quantifying pulp yield, micro fibril angle, density distribution, and grain angle, among others (Tsuchikawa *et al.* 2003). Wood density distribution, grain angle, and chemicals (extractives) could also be quantified by NIR analysis (Gindl and Teischinger 2002; Schimleck and Evans 2003; Via *et al.* 2005).

NIR spectroscopy has yielded promising results when used to classify various types of wood. Brunner *et al.* (1996) used NIR to differentiate woods from 12 tree species, mainly tropical. Schimleck *et al.* (1996) utilized NIR and principal component analysis for eucalypt wood classification. Borga *et al.* (1992) reported that NIR could distinguish milled samples from wet-stored timber of *Pinus sylvestris* L. (Scots pine) into heartwood and sapwood. NIR data from solid wood were successfully used in multivariate statistical models to separate the heartwood from sapwood in specimens of Scots pine (Flæte and Haartveit 2003). Results also indicated that NIR might differentiate between not only wood samples of the same species, but also from different origins (Gierlinger *et al.* 2004). Dawson and Adedipe (2012) combined NIR and fluorescence spectroscopy to separate three Canadian softwoods, namely, balsam fir, western hemlock, and white spruce. Lastly, Nisgoski *et al.* (2015) employed NIR spectroscopy to identify *Cryptomeria japonica* varieties from southern Brazil.

Because of past successes for NIR spectroscopy in wood species differentiation, the prospect of analyzing spectra with multivariate statistical modeling and using the results as a means of identifying and separating green sub-alpine fir lumber from the SPF mix was undertaken as the objective of this project.

EXPERIMENTAL

Fifteen randomly selected trees, five of each species of white spruce, lodgepole pine, and sub-alpine fir, were harvested from the UBC Alex Fraser Research Forest near Williams Lake (British Columbia, Canada) in the fall of 2015. One-meter long bolts from the breast height (1.5 m above the ground) were cut from each tree. The fifteen bolts were thereafter processed into slabs with a Wood-Mizer sawmilling (LT28 Portable Sawmill from Wood-Mizer LLC, USA) system and 2250 small clear specimens of 100×100×20 mm (L × T × R directions) were cut -150 from each bolt. The specimens came in equal numbers of 50 from the three tree cross-sections, namely, all-sapwood, all-heartwood, and mixed.

NIR spectra were collected with QualitySpec® Pro Analytical spectrometer (Diameter, 25 mm; Field of View (FOV) – 13 mm X 11 mm - Analytical Spectral Devices Inc., Boulder, CO, USA) in a diffuse reflectance mode between 400 and 2400 nm wavelength region. Each specimen was placed on a fiber optic probe, and a piece of commercial micro porous Teflon was used as reference. The NIR spot area was approximately 50 × 20 mm. A single spectrum was obtained by averaging 10 independent scans. Two spectra were collected from each sample, one from the top and the other from

the bottom surface (tangential). Then, 3500 of the 4500 spectra captured were used as the calibration set (CS), and the remaining 1000 spectra were used as the validation set (VS). The reflectance data were expressed as apparent absorbance (log(1/reflectance)) were centered, corrected for baseline shifts, and then, they were transformed to first derivatives (Flaete *et al.* 2006).

Principal component analysis (PCA) is a linear modeling method that projects the original variables on to a smaller set of variable scaled principal components in order to compress the data matrix. By applying PCA to the spectral matrix, the large number of highly correlated variables is reduced to a few orthogonal principal components (Flaete *et al.* 2006).

Partial least squares (PLS) regression is a linear modeling method that compresses the spectral data and projects them onto partial least square components. The PLS1 method extracts the spectral information with the largest covariance to one dependent variable (Martens and Naes 1989). In the PLS2 method, two or more dependent variables were modeled simultaneously. The PLS components were created so that they were mutually orthogonal, thus avoiding problems related to co-linearity among the variables in the X-matrix. Compared to traditional statistical modeling based on least squares estimation, independent variables in the X-matrix were not a requirement for the partial least squares method. The PLS regression can also handle situations where the number of variables far exceeded the number of samples, which was typical for modeling with NIR data (Flæte *et al.* 2006).

PLS discriminant analysis involved developing a conventional partial least squares regression model, but instead of a continuous variable, the response one was a binary class indicator variable. If there were only two classes to separate, the PLS model used one response variable, which was coded for class membership as follows: 0 for members of one class, 1 for members of the other one (Flæte *et al.* 2006), and the PLS1 algorithm was then used. If there were three classes or more, the PLS2 was used. Each class was represented by an indicator variable, *i.e.*, a binary variable with value 1 for members of that class, 0 for non-members (Flæte *et al.* 2006). By building a PLS2 model with all indicator variables as Y, one can directly predict the class membership from the X-variables describing the samples. The model was interpreted by viewing predicted *vs.* measured for each class indicator Y-variable (Flæte *et al.* 2006):

- $Y_{pred} > 0.5$ means "member";
- $Y_{pred} < 0.5$ means "non-member"

Because in this work each specimen species was known in advance, three binary variables were constructed, namely, one for each tree species. Each wood specimen was thereafter, assigned to the group for which the model had the highest estimated response.

The PLS discriminate analysis model was validated using a test-set validation. Although the preferred material for a test-set is obtained through new sampling, a frequently used alternative is to split the obtained data into a training and a test set. A model that performs well when validated using test sets has a higher chance of performing well when predicting new samples. The samples were divided into a calibration set and a validation set. The statistical analysis was performed with Unscrambler® (Camo Process AS, Oslo, Norway).

RESULTS AND DISCUSSION

There was a large color similarity among wood specimens from the three species, and that is one reason why SPF is marketed as a single "species". To recognize useful information for identifying wood species, an initial PCA was performed on the first derivatives of the spectra from the 2250 samples. The optimal number of principal components chosen was four, which described 85% of the original variance in X (design matrix in multivariate statistics). A scatter plot of principal components- 1 and 2 is shown in Fig. 1; they explained 72% and 16% of the original variance, respectively.

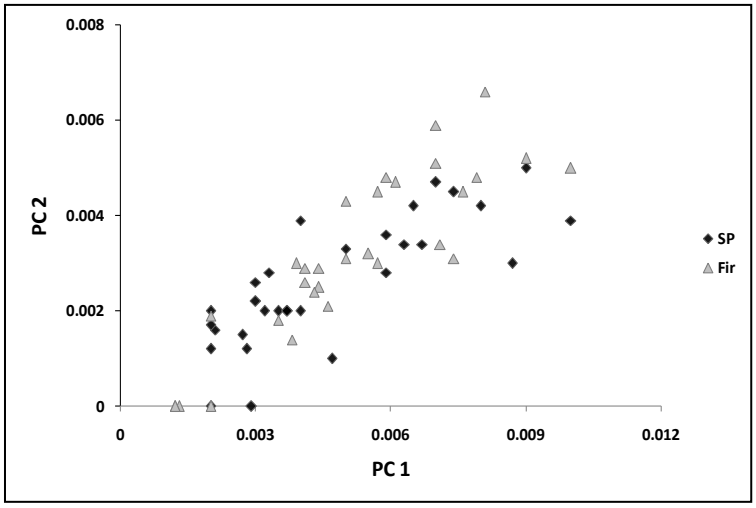

Fig. 1. Scatter plot of the first and second principal components of the NIR spectra

The scatter plot of principal components- 3 and 4 is shown in Fig. 2. Principal component-3explained 9% and principal component-4 explained 7% of the original variance in X.

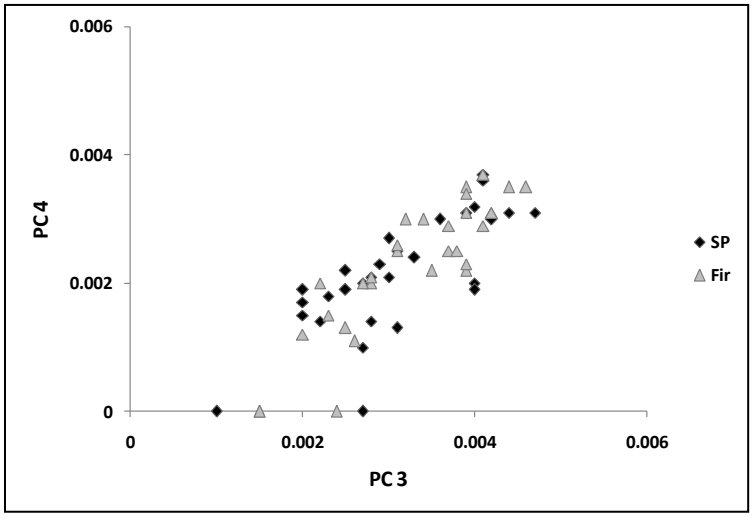

Fig. 2. Scatter plot of the third and fourth principal components of the NIR spectra

It seems that the NIR spectra somehow included relevant information for discrimination of the different tree species, but the information represented only a small fraction of the total variation in those spectra. Brunner *et al.* (1996) used NIR spectroscopy to differentiate 12 species of wood, and found that the comparison of test results was possible only when the samples had been prepared identically. It is possible that pre-treatment of the spectral data (*e.g.*, use of transformations) could have controlled this problem. However, this aspect has to be considered when calibration models for applications were generated where the preparation of wood specimens cannot be performed identically.

For the PLS discriminant analysis model, the optimal number of components used was four. The regression coefficients for the PLS2-regression models of SPF mix are given in Fig. 3. The regression coefficients did not show a strong inverse relationship. A first impression of model performance is given by the fitted values of the response variables. The fitted values based on model calibration are reported in Table 1.

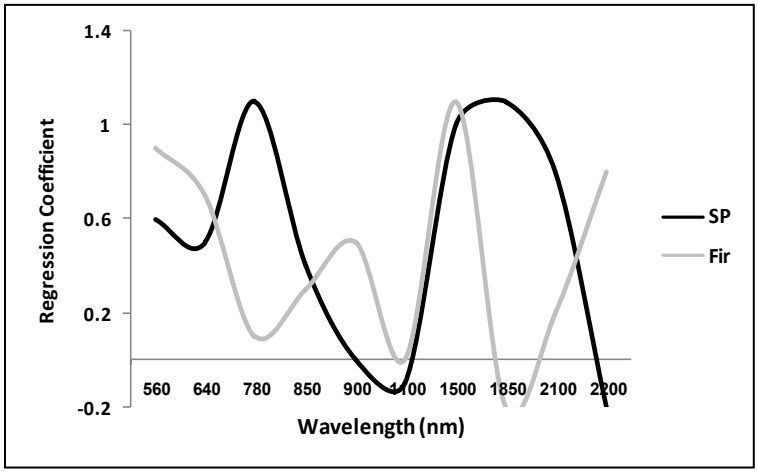

Fig. 3. Regression coefficients from the partial least squares discriminant analysis model

Table 1. Results from Partial Least Squares Discriminant Analysis Model Calibration (3500-Sample Training Set)

Fitted into group	True group	
	SP	Fir
SP	2333	0
Fir	0	1167
Total number	2333	1167
Number correct	2333	1167
Proportion correct	1.00	1.00

The results showed that all the specimens had fitted values that classified them to the correct tree species. The PLS discriminant analysis model was validated using test set validation as described previously (Table 2).

Table 2. Results from Partial Least Squares Discriminant Analysis Model Validation (1000-Sample Test Set)

Fitted into group	True group		Total correct
	SP	Fir	
SP	666	0	666
Fir	0	334	334
Number correct	666	334	1000
Total number	666	334	1000
Proportion correct	1.00	1.00	1.00

It has been argued that validation using data splitting is of little value (Kozak and Kozak 2003), and that observations should be saved for model calibration. Validation by data splitting provides little incremental information compared with the information revealed in standard statistical analysis, and that models should always be validated using external data. The requirements for individual sampling are somewhat unclear. Despite the high proportion of correct classifications, some wood specimens had high-predicted responses for several tree species, while other wood specimens had small-predicted responses for all tree species. This result showed that classification and sorting of tree species are not always straightforward. Still, this research demonstrated the high predictive accuracy of NIR for wood types that are closely related and have similar visual appearance (Tables 1 and 2).

The models will steadily improve as more wood specimens are classified because the classified specimens can be used in calibration sets for new and improved models. This is not surprising, as the regression coefficients for spruce and pine showed a strong inverse relationship along the completely spectral range. The PLS discriminant analysis resulted not only in correct identification of the specimens, but also the separation of sapwood from heartwood and mixed specimens. Despite the small calibration set and the structure of regression coefficients, the partial least squares discriminant analysis model performed well in classification of the sapwood and heartwood *versus* mixed specimens. This underpins the potential for calibrating partial least squares models with higher classification performance if a large number of calibration samples are used.

CONCLUSIONS

The results of near-infrared spectroscopy separation of green chain sub-alpine fir lumber from the spruce-pine-fir (SPF) mix showed that:

1. Near-infrared spectroscopy successfully resulted in models for wood species identification.
2. A small part of the total variance of the NIR spectra contained relevant information or discrimination of sub-alpine fir (F) from the white spruce/lodgepole pine (SP) part of the SPF mix in accordance to the principal component analysis.

REFERENCES CITED

Alkan, S., Zhang, Y., and Lam, F. (2007). "Moisture distribution changes and wetwood behavior in subalpine fir wood during drying using high X-ray energy industrial CT scanner," *Drying Technology* 25(1-3), 483-488. DOI: 10.1080/07373930601184023

Borga, P., Hämäläinen, M., and Theander, O. (1992). "Correlations between near-infrared spectra of wet-stored timber and storage timber in relation to water quality," *Holzforschung* 46(4), 299-303. DOI: 10.1515/hfsg.1992.46.4.299

Brunner, M., Eugster, R., Trenka, E., and Bergamin-Strotz, L. (1996). "FT-NIR spectroscopy and wood identification," *Holzforschung* 50(2), 130-134. DOI: 10.1515/hfsg.1996.50.2.130

Dawson, B., and Adedipe, O. E. (2012). "Rapid spectroscopic separation of three Canadian softwoods," *Wood Science and Technology* 46(6), 1193-1202. DOI: 10.1007/s00226-012-0468-9

Elustondo, D., and Avramidis, S. (2005). "Comparative analysis of three methods for stochastic lumber drying simulation," *Drying Technology* 23(1-2), 131-142. DOI: 10.1081/DRT-200047663

Flæte, P. O., Haartveit, E. Y., and Vadla, K. (2006). "Near infrared spectroscopy with multivariate statistical modeling as a tool for differentiation of wood from tree species with similar appearance," *New Zealand Journal of Forest Science* 36(2-3), 382–392.

Flæte, P. O., and Haartveit, E. Y. (2003)."Differentiation of Scots pine heartwood and sapwood by near infrared spectroscopy," The International Research Group on Wood Preservation Doc. No. IRG/WP 03-10459. /2003. IRG Secretariat, Stockholm.

Gierlinger, N., Schwanninger, M., and Wimmber, R. (2004). "Characteristics and classification of Fourier-Transform near infrared spectra of the heartwood of different larch species (*Larix* sp.),"*Journal of near Infrared Spectroscopy* 12(2), 113-119. DOI:10.1255/jnirs.415

Gindl, W., and Teischinger, A. (2002). "The potential of VIS- and NIR spectroscopy for the non-destructive evaluation of grain-angle in wood," *Wood Fiber Science* 34, 651–656.

Haartveit, E. Y., and Flæte, P. O. (2008)." Near infrared spectroscopy (NIRS) as a tool for effective classification of wood," *Proceedings of the 51st International Convention of Society of Wood Science and Technology, November 10–12, Concepcion, Chile.*

Kozak, A., and Kozak, R. (2003). "Does cross validation provide additional information in the evaluation of regression models?" *Canadian Journal of Forest Research* 33(6), 976–987. DOI: 10.1139/x03-022

Martens, H., and Naes, T. (1989). *Multivariate Calibration*, John Wiley & Sons, Chichester, UK. DOI: 10.1016/0584-8539(90)80069-B

Nisgoski, S., Schardosin, F. Z., Batista, F. R. R., Muñiz, G. I. B., and Carneiro, M. E. (2015). "Potential use of NIR spectroscopy to identify *Cryptomeria japonica* varieties from southern Brazil," *Wood Science and Technology* 50(1), 71-80. DOI: 10.1007/s00226-015-0783-z

Pastore, T. C. M., Willian, J., Braga, B., Coradin, V. T. R., Magalhaes, W. L. E., Okino, E. Y. A., Camargos, J. A. A., Muniz, J. I. B., Bressan, O. A., and Davrieux, F. (2011). "Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: Discrimination of true mahogany, cedar, andiroba, and curupixa," *Holzforschung* 65(1), 73-80. DOI: 10.1515/hf.2011.010

Sandberg, K., and Sterley, M. (2009). "Separating Norway spruce heartwood and sapwood in dried condition with near-infrared spectroscopy and multivariate data analysis," *European Journal of Forest Research* 128(5), 475-481. DOI: 10.1007/s10342-009-0296-0

Schimleck, L. R., and Evans, R. (2003). "Estimation of air-dry density of increment cores by near infrared spectroscopy," *Appita Journal* 56, 312-317.

Schimleck, L. R., Michell, A. J., and Vinden, P. (1996). "Eucalypt wood classification by NIR spectroscopy and principal component analysis," *Appita Journal* 49(5), 319-324.

Tsuchikawa, S., Inoue, K., Noma, J., and Hayashi, K. (2003). "Application of near-infrared spectroscopy to wood discrimination," *Journal of Wood Science* 49(1), 29-35. DOI: 10.1007/s100860300005

Via, B. K., So, C. L., Shupe, T. F., Stine, M., and Groom, L. H. (2005). "Ability of near infrared spectroscopy to monitor air-dry density distribution and variation of wood," *Wood and Fiber Science* 37(3), 394–402.

Watanabe, K., Mansfield, S. D., and Avramidis, S. (2011). "Application of near-infrared spectroscopy for moisture-based sorting of green hem-fir timber," *Journal of Wood Science* 57(4), 288–294. DOI: 10.1007/s10086-011-1181-2

Zhang, Y., and Cai, L. (2008). "Impact of heating speed on permeability of sub-alpine fir," *Wood and Science Technology* 42(3), 241-250. DOI: 10.1007/s00226-007-0172-3

Article submitted: December 28, 2016; Peer review completed: March 17, 2017; Revised version received: March 27, 2017; Accepted: March 28, 2017; Published: April 3, 2017.
DOI: 10.15376/biores.12.2.3720-3727