

Effects of variable selection and processing of NIR and ATR-IR spectra on **the prediction of extractives** content in *Eucalyptus bosistoana* heartwood

Yanjie [Li](#)^{a,b}

Clemens [Altaner](#)^{b,*}

clemens.altaner@canterbury.ac.nz

^aResearch Institute of Subtropical Forestry, Chinese Academy of Forestry, Fuyang, Zhejiang 311400, [PR China](#)

^bSchool of Forestry, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

*Corresponding author.

Abstract

The use of quick attenuated total reflectance infrared (ATR-IR) spectroscopy and near infrared (NIR) spectroscopy to predict extractives content (EC) in heartwood of *E. bosistoana* with partial least squares regression (PLSR) models was studied. Different spectra pre-processing methods and variable selection were tested for calibration optimisation. While variable selection substantially improved the NIR-PLSR models, only small effects were observed for spectra pre-processing methods and ATR-IR-PLSR models. Both of the NIR-PLSR and ATR-IR-PLSR models yielded reliably EC results with high R^2 and low root mean square error (RMSE). NIR based models performed better (RMSE 0.9%) than ATR-IR based models (RMSE 1.6%). Analysis showed that the models were based on IR signals assigned to chemical structures known from eucalyptus heartwood extracts. Combined with PLSR and variable selection, both, ATR-IR and the NIR spectroscopy, can be used to quickly predict EC in *E. bosistoana*, a measure needed in tree breeding and the quality control of for durable timber.

Keywords: [n](#)Natural durability; [s](#)Spectra pre-processing; [s](#)Standard normal variate (SNV); [p](#)Partial least squares regression (PLSR); [v](#)Variable selection (sMC)

1.1 Introduction

Wood is a renewable natural material and used for many purposes ranging from general construction work to reconstituted products. However, like for all natural materials, its properties are highly variable even within a species. A successful timber industry relies on good and uniform wood quality. This can be achieved by segregation [in](#)and tree breeding programs. Both approaches require rapid and cost effective methods to measure large numbers of samples. This is not trivial for many wood properties, but near infrared (NIR) and infrared (IR) spectroscopy have been shown potential as rapid, low-cost, non-destructive measurement techniques of various wood properties [1-3]. NIR spectra are typically acquired in reflection, reducing the demand on sample preparation. Reflection spectra can also be obtained [directly](#) from powders ~~without sample preparation~~ in the IR range with attenuated total reflectance (ATR) IR spectroscopy.

Multivariate data analysis, especially partial least square regression (PLSR), is typically used to obtain quantitative information from NIR and IR spectra [4]. This calibration process is dependent on spectra pre-processing [5] and variable selection [6]. Spectra modifications and variable selection attempt to minimise the effects of noise, baseline variation and additive effects arising from interfering physical and chemical factors [7]. To obtain a reliable calibration model, the samples should be split in calibration and validation data sets. This is often done by Kennard-Stone sampling, which samples based on Euclidian distances [8], for spectroscopy data as random selection can result in less accurate models [9]. In addition, IR spectra contain not only informative but also uninformative variables. Those uninformative variables can influence the robustness of the calibration models. Variable selection can improve the performance of calibration models by only selecting those wavelengths, which explain variation in the target variable [6]. Of the many methods, which have been developed for variable selection, filter methods can be easily combined with PLSR calibration models because variable selection is performed independently of model fitting [6]. The significance multivariate correlation (sMC) filter method not only reduces the effect of unrelated variables but concurrently aids spectra interpretation as it highlights the variables that are most correlated to the response [10].

This study focuses on the quantification of the extractive content (EC) in heartwood of *Eucalyptus bosistoana*, which is fast-growing in New Zealand's climate and produces [class 1](#) durable timber ~~(class 1)~~ [11]. *E. bosistoana* [has](#)ve been selected by the New Zealand Dryland Forests Initiative (NZDFI), which aims to establish a sustainable supply of ground durable timber for local and international markets [12,13]. The EC is the key factor for the natural durability of wood [14,15], which is very time consuming to measure directly [16,17]. Therefore, [e](#) rapid assessment of EC can be a quick indirect assessment of durability allowing quantification of heartwood quality for genetic selection in tree breeding, which benefits form large numbers of individual tree samples. Both, NIR [18-20] and IR [21] were reported suitable for assessing EC in wood. Predictions of mass loss caused by fungi based by IR spectroscopy ~~was~~

were not always accurate [22,23]. Depending on the application, either NIR or IR spectroscopy have been reported to be more accurate. For example, while IR spectroscopy outperformed NIR for the determination of β -carotene content in tomato fruits [24], the opposite was reported for chemical features of hardwoods [25].

The objectives of this work were 1) to investigate the potential of NIR and ATR-IR spectra to quantify the EC in *E. bosistoana* heartwood, 2) optimise the PLSR models based on NIR and ATR-IR spectra and 3) obtain information on the chemical structure of heartwood compounds in *E. bosistoana*.

2.2 Methods and Materials

2.1.2.1 Materials

E. bosistoana trees were planted in 2009 by the New Zealand Dryland Forests Initiative (NZDFI) in Marlborough, New Zealand. Discs with a length of ~10 cm were cut in May 2016 from the bottom of the tree trunk in two different sites (41° 26' S, 173° 56' E and 41° 43' S, 174° 02' E). In total, 100 *E. bosistoana* wood disc samples were collected. Each disc was air-dried (25 °C, 60% RH) for one month. The heartwood was isolated from all wood discs by drilling into the transverse face with a 12 mm diameter drill. Sapwood from ten randomly selected discs was collected in the same way. The drill 'dust' was ground in a Wiley mill fitted with a 20 mesh (0.85 mm) screen. The powdered samples were then oven dried at 60 °C to a stable moisture content (MC) of ~2%.

2.2.2.2 NIR and ATR-IR spectroscopy

NIR and IR spectra were collected with a Bruker Tensor 37 spectrometer controlled with OPUS_7.5.18 software (Bruker Optik GmbH, Germany). NIR spectra of wood powder were taken with a fibre-optics probe (Model N-500, Bruker Optik GmbH, Germany) at wavelength ranging from 9000 to 4000 cm^{-1} at 4 cm^{-1} intervals averaging 32 scans. IR spectra were obtained with an ATR accessory in combination with a RT-DLaTGS detector (Bruker Optik GmbH, Germany) in the range from 4000 to 800 cm^{-1} at 4 cm^{-1} intervals averaging 32 scans. Three spectra were taken from each wood powder in both regions, and the corresponding spectra were averaged.

2.3.2.3 Extractive content

Approximately 5–8 g of oven dry wood powder was precisely weighted into a stainless-steel cell and extracted with ethanol in an Accelerated Solvent Extractor (ASE) (Thermo Scientific) using the following extraction conditions: static time of 15 mins, temperature of 70 °C, 100% rinse volume and 2 extraction cycles.

Dry aluminium foil trays of known mass were used to hold the extractive solutions and left in a fume hood overnight to evaporate the ethanol. The extracts were subsequently oven dried at 105 °C. The dry mass of each extract was measured and the EC was calculated on a dry mass basis.

NIR and ATR-IR spectra were collected of dry ethanol heartwood and sapwood extracts from ten selected trees and averaged.

2.4.2.4 Data processing

The Kennard-Stone sampling method was used to divide the data into two subsets: 80 samples were selected for a calibration data set, and the remaining 20 samples were used as validation data set. The R software (version 3.1.2) [26] was used for data processing. The prospectr package [27] was used for NIR spectra manipulation and Kennard-Stone sampling. Three spectral pre-processing methods were tested: standard normal variate (SNV), 1st derivative and 2nd derivative. The Savitzky-Golay algorithm with a second-order polynomial and a window size of 15 was used to calculate the 1st and 2nd derivatives. The pls package [28] was used for developing the PLSR calibration models and optimal components selection with leave-one-out cross-validation. The sMC (alpha = 0.05) algorithm implemented by the plsVarSel package [6] was applied to both the NIR and ATR-IR spectra to a) study the effect of spectra pre-processing on the most important variables for PLSR models for heartwood EC and b) variable selection.

3.3 Results and Discussion

3.1.3.1 Preliminary Data Analysis

The EC ranged between 0.96% and 14.67% in the calibration data set with an average of 5.64% (Table 1). The range was with 2.08% to 7.19% smaller for the validation data set.

Table 1 Summary statistics of ethanol soluble extractive content (EC) in heartwood of 7 year-old *E. bosistoana* for the used datasets; CV: Coefficient of variation; n: number of selected samples.

alt-text: Table 1

Calibration data

Validation data

		(n = 80)	(n = 20)
EC	Max (%)	14.67	7.19
	Mean (%)	5.64	4.53
	Min (%)	0.96	2.08
	CV	0.61	0.41

NIR and ATR-IR spectra of *E. bosistoana* are dominated by cell walls and extractives are only a minor component. To identify characteristic signals for heartwood extractives the averaged NIR and ATR-IR spectra of wood powders were compared to those of dry extractives (Fig. 1). To resolve some signal overlap, the NIR spectra were converted into their 2nd derivatives [29]. The heartwood extractives showed unique signals at ~6900, 6017 and 4659 cm⁻¹ in the 2nd derivative NIR spectra. These were close to signals assigned to the 1st overtones of stretching vibrations of phenolic O-H [30] and C_{ar}-H groups as well as combinations of vibrations C_{ar}-H and C=O groups related to extractives in *E. globulus* [31], respectively. In the ATR-IR spectra, the heartwood extractives showed characteristic signals at 1712, 1330–1300, ~1174 and 860 cm⁻¹. Signals ~1712 cm⁻¹ were assigned to stretching vibrations of C=O groups of ketones, esters and conjugated aldehydes [32], while a signal at ~1330–1300 cm⁻¹ was likely to be related to phenolic O-H groups [33]. Two broader peaks at around 1350 cm⁻¹ and between 1290 and 1150 cm⁻¹ were reported to arise from combinations of C-O stretching and O-H deformation vibrations of hydrolysable and condensed tannins, respectively [34]. Vibrations from C-H groups were reported at 860 cm⁻¹ [33]. These assignments fit well to what is known on the chemical components in heartwood extracts of eucalypts, which were generally termed as polyphenol compounds including hydrolysable and condensed tannins [14,35]. Sapwood of *E. nitens* was shown to contain only small amounts of extractable phenolic compounds [36,37].

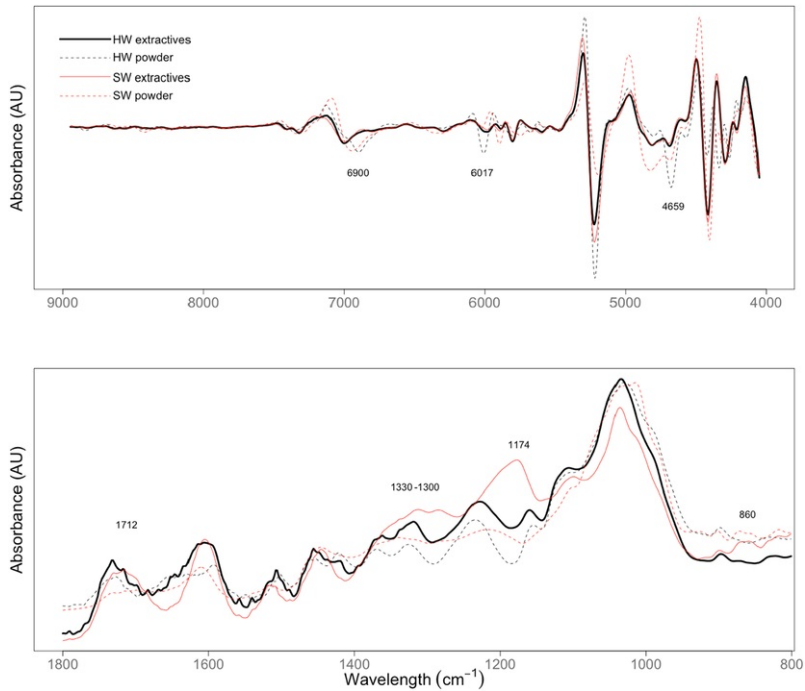


Fig. 1, Fig. 1 2nd derivate-NIR (top) and ATR-IR (bottom) spectra of heartwood (HW) and sapwood (SW) extractives as well as wood powders of *E. bosistoana*; AU: arbitrary units.

alt-text: Fig. 1

As the characteristic heartwood signals overlapped with peaks from the cell wall polymers, it was difficult to quantify the extractive content directly from the spectra. PLSR methods are typically used for quantitative analysis of IR spectra [4].

3.2.3.2 Effects of sSpectra pPre-pProcessing on PLSR mModels

The spectra can be negatively influenced by noise and baseline shifts. Spectral pre-processing methods are often used to reduce these effects on PLSR models [5]. Four pre-processing methods were applied to both NIR and ATR-IR spectra of *E. bosistoana* wood powder and their effect on PLSR models for EC were assessed (Fig. 2). Regardless of the pre-processing method, the PLSR models based on the NIR spectra performed better, with the residual mean square error (RMSE_{CV}) ranging between 0.99% and 1.25%, than those based on ATR-IR spectra (RMSE_{CV} 1.67 to 1.74%). This effect was consistent with earlier reports showing that NIR spectra allowed more accurate prediction of wood properties than ATR-IR [25,38,39]. SNV and 2nd derivative transformations did not improve model accuracy. For NIR spectra the models based on raw spectra and their 1st derivative performed best, however more components were needed using the raw spectra (Table 2). The applied spectra pre-processing methods had no effect on the performance of the ATR-IR models. The accuracy of the models were comparable to those reported for quantification of several chemical wood features by based on NIR and ATR-IR spectra [25].

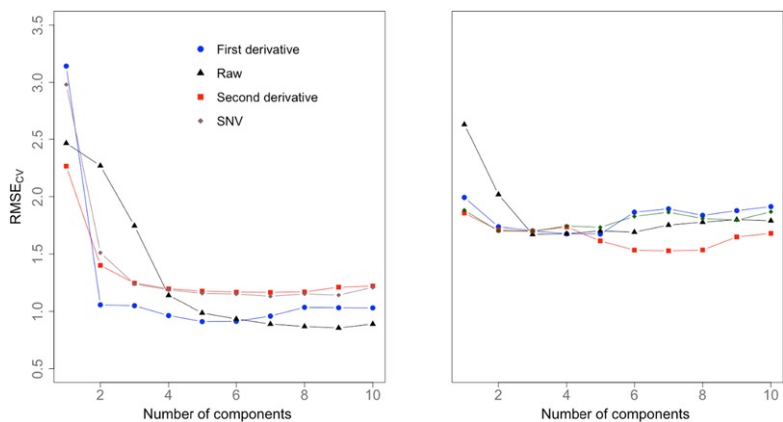


Fig. 2, Fig. 2 Residual mean square error of calibration using leave-one-out cross validation ($RMSE_{CV}$) for PLSR models predicting EC in *E. bosistoana* heartwood for different spectra pre-processing methods of NIR spectra (left) and ATR-IR spectra (right). SNV: standard normal variate, Raw: spectra without pre-processing.

alt-text: Fig. 2

Table 2 Characteristics of PLSR regression models for NIR and ATR-IR spectra for EC in *E. bosistoana* heartwood for different spectra pre-processing methods. R^2 : coefficient of determination and RMSE: root-mean-square error. The subscripts CV and p denote models based on a calibration data set using leave-one-out cross-validation and the predictions when the model was applied to the validation data set, respectively. SNV: standard normal variate, Raw: spectra without pre-processing.

alt-text: Table 2

Pre-treatment spectra		Calibration			Validation	
		R^2_{CV}	$RMSE_{CV}$	Ncomp	R^2_p	$RMSE_p$
NIR	Raw (None)	0.92	0.99	5	0.83	0.8
	SNV	0.90	1.25	3	0.88	1.28
	1st derivative	0.91	1.06	2	0.79	0.89
	2nd derivative	0.87	1.24	3	0.68	1.08
ATR-IR	Raw (NO)	0.72	1.67	3	0.72	1.38
	SNV	0.70	1.71	2	0.67	1.52
	1st derivative	0.71	1.70	3	0.70	1.43
	2nd derivative	0.71	1.70	3	0.70	1.44

3.3.3.3 Interpretation of PLSR Models Based on Different Pre-processing Methods

Interpretation of PLSR models is difficult as the model algorithms can interfere with the correlation between the spectra frequency and the target variable [10]. The sMC algorithm has been developed to facilitate PLSR model interpretation. The sMC method was used to investigate the effect of 4 different spectra pre-processing on the significant variables to predict EC in *E. bosistoana*, for both, the NIR and ATR-IR region.

Spectra pre-processing influenced the variables, which correlated to the EC (Figs. 3 and 4). The figures show the average spectra after signal processing as well as the explained variance of the individual frequencies (blue solid line). Compared to the unmodified spectra similar frequencies but with different relative importance were identified by sMC after SNV normalisation. These signals also appeared sharper. An exception was observed for the region from 4785 to 4485 cm^{-1} , which was not contributing after SNV normalisation. Water has a strong signal at $\sim 5000\text{ cm}^{-1}$ and wood is hygroscopic. It might be that the SNV normalisation of this variable region affected the neighbouring signal

at 4785-4485 cm^{-1} .

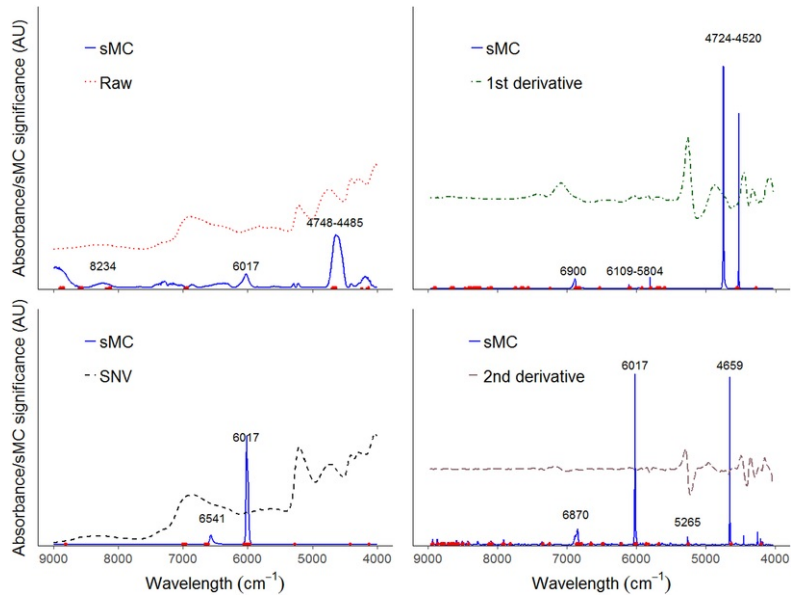


Fig. 3. Influence of pre-processing on NIR spectra, the correlation to EC of *E. bosistoana*, and the variables selected by the sMC algorithm (upper left: raw spectra, upper right: 1st derivative, bottom left: SNV, bottom right: 2nd derivative). Dashed lines: average NIR spectra; blue solid lines: sMC - explained variance; red points: variables selected by the sMC algorithm; SNV: standard normal variate; Raw: spectra without pre-processing. AU: arbitrary unit. [\(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.\)](#)

alt-text: Fig. 3

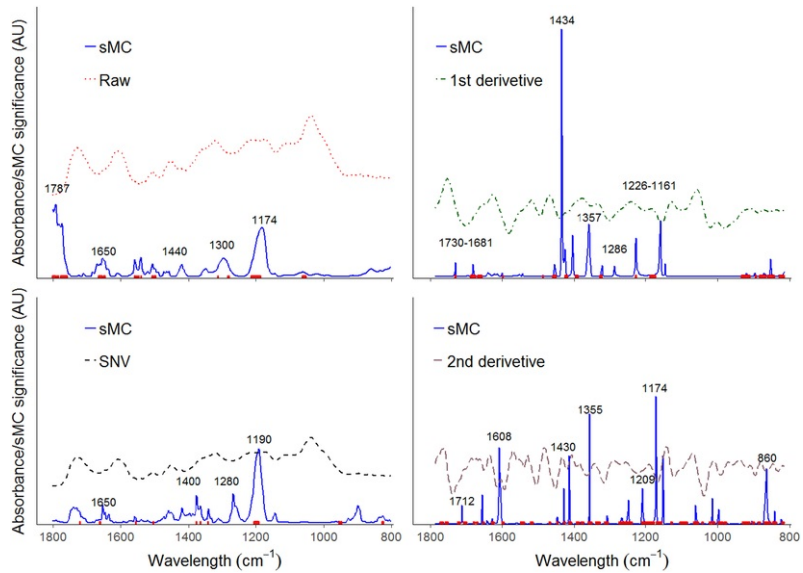


Fig. 4. Influence of pre-processing on ATR-IR spectra, the correlation to EC of *E. bosistoana*, and the variable selected by the sMC algorithm (upper left: raw spectra, upper right: 1st derivative, bottom left: SNV, bottom right: 2nd derivative). Dashed lines: average ATR-IR

spectra; blue solid lines: sMC - explained variance; red points: variables selected by the sMC algorithm; SNV: standard normal variate; Raw: spectra without pre-processing. AU: arbitrary unit. [\(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.\)](#)

alt-text: Fig. 4

Transformation of spectra into their 1st derivative results in a single band being converted into a split peak with extremes at the flanks of the original signal [40]. This was reflected in the explained variance identified with the sMC algorithm. For example, the identified region around 1174 cm⁻¹ in the unmodified and SNV spectra was represented by a double peak at 1226 and 1161 cm⁻¹ or the region 4785–4485 cm⁻¹ as a double peak at 4724 and 4520 cm⁻¹. For NIR spectra, the spectral regions explaining the variance after conversion into the 1st derivative spectra were more similar to the unmodified than the SNV normalised spectra. As expected, the signals appeared sharper as derivatives are a common means of resolving signal overlap.

2nd derivative spectra are often used to resolve signal overlap and positive signals are converted into negative peaks. The spectral regions identified by the sMC mirrored those of the unmodified spectra but with sharper and more numerous peaks.

In general, the signals identified by the PLSR/sMC models included those, which were characteristic for heartwood extract (Fig. 1). These include ~6900, 6017, 4659, 1712, 1330–1300 and ~1174 cm⁻¹, which were associated to the phenolic O

H, C_{ar}-H groups, ketones, esters, conjugated aldehydes and condensed tannins that occur in heartwood extractives.

In the region from 9000 to 7000 cm⁻¹ numerous signals have been reported, for example the 1st overtones and 2nd overtones of O

H and N

H stretching vibrations as well as the 1st overtones of C

H combination bands. For example signals at 8650 to 8450 cm⁻¹ were assigned to the 1st and 2nd overtone of C

H stretching vibrations from methyl groups [29].

For ATR-IR spectroscopy, without any pre-processing method, the significant variables were mainly in the region of 1800 to 1400 cm⁻¹. Several peaks on the ATR-IR spectra were found, including the peaks at 1787, 1650, 1440, 1300, 1174 and 860 cm⁻¹. Peaks at 1787 cm⁻¹ were associated with the 1st overtone of C

H stretching of CH₂ groups and a signal at 1650 cm⁻¹ was reported characteristic of aromatic skeletal C

C stretching in extractives compounds [32]. C

H bending was reported at around 1440 cm⁻¹ while the CH₃ stretching vibrations were associated with peaks at 1317 and 1174 cm⁻¹ [41,42].

3.4.3.4 Variable Selection

Typically, only some frequencies of IR spectra are strongly correlated to the target variable [43]. Not all regions of the spectra contribute to the prediction models and some frequencies might even reduce the precision of the models [6]. Consequently selecting the most relevant variables can improve model quality. Table 3 presents PLSR model characteristics after variable selection with the sMC algorithm [10], one of the numerous algorithms which have been proposed for this purpose [6]. The number of variables was significantly reduced by the sMC algorithm for both NIR and ATR-IR PLSR while the optimal number of components remained the same as for the models based on the full spectra. Considering spectra pre-processing methods, conversion of the spectra into their 1st or 2nd derivative resulted in more selected variables compared to the unmodified or baseline (SNV) corrected spectra. After variable selection, the RMSE for NIR spectra based models was reduced from 0.99 to 0.91 (This change is incorrect. Prefer not to replace - by to) 1.25% to 0.91–1.16% dependent of the spectra pre-processing. The largest reductions in RMSE were observed for the 2nd derivative spectra (Table 2 and 3). No marked improvement of the RMSE for ATR-IR spectra based models was achieved by variable selection, ranging from 1.62 to 1.67% after variable selection compared to 1.67 to 1.71% for the full models.

Table 3 NIR and ATR-IR PLSR regression models based on different pre-processing spectra for calibration and validation of EC in heartwood of *E. bosistoana* with sMC variable selection. R²_{CV}: coefficient of determination using cross-validation. R²_p: coefficient of determination of predictions when the model was applied to the validation data set, respectively; RMSE_{CV}: root-mean-square error of cross-validation. RMSEP: root mean-square error of prediction when the model was applied to the validation data set. SNV: standard normal variate, Raw: spectra without any pre-processing.

alt-text: Table 3

Spectra	Total number of variables	Pre-treatment	Variables selected by sMC	Calibration			Validation	
				R ² _{CV}	RMSE _{CV}	Ncomp	R ² _P	RMSE _P
NIR	1296	Raw (None)	37	0.93	0.91	5	0.82	0.81
	1296	SNV	71	0.92	1.16	3	0.86	1.47
	1282	1 st derivative	137	0.91	1.06	2	0.81	1.58
	1282	2 nd derivative	84	0.91	1.03	3	0.82	0.84
ATR-IR	520	Raw (None)	64	0.72	1.65	3	0.71	1.42
	520	SNV	23	0.72	1.66	2	0.7	1.43
	506	1 st derivative	85	0.72	1.67	3	0.67	1.51
	506	2 nd derivative	166	0.73	1.62	3	0.72	1.40

The models based on the unmodified NIR spectra showed the best results for both, the calibration and validation sets but also needed the most components to reach this level. SNV normalisation reduced the model accuracy both in NIR and ATR-IR spectra (Table 3). The 2nd derivative spectra yielded robust and more accurate prediction results compared to the 1st derivative spectra for both, the NIR and ATR-IR technique.

The chosen variables comprised signals (Figs. 3 and 4), which were related to heartwood extracts (~6900, 6017, 4659, 1730 and 1174 cm⁻¹) (Fig. 1) as well as others (8234 and 6541 cm⁻¹) which might be associated with cell wall polymers. For example, the peaks at 8234 and 6541 cm⁻¹ were associated to the 2nd overtone of the C-H and 1st overtone of the O-H stretching vibrations of cellulose in wood [30,44]. EC is a relative measure in respect to the cell wall material in a sample. Therefore, a calibration needs also to consider the amount of cell wall material present in the spectra or in other words, the spectra need to be normalised for the absolute amount of matter in the light beam.

For the NIR spectra the sMC algorithm selected all three major extractive signals (~6900, 6017 and 4659 cm⁻¹) after 1st and 2nd derivative conversion (Fig. 3). Interestingly only two of the three signals were selected for the unmodified (~6900 and 4659 cm⁻¹) and SNV normalised spectra (~6900 and 6017 cm⁻¹). For ATR-IR spectra, extract specific signals including 1730, 1174 and 860 cm⁻¹ were selected by the sMC algorithm. However, also a relatively large proportion of unidentified signals was selected, potentially also including cell wall signals (Fig. 4).

The density plots of the residuals between measured and predicted EC values of *E. bosistoana* heartwood using the sMC-PLSR NIR and sMC-ATR-IR models are displayed in Fig. 5. The residuals of the validation data sets visualise the accuracy (bias) and precision (deviation around the mean). Both, the NIR and ATR-IR based models gave accurate predictions with no obvious bias. The NIR models were more precise than ATR-IR model having residuals closer to the mean, similar results were found by Zhou et al. [25].

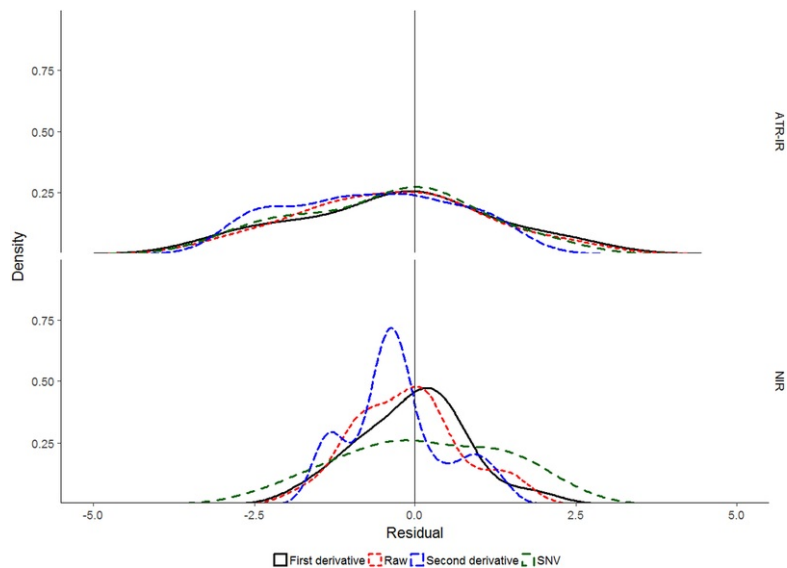


Fig. 5. Density plot of predicted EC (%) residuals (validation data sets) of *E. bosistoana* heartwood by NIR and ATR-IR models with wavenumbers selected by sMC in the validation sets. SNV: standard normal variate, Raw: spectra without any pre-processing.

alt-text: Fig. 5

4.4 Conclusions

NIR and ATR-IR PLSR regression models were able to predict the EC in heartwood of *E. bosistoana* from wood powder samples and can serve as fast and non-destructive methods for tree breeders to screen eucalyptus for heartwood quality. Both, NIR and ATR-IR based models were accurate with no bias but NIR based models were more precise. Variables selection with the sMC algorithm improved the precision (RMSE) of the NIR based models to 0.9%, a useful result considering the EC range of 0.96% to 14.67% in the heartwood of 7 year-old *E. bosistoana*. Variable selection had no benefit on the accuracy of models based on ATR-IR spectra. Spectra pre-processing methods did not offer substantial advantages for both the NIR and ATR-IR spectra. However, 2nd derivative transformation of NIR spectra combined with variable selection reduced the required number of components compared to unmodified spectra. The PLSR models selected extractive signals that related to chemical structures known of heartwood extracts including ~6900 (phenolic O

H groups), 6017 (C_{ar}-H groups), 4659 (C_{ar}-H and C

O groups), 1330-1300 (phenolic O

H groups), and ~1174 cm⁻¹ (condensed tannins).

Acknowledgements

This work was funded by the MBIE Partnership for Specialty Wood Products (Contract FFRX1501). We like to thank Paul Millen for supplying the wood samples.

Conflicts of Interest

We declare that we have no conflicts of interest.

References

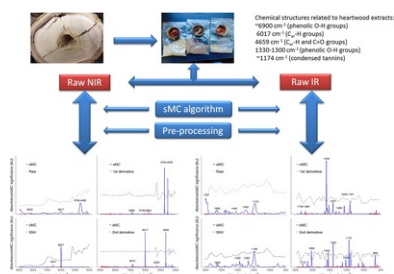
[1] A.K. Moore and N.L. Owen, Infrared spectroscopic studies of solid wood, *Applied Spectroscopy Reviews* *Appl. Spectrosc. Rev.* **36**, 2001, 65-86.

[2] S. Tsuchikawa and M. Schwanninger, A review of recent near-infrared research for wood and paper (Part 2), *Appl. Spectrosc. Rev.* **48**, 2013, 560-587.

- [3] S. Tsuchikawa, A review of recent near infrared research for wood and paper, *Appl. Spectrosc. Rev.* **42**, 2007, 43-71.
- [4] H. Bjorsvik and H. Martens, Data analysis: calibration of NIR instruments by PLS regression, In: *Handbook of Near-Infrared Analysis*, 2nd edn, 2001, Marcell Dekker, Inc; New York.
- [5] Å. Rinnan, F. van den Berg and S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, *Trends Anal. Chem.* **28**, 2009, 1201-1222.
- [6] T. Mehmood, K.H. Liland, L. Snipen and S. Sæbø, A review of variable selection methods in partial least squares regression, *Chemom. Intell. Lab. Syst.* **118**, 2012, 62-69.
- [7] L. Xu, Y.-P. Zhou, L.-J. Tang, H.-L. Wu, J.-H. Jiang, G.-L. Shen and R.-Q. Yu, Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration, *Anal. Chim. Acta* **616**, 2008, 138-143.
- [8] R.W. Kennard and L.A. Stone, Computer Aided design of experiments, *Technometrics* **11**, 1969, 137-148.
- [9] L. Zhang, G. Li, M. Sun, H. Li, Z. Wang, Y. Li and L. Lin, Kennard-Stone combined with least square support vector machine method for noncontact discriminating human blood species, *Infrared Phys. Technol.* **86**, 2017, 116-119.
- [10] T.N. Tran, N.L. Afanador, L.M. Buydens and L. Blanchet, Interpretation of variable importance in partial least squares with significance multivariate correlation (sMC), *Chemom. Intell. Lab. Syst.* **138**, 2014, 153-160.
- [11] K.R. Bootle, Wood in Australia, In: *Types, Properties, and Uses*, 2nd ed., 2005, McGraw-Hill; Australia.
- [12] C.M. Altaner, T.J. Murray and J. Morgenroth, Durable Eucalypts on Drylands: Protecting and Enhancing Value: Workshop Proceedings, 2017, New Zealand School of Forestry, University of Canterbury; NZ.
- [13] J.C.F. Walker, Developing a Eucalypt Resource: Learning From Australia and Elsewhere, 2011, Wood Technology Research Centre; Blenheim, N.Z., 158.
- [14] P. Rudman, Durability in the genus *Eucalyptus*, *Aust. For.* **28**, 1964, 242-257.
- [15] L.F. Hawley, L.C. Fleck and C.A. Richards, The relation between durability and chemical composition in wood, *Ind. Eng. Chem. Res.* **16**, 1924, 699-700.
- [16] A.M. Harju and M. Venäläinen, Measuring the decay resistance of Scots pine heartwood indirectly by the Folin-Ciocalteu assay, *Can. J. For. Res.* **36**, 2006, 1797-1804.
- [17] D. ASTM, Standard Test Method of Accelerated Laboratory Test of Natural Decay Resistance of Woods, 2005, American Society for Testing Materials; West Conshohocken.
- [18] P. Geladi, D. Eriksson and T. Ulvcrona, Data analysis of hyperspectral NIR image mosaics for the quantification of linseed oil impregnation in Scots pine wood, *Wood Sci. Technol.* **48**, 2014, 467-481.
- [19] A. Ribeiro da Silva, T.C. Pastore Monteiro, J.W. Batista Braga, F. Davrieux, E.Y. Arakaki Okino, V.T. Rauber Coradin, J.A. Alves Camargos and A.G.S. Do Prado, Assessment of total phenols and extractives of mahogany wood by near infrared spectroscopy (NIRS), *Holzforchung* **67**, 2013, 1-8.
- [20] P. Maniwaru, K. Nakano, D. Boonyakiat, S. Ohashi, M. Hiroi and T. Tohyama, The use of visible and near infrared spectroscopy for evaluating passion fruit postharvest quality, *J. Food Eng.* **143**, 2014, 33-43.
- [21] R. Meder, S. Gallagher, K.L. Mackie, H. Böhler and R.R. Meglen, Rapid determination of the chemical composition and density of *Pinus radiata* by PLS modelling of transmission and diffuse reflectance FTIR spectra, *Holzforchung* **53**, 1999, 261-266.
- [22] D. Bush, K. McCarthy and R. Meder, Genetic variation of natural durability traits in *Eucalyptus cladocalyx* (sugar gum), *Ann. For. Sci.* **68**, 2011, 1057-1066.
- [23] T. Jones, R. Meder, C. Low, D. O'Callahan, C. Chittenden, N. Ebdon, A. Thumm and M. Riddell, Natural durability of the heartwood of coast redwood [*Sequoia sempervirens* (D. Don) Endl.] and its prediction using near infrared spectroscopy, *J. Near Infrared Spectrosc.* **19**, 2011, 381-389.
- [24] M. Baranska, W. Schütze and H. Schulz, Determination of lycopene and β -carotene content in tomato fruits and related products: comparison of FT-Raman, ATR-IR, and NIR spectroscopy, *Anal. Chem.* **78**, 2006, 8456-8461.
- [25] C. Zhou, W. Jiang, B.K. Via, O. Fasina and G. Han, Prediction of mixed hardwood lignin and carbohydrate content using ATR-FTIR and FT-NIR, *Carbohydr. Polym.* **121**, 2015, 336-341.
- [26] R Core Team, R: A Language and Environment for Statistical Computing, 2017, R Foundation for Statistical Computing; Vienna, Austria.

- [27] A. Stevens and L. Ramirez-Lopez, An introduction to the prospectr package, In: *R Package Vignette*, **3**, 2014, (Report No.: R Package Version 0.1).
- [28] B. Mevik, R. Wehrens and L. Hovde, Partial Least Squares and Principal Component Regression, R package version 2.5-02015.
- [29] M. Schwanninger, J.C. Rodrigues and K. Fackler, A review of band assignments in near infrared spectra of wood and wood components, *J. Near Infrared Spectrosc.* **19**, 2011, 287-308.
- [30] K. Fackler and M. Schwanninger, Polysaccharide Degradaation and Lignin Mmodification during brown Rrot of Sspruce Wwood: Aa Ppolarised Fourier Ttransform near Iinfrared Sstudy, *J. Near Infrared Spectrosc.* **18**, 2010, 403-416.
- [31] A.J. Michell and L.R. Schimleck, NIR spectroscopy of woods from *Eucalyptus globulus*, *Appita J.* **49**, 1996, 23-26.
- [32] O. Faix, Classification of lignins from different botanical origins by FT-IR spectroscopy, *Holzforchung:Holzforchung* **45**, 1991, 21-28.
- [33] M. Schwanninger, J. Rodrigues, H. Pereira and B. Hinterstoisser, Effects of short-time vibratory ball milling on the shape of FT-IR spectra of wood and cellulose, *Vib. Spectrosc.* **36**, 2004, 23-40.
- [34] A. Edelmann and B. Lendl, Toward the optical tongue: Fflow-through sensing of tannin- protein interactions based on FTIR spectroscopy, *J. Am. Chem. Soc.* **124**, 2002, 14741-14747.
- [35] W. Hillis, Formation and properties of some wood extractives, *Phytochemistry:Phytochemistry* **11**, 1972, 1207-1218.
- [36] K.M. Barry, R.B. Pearce and C.M. Mohammed, Properties of reaction zones associated with decay from pruning wounds in plantation-grown *Eucalyptus nitens*, *For. Pathol.* **30**, 2000, 233-245.
- [37] A. Eyles, N.W. Davies and C. Mohammed, Wound wood formation in *Eucalyptus globulus* and *Eucalyptus nitens*: anatomy and chemistry, *Can. J. For. Res.* **33**, 2003, 2331-2339.
- [38] B.K. Via, Prediction of oriented strand board wood strand density by near infrared and Fourier transform infrared reflectance spectroscopy, *J. Near Infrared Spectrosc.* **18**, 2010, 491-498.
- [39] B. Sun, J. Liu, S. Liu and Q. Yang, Application of FT-NIR-DR and FT-IR-ATR spectroscopy to estimate the chemical composition of bamboo (*Neosinocalamus affinis* Keng), *Holzforchung:Holzforchung* **65**, 2011, 689-696.
- [40] L. Salmén and E. Bergström, Cellulose structural arrangement in relation to spectral changes in tensile loading FTIR, *Cellulose:Cellulose* **16**, 2009, 975-982.
- [41] H. Yang, J. Irudayaraj and M.M. Paradkar, Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy, *Food-Chem:Food Chem.* **93**, 2005, 25-32.
- [42] G.E. Acquah, B.K. Via, O.O. Fasina and L.G. Eckhardt, Rapid Qquantitative Aanalysis of forest Bbiomass Uusing Fourier Ftransform Iinfrared Sspectroscopy and Ppartial LLeast Ssquares RRegression, *J. Anal. Methods Chem.* **2016**, 2016.
- [43] F. Westad and H. Martens, Variable selection in near infrared spectroscopy based on significance testing in partial least squares regression, *J. Near Infrared Spectrosc.* **8**, 2000, 117-124.
- [44] R. Jackson, Determination of total hydroxyl content of cellulose esters by near-infrared spectroscopy, *TAPPI J.* **51**, 1968, 560-563.

Graphical Abstract



alt-text: Unlabelled Image

Highlights

- NIR based models performed better (RMSE 0.9%) than ATR-IR based models (RMSE 1.6%).
 - Second derivative spectra combined with sMC reduced the number of components.
 - Signals related to EC chemical structures: ~6900, 6017, 4659, 1330, 1300 and ~1174 cm^{-1} were found.
-

Queries and Answers

Query:

Your article is registered as a regular item and is being processed for inclusion in a regular issue of the journal. If this is NOT correct and your article belongs to a Special Issue/Collection please contact v.david@elsevier.com immediately prior to returning your corrections.

Answer: Yes

Query:

Please confirm that given names and surnames have been identified correctly and are presented in the desired order, and please carefully verify the spelling of all authors' names.

Answer: Clemens M. Altaner

Query:

The author names have been tagged as given names and surnames (surnames are highlighted in teal color). Please confirm if they have been identified correctly.

Answer: Yes

Query:

Please check whether the designated corresponding author is correct, and amend if necessary.

Answer: yes

Query:

Have we correctly interpreted the following funding source(s) and country names you cited in your article: "MBIE Partnership for Specialty Wood Products".

Answer: Yes