

Predicting the cell-wall compositions of *Pinus radiata* (radiata pine) wood using ATR and transmission FTIR spectroscopies

Leona M. Fahey  · Michél K. Nieuwoudt  · Philip J. Harris 

Received: 4 July 2017 / Accepted: 19 September 2017 / Published online: 3 October 2017
© Springer Science+Business Media B.V. 2017

Abstract Because plant cell walls vary in their polysaccharide compositions and lignin contents, their monosaccharide compositions and lignin contents are often determined, but these analyses are time consuming and laborious. We therefore investigated Fourier transform infrared (FTIR) spectroscopy coupled with partial least squares (PLS) regression analysis as a way of rapidly predicting the monosaccharide compositions and lignin contents of the cell walls of compression wood (CW) and opposite wood (OW) of the gymnosperm *Pinus radiata*. The effects were investigated of sample moisture content (ambient or dry) and sample particle size (large particles, < 0.422 mm or small particles, < 0.178 mm) of milled wood on attenuated total reflectance (ATR) and transmission FTIR spectra, as well as the PLS-1 models and subsequent predictions. PLS-1 models were built using mixtures of CW and OW as the

training set, to provide a linear range of monosaccharide compositions and lignin contents. Models were externally validated by predicting another set of wood mixtures before predicting CW and OW of a separate test set. Most of the monosaccharide amounts in the separate test set were best predicted by ATR spectroscopy of ambient large particles, achieving the lowest standard error values for the monosaccharides arabinose (0.36%), xylose (1.05%), galactose (1.79%), glucose (6.32%), and 4-*O*-methylglucuronic acid (0.20%). The results show the feasibility of using ATR spectroscopy of ambient large particles for the rapid prediction of monosaccharide compositions and lignin contents of plant cell walls.

Keywords *Pinus radiata* (radiata pine) · Monosaccharide composition · Lignin · Attenuated total reflectance Fourier transform infrared (ATR FTIR) spectroscopy · Transmission Fourier transform infrared (transmission FTIR) spectroscopy · Partial least squares (PLS) regression

Electronic supplementary material The online version of this article (doi:10.1007/s10570-017-1506-4) contains supplementary material, which is available to authorized users.

L. M. Fahey · P. J. Harris (✉)
School of Biological Sciences, The University of
Auckland, Private Bag 92019, Auckland 1142,
New Zealand
e-mail: p.harris@auckland.ac.nz

M. K. Nieuwoudt
School of Chemical Sciences, The University of
Auckland, Private Bag 92019, Auckland 1142,
New Zealand

Abbreviations

ATR FTIR	Attenuated total reflectance Fourier transform infrared
CW	Compression wood
NIPALS	Non-linear iterative partial least squares
OW	Opposite wood
PLS	Partial least squares

RMSECV	Root mean squared error of cross validation
RMSEP	Root mean squared error of prediction

Introduction

Plant cell walls are of considerable economic importance. They form the basis of the lumber, paper and pulp industries, and underpin natural-textile industries. The cell walls of food plants are a major part of dietary fibre in human nutrition, and those of forage plants form the fibre in ruminant nutrition. Furthermore, new industries based on plant cell walls have emerged: plant biomass containing high proportions of plant cell walls is the feedstock for the production of second-generation biofuels (Pauly and Keegstra 2008) and a source of raw material for biocomposites (John and Thomas 2008).

Plant cell walls are composed of cellulose microfibrils embedded in a matrix containing mostly polysaccharides, but which may also contain phenolic compounds, including lignin, as well as structural proteins and glycoproteins (Harris 2005; Harris and Stone 2008). There are two major cell wall types: primary walls that are deposited while cells are still enlarging, and secondary walls that are deposited on the primary walls when cell expansion is completed. Primary and secondary cell walls often have different compositions. At maturity, some cell types, such as parenchyma, often have only a non-lignified primary cell wall, whereas others, such as sclerenchyma fibres, have both a primary and a secondary wall layer, both of which are lignified. Non-lignified primary walls predominate in plant foods for human consumption, but for all other applications, secondary walls usually predominate, and are usually lignified.

The types and proportions of the matrix polysaccharides vary with cell type, developmental stage, and plant taxon (Harris 2005; Harris and Stone 2008). For example, the primary walls of parenchyma cells of eudicotyledons (e.g. apples), many monocotyledons (non-commelinid monocotyledons, e.g. onions) and coniferous gymnosperms (e.g. pines) contain high proportions of pectic polysaccharides, and lower proportions of xyloglucans, whereas those of the grasses, including the cereals, (within the commelinid

monocotyledons) contain mostly heteroxylans (glucuronoarabinoxylans) (GAXs), with smaller proportions of pectic polysaccharides and xyloglucans, and variable proportions of (1 → 3), (1 → 4)-β-D-glucans. The lignified secondary walls of these taxa usually contain mostly heteroxylans, and smaller proportions of heteromannans. However, the lignified secondary walls of coniferous gymnosperms contain mostly heteromannans (galactoglucomannans), with smaller proportions of heteroxylans [arabino(4-O-methyl-glucurono)xylans] (Harris and Stone 2008; Scheller and Ulvskov 2010). In addition to the polysaccharides, the presence and concentration of lignin in plant cell walls is of structural importance to the plant, but is critical in the optimum utilisation of plant cell walls as forage for ruminants where the lignin affects the digestibility of plant material, and as feedstock for bio-industries where delignification is often a necessary pre-processing treatment (Himmel et al. 2007).

Because plant cell walls vary in their polysaccharide compositions, an initial approach in analysing these is to determine their monosaccharide compositions, by acid hydrolysing them to release the component monosaccharides, which are then separated and quantified. The monosaccharides are often derivatized, for example to alditol acetates, and then separated by capillary gas chromatography (Blakeney et al. 1983), but more recently, they have frequently been separated without derivatization using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) (Brennan et al. 2012; Currie and Perry 2006). A two-stage sulfuric acid hydrolysis method is commonly used as it hydrolyses all of the cell-wall polysaccharides, including cellulose (Saeman et al. 1963). This method has the additional advantage that lignin forms an insoluble residue and can be dried and weighed; this is referred to as Klason lignin (Dence 1992). More recently, a faster method of determining lignin has been introduced in which lignin is dissolved in acetyl bromide and quantified by ultraviolet (UV) spectrophotometry (Dence 1992; Fukushima and Hatfield 2004).

These methods are time consuming, labour intensive, and not well suited to large scale screening of samples. An alternative, more rapid method, is Fourier transform infrared (FTIR) spectroscopy in the mid-infrared region, which is now extensively used in plant

cell wall analysis and is a powerful high-throughput analytical technique. FTIR spectroscopy has been used qualitatively to compare the compositions of plant cell walls, and quantitatively, when coupled with multivariate data analyses such as partial least squares (PLS) regression or principal component regression (PCR), to predict the monosaccharide compositions and lignin contents (Baker et al. 2014; Kačuráková and Wilson 2001; Lupoi et al. 2013; Xu et al. 2013). PLS regression is the more frequently chosen method for prediction purposes, although PCR may be the better method for model interpretation (Hair et al. 2010). The amounts of IR radiation absorbed or transmitted by samples are influenced by the way the sample is prepared, for example, the moisture content of the samples, the particle size, and the method used to mill the samples, all of which can affect the intensities of the FTIR spectra. With transmission FTIR spectroscopy, it also depends on the amount of sample analysed. Traditionally, transmission spectra are recorded on dry samples in a potassium bromide (KBr) matrix, which is very hygroscopic, and can result in variable intensities for water bands that may overlap bands of interest. In obtaining transmission spectra of milled wood using KBr as a matrix, the intensities of major bands have been found to decrease as the particle size increased (Faix and Böttcher 1992; Schwanninger et al. 2004), and increasing the amount of sample relative to that of KBr was found to lower the relative band intensities (Faix and Böttcher 1992). In addition, the shapes of the bands in the IR spectra of wood have been found to alter when ball-milling is used as the method of reducing particle size (Schwanninger et al. 2004).

Attenuated total reflectance (ATR) FTIR spectroscopy is an important alternative to transmission spectroscopy, partly because it removes the need for time-consuming preparation of the sample in a KBr matrix, but also because it allows the analysis of both powdered and solid samples. In ATR spectroscopy, instead of the IR radiation passing through the sample, the sample is analysed using a small evanescent wave arising from the total internal reflection on a crystal and protruding $\sim 4\text{--}6\ \mu\text{m}$ above the crystal directly into the sample. However, the success of this method depends on sufficient contact between the sample and the ATR crystal; the degree of compaction and particle size would be expected to have an effect on the degree of contact with the crystal. As far as the authors are

aware, no systematic study has been published on the effects of the extent of sample moisture content or particle size on the accuracy of ATR predictions of the monosaccharide compositions and lignin contents of milled, lignified biomass, such as woods, with PLS.

In the present study, monosaccharide compositions and lignin contents of wood cell walls of the coniferous gymnosperm *Pinus radiata* (radiata pine) were determined. *P. radiata* is a major forestry plantation species (Walker 2013), and is used for timber, pulp and paper, as well as for the production of second-generation biofuels (Araque et al. 2008). The present study included an analysis of compression wood (CW) from *P. radiata*. CW is a reaction wood formed on the underside of coniferous gymnosperm stems tilted from the vertical, for example by wind, snow, or slope, and is formed to correct this orientation. CW cell walls contain more lignin and less cellulose than those of opposite wood (OW), which is formed on the opposite side of the stem and is anatomically and compositionally similar to normal wood (Brennan et al. 2012; Timell 1986). In addition, CW cell walls contain significant proportions of the polysaccharides (1 \rightarrow 4)- β -galactans and (1 \rightarrow 3)- β -glucans (Brennan et al. 2012; Timell 1986; Zhang et al. 2016). OW also contains (1 \rightarrow 4)- β -galactans, but in much lower proportions (Chavan et al. 2015). The differences in compositions between CW and OW cell walls of *P. radiata* allow ATR spectroscopy to successfully discriminate the two wood types, and predict the galactosyl and lignin content of both (McLean et al. 2014).

As part of our study, we exploited the known compositional differences between CW and OW of *P. radiata*, to provide mixed wood samples that covered a wide, linear range of monosaccharide compositions and lignin contents. PLS-1 models were built using both ATR and transmission spectra collected from these samples, and then externally validated by predicting the monosaccharide compositions and lignin contents of an external validation set, that was a similar mixed wood sample set. The externally validated PLS-1 models were then further applied to predict the cell-wall compositions of CWs and OWs of a separate test set, grown in a semi-controlled outdoor environment. In this study, we also examined the effects of these factors on the accuracy of the prediction of all monosaccharides and lignin contents: the effect of moisture content by using wood samples

that were either equilibrated to laboratory atmosphere (ambient moisture content) or dry, the effect of particle size by using 40 (large) and 80 (small) mesh sized particles, and the effect of sampling methods by collecting both ATR and transmission spectra.

Materials and method

Wood material

Three sets of wood samples from *Pinus radiata* D. Don were used: a training set, an external validation set, and a separate test set. The training and external validation sets were from 8-month-old sapling trees of two clones [Clone A or 96047/98015 (training set) and Clone K or 95044 (external validation set) from ArborGen Australasia, TeTeko, Whakatane, New Zealand], grown from seedlings in rigid plastic tubes in an unheated glasshouse at the University of Canterbury, Christchurch, New Zealand (Apiolaza et al. 2011; Brennan et al. 2012). They were tilted by staking at $\sim 45^\circ$ from the vertical to induce the formation of CW. The separate test set was from 21-month-old saplings grown in a semi-controlled outdoor environment at Harewood, Christchurch, New Zealand. The saplings (20 clones, one ramet per clone from Forest Genetics Ltd., Rotorua, New Zealand) were 9-months-old when planted, grown straight for 3 months before being tilted by staking at $\sim 30^\circ$ from the vertical. They were grown in planter bags with potting mix containing slow-release fertilizer and were drip irrigated.

After harvesting the trees, one segment (~ 15 cm long) for the training and external validation sets and one segment (~ 10 cm long) for the separate test set were cut from each stem above the soil line, debarked and dried at 35°C to constant weight. The CW of the training and external validation sets was identified visually by its darker colour and separated from the OW after drying, taking care to avoid the pith and side wood (between the CW and OW). For the separate test set, the CW and OW were separated by sawing the segment lengthways before drying. Longitudinal, tangential strips (~ 2 mm thick) were cut from the outermost growth ring of CW and OW samples using a band saw. The outermost strip was excluded from the study to avoid the phloem and cambium.

The CW and OW from each ramet were milled separately using a Wiley[®] mini-mill (Thomas

Scientific, Swedesboro, NJ, USA) fitted with a 40 mesh screen (0.422 mm pore size; large particles), extracted with dichloromethane (DCM) in a Soxhlet extractor overnight at 60°C (TAPPI T 264 cm-97 1997), dried to constant weight and stored at 30°C under vacuum over phosphorus pentoxide (P_2O_5). Mixtures of milled, extracted and dried CW and OW were made for analysis. The training (T) and external validation (V) sets were each divided into two equal masses to give two technical replicates of each set (T1 and T2; large particles, and V1 and V2; large particles). The training and external validation sets were mixtures obtained by mixing CW and OW in 10 and 20% dry mass increments, 11 and 6 samples, respectively. A portion of the large particle size training (T1) and external validation (V1) wood mixture sets, and of the separate test set were further ground to a smaller particle size. Grinding of each sample was carried out for ~ 10 min using a ceramic pestle and mortar containing liquid nitrogen. After equilibrating overnight at 30°C , the ground wood was sieved on an 80 mesh screen (0.178 mm pore size; small particles) on a vortex mixer. Particles that failed to pass through the screen were re-ground in the same manner until $< 1\%$ of the initial sample remained. The small particles were dried and stored as indicated above.

To investigate the effect of moisture on ATR spectra, the samples (large and small particles) were examined both dry and with ambient moisture content. “Ambient” samples were equilibrated to laboratory atmosphere moisture content (6–8% water content) to constant weight, and “dry” samples were dried at 30°C under vacuum over P_2O_5 to constant weight and stored under the same conditions. Transmission spectra were recorded only using the dry samples to avoid clouding of the hygroscopic KBr pellets.

For PLS analysis, four separate sets of ATR spectra (ambient and dry, large particles; ambient and dry small particles) were used for each of the training, external validation and separate test sets, and two separate sets of transmission spectra (dry, large and small particles) for each of the above sample sets.

Determination of monosaccharide compositions and Klason lignin contents

All three sample sets with large particles were used. For the training and external validation sets, both

technical replicates were analysed of the 11 and 6 samples, respectively. Particles were dried (16 h, 105 °C) and two replicates (100 mg each) hydrolysed using a scaled down two-stage sulfuric acid hydrolysis method in the manner of TAPPI standards T 249 cm-85 (1985) and T 222 om-98 (1998). Sulfuric acid (1 ml, 72% w/w) was added, the suspension stirred every 15 min for 1 h at 30 °C, water (28 ml) added, and autoclaved (121 °C, 1 h, 103.4 kPa). The hydrolysates were filtered under vacuum through sintered glass filter crucibles (porosity 4) (ROBU[®] Glasfilter-Geraet GmbH, Hattert, Germany) of accurately known weights. Monosaccharide analysis was done on the filtrate after further diluting ($\times 1250$) with water. The residue was washed, and dried (at 105 °C) to constant weight to give the Klason lignin content.

Monosaccharides were separated and quantified using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) on a Dionex BioLC system (Dionex, Sunnyvale, CA, USA) with an ED50 electrochemical detector and GP50 gradient pump. A CarboPac PA20 guard column (3 \times 30 mm) and a CarboPac PA20 analytical column (3 \times 150 mm) were kept at 25 °C with a TCC-100 thermostatted column compartment. Separations were with isocratic elutions: 1 mM NaOH for 35 min to separate the neutral monosaccharides, followed by 200 mM NaOAc in 200 mM NaOH for 10 min to separate the uronic acids. The column was washed with 200 mM NaOH for 5 min and equilibrated for a further 5 min with 1 mM NaOH. The injection volume was 20 μ l with a flow rate of 0.4 ml/min.

The retention times of the monosaccharides, including the internal standard 2-deoxy-D-galactose, were determined by running solutions of individual monosaccharides and were: 2-deoxy-D-galactose (~ 5 min), L-arabinose (Ara, ~ 9 min), D-galactose (Gal, ~ 11 min), D-glucose (Glc, ~ 13 min), D-xylose (Xyl, ~ 15 min), D-mannose (Man, ~ 16 min), D-galacturonic acid (GalA, ~ 36.2 min), and D-glucuronic acid (GlcA, ~ 36.6 min). The retention time of 4-O-methyl-D-glucuronic acid (4OMeGlcA, ~ 35.9 min) was determined by chromatography of a 2 M trifluoroacetic acid hydrolysate of 4-O-methylglucuronoxylan from birchwood (Sigma-Aldrich, St. Louis, MO, USA) and the extracellular polysaccharide of *Rhizobium japonicum* strain 71 A (Smith and Harris 1995). The detector response of 4OMeGlcA was assumed to be the same as that of

GlcA. Calibration mixtures of these monosaccharides (2.0–300.0 μ g ml⁻¹) were analysed using the same elution profile to determine the PAD response for different amounts. The response from the PAD for the monosaccharides was linear across this range and analyses of wood hydrolysates was carried out within this range.

Prior to running hydrolysate samples, a water blank was run, followed by an external standard solution containing the same monosaccharides at concentrations approximating a softwood hydrolysate. This contained the internal standard 2-deoxy-D-galactose (10.0 μ g ml⁻¹), L-Ara (5.0 μ g ml⁻¹), D-Gal (20.0 μ g ml⁻¹), D-Glc (150.0 μ g ml⁻¹), D-Xyl (30.0 μ g ml⁻¹), D-Man (30.0 μ g ml⁻¹), D-GalA (5.0 μ g ml⁻¹) and D-GlcA (5.0 μ g ml⁻¹). Each hydrolysate sample was run in duplicate. The monosaccharide composition was determined using the response factor for each monosaccharide relative to 2-deoxy-D-galactose, used as the internal standard.

ATR FTIR spectroscopy

ATR FTIR spectra of wood samples from each sample set were recorded for each of the four sample conditions (both ambient and dry, large particles; both ambient and dry, small particles). The spectra were collected using a Thermo Electron Nicolet[™] 8700 FTIR spectrometer (Thermo Fisher Scientific, Waltham, MA., USA) equipped with the Smart Orbit ATR accessory and diamond crystal of ~ 1.5 mm sampling area, with single bounce at 45°. The spectroscopic software Omnic version 7.3 was used to record and process the spectra. The spectrometer included a Globar IR source, KBr beam splitter and DTGS TEC (deuterated triglycine sulphide, thermoelectrically cooled) detector. A background spectrum was collected before collecting replicate spectra of three portions of the training (T1 and T2; large particles, and T1; small particles) and external validation (V1 and V2; large particles, and V1; small particles) sets, and each separate test set sample (large and small particles). Each spectrum was an average of 64 scans over the range 4000–525 cm⁻¹ at a spectral resolution of 8 cm⁻¹.

The spectra were ATR corrected, truncated to include only the fingerprint region (1800–725 cm⁻¹), followed by a 2-point linear baseline correction at 1800 and 725 cm⁻¹. All spectra from a given spectral

data set were entered as a single data matrix into The Unscrambler® X version 10.3 (Camo Software AS., Oslo, Norway) software package for PLS regression analysis. All spectra were normalised to the strongest absorption band $\sim 1026 \text{ cm}^{-1}$. Replicate spectra were averaged for each of the training and external validation sets (large particles, six spectra; small particles, three spectra) and separate test set (large particles, three spectra; small particles, three spectra).

Transmission FTIR spectroscopy

Transmission FTIR spectra were recorded on only the dry large and small particles of the training (T1), external validation (V1), and separate test sets. These were dried at $105 \text{ }^\circ\text{C}$ for 16 h, alongside the spectroscopic grade KBr powder (Sigma Aldrich, St. Louis, MO, USA) used for preparing the KBr pellets. Each sample ($\sim 2 \text{ mg}$) was thoroughly mixed with dried KBr powder ($\sim 200 \text{ mg}$) and pressed into a pellet using a 13 mm stainless steel die set (Specac Ltd, Orpington, Kent, England) under vacuum and $8 \times 10^3 \text{ kg}$ of pressure for 5 min. One spectrum was recorded in transmission mode for each pellet using a PerkinElmer Spectrum 400 FT-IR spectrometer (PerkinElmer, Waltham, MA., USA), with SpectrumOne spectroscopic software. A background spectrum of a pure KBr pellet was recorded before the sample spectrum. Each spectrum was an average of 4 scans over the range $4000\text{--}400 \text{ cm}^{-1}$ at a spectral resolution of 4 cm^{-1} .

The transmission spectra were truncated and baseline corrected as described above for the ATR spectra. All spectra of a given spectral data set were entered as separate single data matrix into the Unscrambler® X software and normalised to the $-\text{CH}_3$ deformation mode at 1375 cm^{-1} as this band did not differ significantly among wood types.

PLS-1 regression analysis

Nine PLS-1 models, where variables were modelled separately, were constructed for each of the four spectral data sets. Reference to PLS in the present study can be assumed to mean PLS-1. The Unscrambler® X software was used to perform PLS regression analysis on the four different ATR spectral data sets (ambient and dry, large particles; ambient and dry, small particles) and two different transmission

spectral data sets (dry, large and small particles). PLS calibration models were constructed using the non-linear iterative partial least squares (NIPALS) algorithm for each of the eight monosaccharides and lignin. For PLS analysis of the ATR spectral data sets, the matrix was divided into 11 training set samples, 6 external validation set samples, and 40 separate test set samples. For PLS analysis using transmission spectral data, the matrix was divided into 11 training set samples, 6 external validation set samples, and 40 separate test set samples. All data were mean centred for the PLS regression analyses and each calibration model was internally validated using full cross-validation, in which one sample is systematically left out from the calibration set and predicted by a sub-model of the remaining samples, thus using the same number of calibration sub-models as there were samples to predict the left-out samples.

The PLS calibration models were then externally validated by predicting the cell-wall composition of the external validation set. The calibration models were then subjected to a further and more rigorous test by predicting the cell-wall compositions of the separate test set.

The PLS models were selected in the following way. Each PLS model had a maximum of seven factors (principal components) calculated and the optimum number of factors to be considered was chosen by visually inspecting the explained variance for both the calibration and internal cross validation. The highest factor to be considered optimum was the factor beyond which the internal cross validation began to explain less of the variance for the model than the previous factor (i.e. when the explained variance of the internal cross-validation began to decrease).

Once the highest possible number of factors for consideration for each model was decided, the standard error (SE) of the predicted monosaccharide and lignin contents of the external validation and separate test sets were calculated using the equation

$$SE = \sqrt{\frac{\sum(x - y)^2 - (\sum(y - x))^2/n}{n - 1}}$$

where x = predicted values, y = reference values, and n = number of samples.

The coefficient of determination (R^2) of the calibration model, the root mean square error of cross validation (RMSECV) and the root mean square error

of prediction (RMSEP) were calculated by The Unscrambler® X. The SE of the separate test set samples was used to select the optimum number of factors for each model.

Results

Monosaccharide compositions and lignin contents

For the mixed CW and OW samples in the training and external validation sets, there was a strong linear relationship between the increasing amounts of CW and the individual monosaccharides and lignin contents, as reflected by the high R^2 values in Table 1 and Online Resource 1. The minimum and maximum contents for individual monosaccharides and the lignin content of mixed CW and OW samples in the training and external validation sets, and the lowest and highest values for individual monosaccharides and lignin in the pure CWs and OWs in the separate test set, are shown in Table 1. The lowest and highest values obtained in the separate test set represent the range of wood cell-wall compositions in the pure CWs and OWs of the 20 different *P. radiata* clones. Ideally, the composition in the training set should span the

range of compositions to be predicted, but as shown in Table 1, the separate test set had a slightly broader range of compositions.

ATR and difference spectra of pure CW and OW

Representative ATR spectra of dry, small particles from the training set pure CW and OW are shown in Fig. 1a. These spectra show that some of the bands vary in relative intensity, which reflects differences in cell-wall compositions. To show the differences between the spectra of the two wood types more clearly, a mathematical difference spectrum was calculated by subtracting the mean OW spectrum from the mean CW spectrum (Fig. 1b). Positive bands above zero indicate components in greater proportions in CW and negative bands below zero indicate components in greater proportions in OW (Fig. 1b). The higher proportion of lignin in compression wood than in opposite wood is reflected by the majority of positive bands above zero in the difference spectra being assigned to lignin. The higher proportion of polysaccharides in opposite wood than in compression wood is reflected by the negative bands below zero in the difference spectra being assigned to polysaccharides. Bands that vary in intensity in representative

Table 1 Minimum and maximum^a wood cell-wall composition values of mixed *Pinus radiata* samples in the training and external validation sets, and the linear relationship (R^2) with the percentage of compression wood in each mixed sample. Lowest and highest^a values of the range of individual monosaccharides and lignin contents of CWs and OWs of *P. radiata* in the separate test set

	Ara	Xyl	Gal	Glc	Man	GalA	GlcA	4OMeGlcA	Lignin
Training set									
Minimum	1.4	4.8	2.1	34.4	6.0	0.5	0.1	0.6	25.5
Maximum	2.2	7.6	13.7	45.4	9.0	1.1	0.2	1.2	37.1
R^2	0.94	0.95	0.99	0.90	0.96	0.71	0.20	0.95	0.98
External validation set									
Minimum	1.3	3.9	1.8	29.1	5.2	0.6	0.1	0.5	24.3
Maximum	2.4	8.5	12.4	46.9	9.5	1.2	0.2	1.4	37.3
R^2	0.92	0.95	1.00	0.94	0.91	0.86	0.29	0.99	1.00
Separate test set									
Compression wood (CW)									
Lowest	1.0	3.2	9.1	17.7	3.7	0.3	0.1	0.4	33.4
Highest	1.9	6.2	16.8	44.0	8.9	0.8	0.3	1.0	38.8
Range	0.9	3.0	7.7	26.3	5.2	0.5	0.2	0.6	5.4
Opposite wood (OW)									
Lowest	1.7	6.0	1.3	30.3	6.2	0.4	0.1	1.0	22.1
Highest	3.1	11.3	2.6	56.7	12.2	1.4	0.2	1.9	28.8
Range	1.4	5.3	1.3	26.4	6.0	1.0	0.1	0.9	6.7

^aExpressed as mean percentage of oven-dry weight of wood

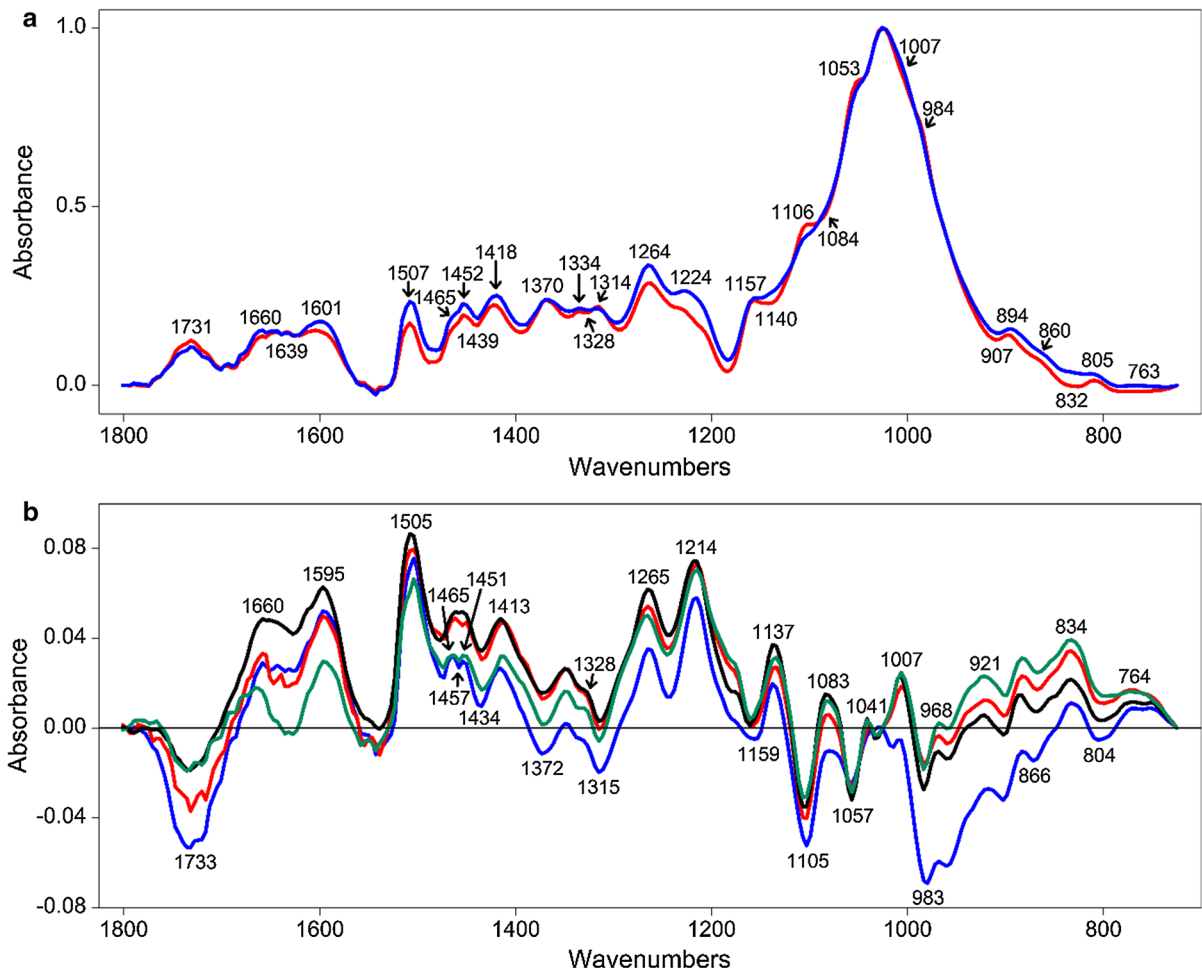


Fig. 1 (a) Representative ATR FTIR spectra of dry, small particles from the training set pure CW (blue) and OW (red) of *Pinus radiata*, and (b) ATR FTIR difference spectra of large

particles, ambient (blue) and dry (red), and small particles, ambient (black) and dry (green), from the training set pure CW and OW of *P. radiata*

ATR spectra (Fig. 1a), and positive and negative bands in the difference spectra (Fig. 1b), along with their possible assignments, are listed in Table 2. To show any differences due to moisture content (ambient or dry) or particle size (large or small), difference spectra were also calculated for the training set CW and OW of both ambient and dry, large particles, and ambient, small particles (Fig. 1b). Overall, the shape of the difference spectra of the four sample conditions were similar, but with bands that vary in intensity. However, for the difference spectrum of the ambient, large particles, there was a marked difference from the other samples between 834 and 1050 cm^{-1} (Fig. 1b).

Effect of moisture content and particle size on ATR and transmission spectra

ATR spectra of the training set CW and OW are shown in Fig. 2a–d. To show the effect of sample moisture content on spectra, spectra of large (Fig. 2a) and small (Fig. 2b) particles in ambient and dry conditions were overlaid. The effect of moisture on the spectra was not consistent as neither ambient nor dry samples, in large or small particle size, consistently had the strongest band intensities. However, differences in band intensities were found throughout the spectra, particularly in the region 1530–1700 cm^{-1} , where adsorbed water is assigned to 1639 cm^{-1} , and between 725 and 1025 cm^{-1} , which has bands assigned to both

Table 2 Band assignments corresponding to spectral differences between compression and opposite woods of *Pinus radiata* in ATR FTIR spectra, and to peaks (+) above and troughs (–) below zero (0.00) in the ATR FTIR difference spectra

Wavenumber (cm ⁻¹)	Band origin	Possible assignments	References
1731/1733 (–)	C=O stretching in acetyl groups and carboxylic acids	Acetyl groups on heteromannans and heteroxylans	Harrington et al. (1964) and Marchessault (1962)
1660/1660 (+)	C=O stretching	Lignin	Faix (1991)
1639/–	O–H bending	Adsorbed water	Harrington et al. (1964) and Marchessault (1962)
1601/1595 (+)	Aromatic skeletal vibrations	Lignin	Faix (1991), Harrington et al. (1964) and Marchessault (1962)
1507/1505 (+)	Aromatic skeletal vibrations	Lignin	Faix (1991), Harrington et al. (1964) and Marchessault (1962)
1465/1465 (+)	CH ₂ deformation vibration and benzene vibration	Lignin and heteroxylans	Harrington et al. (1964) and Marchessault and Liang (1962)
–/1457	CH ₂ deformation vibration and benzene vibration	Lignin and heteroxylans	Harrington et al. (1964) and Marchessault and Liang (1962)
1452/1451(+)	OH in-plane bending modes	Polysaccharides	Liang and Marchessault (1959)
1439/1434	Aromatic skeletal vibrations and CH ₂ bending	Lignin and polysaccharides	Hergert (1971) and Liang and Marchessault (1959)
1418/1413 (+)	CH ₃ bending vibration in lignin, and CH ₂ scissor vibrations in cellulose	Lignin and cellulose	Fengel and Ludwig (1991) and Harrington et al. (1964)
1370/1372	CH bending vibration of CH ₃	Polysaccharides	Liang and Marchessault (1959)
1334/–	OH in-plane bending and CH ₂ wagging	Polysaccharides	Liang and Marchessault (1959) and Marchessault and Liang (1962)
1328/1328 (+)	Phenolic OH and G-ring condensed	Lignin	Faix (1991)
1314/1315	CH ₂ wagging	Cellulose	Liang and Marchessault (1959)
1264/1265 (+)	C–O stretching vibration	Lignin	Harrington et al. (1964) and Marchessault (1962)
1224/–	C–C, C–O, C=O stretching	Lignin	Faix (1991) and Harrington et al. (1964)
–/1214 (+)	C–O stretching, aromatic (phenol)	Lignin	Hergert (1960)
1157/1159	C–O–C antisymmetric bridge stretching vibration	Polysaccharides	Liang and Marchessault (1959) and Marchessault and Liang (1962)
1140/1137 (+)	Aromatic C–H in-plane deformation, secondary alcohols, and C=O stretch	Lignin	Faix (1991)
1106/1105 (–)	O–H association band	Polysaccharides	Higgins et al. (1961)
1084/1083 (+)	C–O deformation in secondary alcohols and aliphatic ethers	Lignin	Faix (1991) and Hergert (1960)
1053/1057 (–)	C–O stretching vibration	Polysaccharides	Higgins et al. (1961) and Liang and Marchessault (1959)
–/1041	C–O stretching	Lignin and polysaccharides	Hergert (1960) and Marchessault and Liang (1962)
1007/1007	C–O stretching	Polysaccharides	Liang and Marchessault (1959)
984/983 (–)	C–O stretching	Polysaccharides	Higgins et al. (1961) and Liang and Marchessault (1959)
–/968	–HC=CH out of plane deformation	Lignin	Faix (1991)
–/921	C–H out of plane; aromatic	Lignin	Faix (1991)
907/–	C–H deformation	Polysaccharides	Kačuráková et al. (2000) and Müller et al. (2009)

Table 2 continued

Wavenumber (cm ⁻¹)	Band origin	Possible assignments	References
894/–	Anomeric vibration at β -glycosidic linkage	Polysaccharides	Higgins et al. (1961) and Marchessault and Liang (1962)
–/866	Deformation of the equatorial C2–H bond of mannosyl residues	Heteromannans	Åkerholm and Salmén (2001), Kato et al. (1973) and Marchessault (1962)
860/–	C–H out of plane in position 2,5, and 6	Lignin G-units	Faix (1991) and Hergert (1960)
832/834 (+)	C–H out of plane in all positions	Lignin H-units	Faix (1991)
805/804	In-phase ring stretching of mannosyl residues	Heteromannans	Åkerholm and Salmén (2001), Kato et al. (1973) and Marchessault (1962)
763/764 (+)	Aromatic C–H out-of-plane deformation	Lignin G-units and β -galactosyl residues	Hergert (1971), Marchessault (1962) and Mathlouthi and Koenig (1987)

polysaccharides and lignin (Table 2). To show the effect of particle size on spectra, spectra of ambient (Fig. 2c) and dry (Fig. 2d) samples of large and small particles were overlaid. This effect was more apparent than the effect of moisture, as large particles in both ambient and dry conditions had the strongest band intensities throughout the spectra.

Transmission spectra were obtained of the dry, large and small particles of the training set CW and OW to examine possible effects of particle size on the spectra, and are shown in Fig. 2e. Band intensities in the region 1374–1800 cm⁻¹ were higher for the larger particles, whereas band intensities in the region 965–1290 cm⁻¹ were higher for smaller particles, and the overlapping components were better resolved.

PLS analysis

The R² and RMSECV values of each PLS model for the monosaccharide compositions and lignin contents of the training set are shown in Table 3. The RMSEP and SE values for predictions of the external validation and separate test sets compositions are shown in Tables 3 and 4, respectively. A good PLS model is indicated by a high R² value and low RMSECV (from internal cross validation), RMSEP and SE values (from predictions of other sample sets). The factor chosen for each PLS model considered this and was selected based on the SE value of the separate test set (Tables 3 and 4). The internal cross validation for the model built for the minor component, GlcA, using ATR spectra of ambient, large particles was negative at factor 1, and therefore no model or predictions were

made for this monosaccharide. The R² values for all models of GlcA were consistently low, whereas the RMSEP and SE values for Glc were consistently high. Overall, the effects of moisture content, particle size and sampling method on the R² and RMSECV values of the training set, and RMSEP and SE values of the external validation and separate test sets were small, and in some cases, there was no effect. The following sections describe the effects of moisture content, particle size and sampling method on PLS models built using the training set, and predictions of the external validation and separate test sets. Differences in the R² (> 0.2), RMSECV (> 1%), RMSEP (> 2%) and SE (> 2%) values are highlighted to indicate the largest effects of moisture content, particle size and sampling method. The ranges of RMSEP and SE values (< 1%) for the external validation and separate test sets are also given.

Effect of moisture content

The moisture content of large or small particles had only a small effect on most PLS models built using ATR spectra of ambient or dry, large particles, and ambient or dry, small particles of the training set as similar R² and RMSECV values were obtained (Table 3). However, the R² value of the model built using dry, large particles for the minor component, GalA, was 0.22 lower.

Predictions of the individual monosaccharide and lignin contents of the external validation sets using ATR spectra of ambient or dry, large particles, or ambient or dry, small particles are given in Table 3,

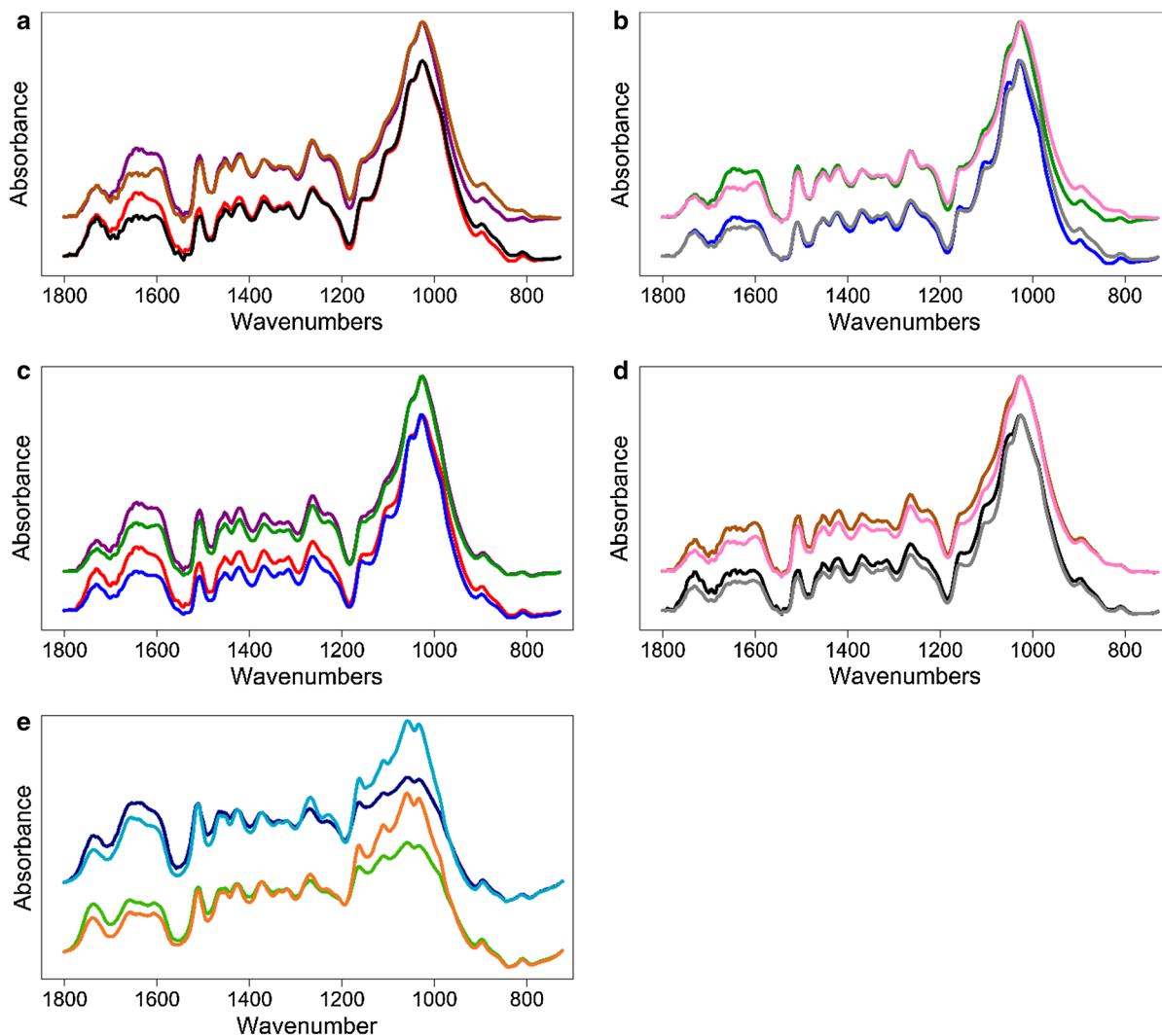


Fig. 2 Overlaid ATR FTIR spectra of compression wood (CW) (upper two spectra) and opposite wood (OW) (lower two spectra) of *Pinus radiata* comparing the spectral differences of (a) large particles of ambient (purple) and dry (brown) CW, and ambient (red) and dry (black) OW; (b) small particles of ambient (dark green) and dry (pink) CW, and ambient (blue) and dry (grey) OW; (c) ambient large (purple) and small (dark green) particles of CW, and ambient large (red) and small (blue)

particles of OW; and (d) dry large (brown) and small (pink) particles of CW, and dry large (black) and small (grey) particles of OW. Overlaid transmission FTIR spectra of CW (upper two spectra) and OW (lower two spectra) of *P. radiata* comparing the spectral differences of (e) dry large (navy) and small (light blue) particles of CW, and dry large (light green) and small (orange) particles of OW

and the predicted values are shown plotted against the measured values in Online Resource 2. Overall, moisture content mostly had only a small effect on predictions with most RMSEP and SE values of ambient or dry large particles affected by 0.02–0.54 and 0.03–0.55%, respectively, and ambient or dry small particles affected by 0.06–0.97 and 0.01–0.46%, respectively. The largest effect was seen in the

RMSEP value for lignin using ambient, small particles which was 3.46% higher compared to dry, small particles.

Predictions of the individual monosaccharide and lignin contents of the separate test set using ATR spectra of ambient or dry, large particles, or ambient or dry, small particles are shown in Table 4, and the predicted values are shown plotted against the

Table 3 Summary of PLS-1 predictions of wood cell-wall compositions of the external validation set from samples of differing moisture content (ambient and dry) and particle size (large and small) from ATR and transmission FTIR spectroscopies

	Large/Ambient					Small/Ambient				
	Factor	R ²	RMSECV	RMSEP	SE	Factor	R ²	RMSECV	RMSEP	SE
ATR FTIR										
Arabinose	1	0.96	0.09	0.42	0.19	2	0.92	0.11	0.42	0.24
Xylose	1	0.96	0.28	1.15	0.81	2	0.94	0.37	1.11	0.84
Galactose	3	0.99	0.86	1.69	1.73	2	0.99	0.57	4.09	3.03
Glucose	1	0.91	1.65	3.35	2.86	2	0.91	1.77	3.50	3.31
Mannose	1	0.93	0.35	1.08	0.79	2	0.95	0.32	1.53	1.05
GalA	1	0.86	0.10	0.33	0.07	1	0.56	0.15	0.45	0.19
GlcA	–	–	–	–	–	1	0.52	0.05	0.15	0.12
4OMeGlcA	1	0.89	0.09	0.18	0.11	2	0.98	0.05	0.41	0.23
Lignin	2	0.95	1.16	1.35	1.19	2	0.99	0.73	5.19	3.34
	Large/Dry					Small/Dry				
	Factor	R ²	RMSECV	RMSEP	SE	Factor	R ²	RMSECV	RMSEP	SE
ATR FTIR										
Arabinose	1	0.93	0.09	0.44	0.22	2	0.91	0.13	0.31	0.23
Xylose	2	0.95	0.34	0.79	0.73	2	0.93	0.44	0.79	0.79
Galactose	2	0.95	1.14	1.34	0.68	2	0.98	0.78	2.79	2.57
Glucose	1	0.87	1.79	3.89	3.41	1	0.75	2.46	5.23	3.54
Mannose	2	0.96	0.31	0.71	0.66	2	0.94	0.38	0.56	0.92
GalA	1	0.64	0.14	0.29	0.14	1	0.53	0.16	0.29	0.10
GlcA	1	0.23	0.07	0.07	0.08	1	0.40	0.06	0.09	0.08
4OMeGlcA	2	0.91	0.08	0.10	0.14	2	0.97	0.06	0.24	0.10
Lignin	2	0.94	1.25	1.03	0.89	3	0.99	0.73	1.73	1.80
	Large/Dry					Small/Dry				
	Factor	R ²	RMSECV	RMSEP	SE	Factor	R ²	RMSECV	RMSEP	SE
Transmission FTIR										
Arabinose	2	0.88	0.13	0.43	0.42	1	0.85	0.14	0.69	0.28
Xylose	3	0.92	0.42	1.48	1.61	1	0.84	0.51	2.06	1.22
Galactose	2	0.86	1.91	3.18	3.48	4	1.00	0.36	1.96	0.87
Glucose	2	0.95	1.18	6.09	6.66	1	0.79	2.34	7.36	4.41
Mannose	3	0.92	0.40	1.44	1.53	1	0.87	0.46	2.08	1.15
GalA	2	0.77	0.13	0.23	0.23	1	0.61	0.16	0.34	0.15
GlcA	1	0.26	0.07	0.06	0.06	1	0.32	0.07	0.10	0.08
4OMeGlcA	4	0.95	0.10	0.12	0.12	2	0.96	0.07	0.23	0.10
Lignin	2	0.85	1.99	4.01	4.39	2	0.96	1.34	1.66	1.08

measured values in Online Resource 3. Overall, moisture content mostly had only a small effect on predictions with most RMSEP and SE values of

ambient or dry large particles affected by 0.00–0.68 and 0.01–0.82%, respectively, and ambient or dry small particles affected by 0.02–0.94 and 0.00–0.56%,

Table 4 Summary of PLS-1 predictions of wood cell-wall compositions of *Pinus radiata* samples from the separate test set of differing moisture content (ambient and dry) and particle size (large and small) from ATR and transmission FTIR spectroscopies

	Large/Ambient			Small/Ambient		
	Factor	RMSEP	SE	Factor	RMSEP	SE
ATR FTIR						
Arabinose	1	0.43	0.36	2	0.44	0.37
Xylose	1	1.72	1.05	2	1.22	1.12
Galactose	3	1.86	1.79	2	3.74	2.27
Glucose	1	10.52	6.32	2	6.86	6.85
Mannose	1	2.06	1.51	2	1.72	1.59
GalA	1	0.43	0.28	1	0.26	0.24
GlcA	–	–	–	1	0.13	0.07
4OMeGlcA	1	0.25	0.20	2	0.40	0.22
Lignin	2	4.01	1.67	2	4.33	1.97
	Large/Dry			Small/Dry		
	Factor	RMSEP	SE	Factor	RMSEP	SE
ATR FTIR						
Arabinose	1	0.43	0.38	2	0.38	0.38
Xylose	2	1.12	1.13	2	1.24	1.17
Galactose	2	2.54	2.56	2	2.80	2.83
Glucose	1	7.39	7.14	1	7.34	7.12
Mannose	2	1.42	1.44	2	1.61	1.60
GalA	1	0.28	0.27	1	0.24	0.23
GlcA	1	0.07	0.07	1	0.08	0.07
4OMeGlcA	2	0.23	0.23	2	0.26	0.22
Lignin	2	1.52	1.54	3	2.22	1.76
	Large/Dry			Small/Dry		
	Factor	RMSEP	SE	Factor	RMSEP	SE
Transmission FTIR						
Arabinose	2	0.68	0.47	1	0.82	0.41
Xylose	3	2.08	1.59	1	2.47	1.47
Galactose	2	6.30	3.58	4	2.70	1.95
Glucose	2	8.11	7.88	1	8.54	7.12
Mannose	3	2.16	1.68	1	2.63	1.74
GalA	2	0.26	0.26	1	0.27	0.21
GlcA	1	0.10	0.07	1	0.12	0.05
4OMeGlcA	4	0.28	0.28	2	0.38	0.23
Lignin	2	6.51	3.87	2	3.53	1.49

respectively. However, ambient, large particles had RMSEP values for Glc and lignin that were higher by 3.13 and 2.49%, respectively, compared with dry, large particles. Ambient, small particles had a RMSEP value for lignin that was higher by 2.11% compared with dry, small particles.

Effect of particle size

Particle size had only a small effect on most PLS models built using ATR spectra of ambient, large or small particles, and dry, large or small particles of the training set as similar R^2 and RMSECV values were

obtained (Table 3). However, the R^2 value of the model built using ambient, small particles for the minor component, GalA, was 0.30 lower. Particle size also had only a small effect on the PLS models built using transmission spectra of dry, large or small particles as similar R^2 and RMSECV values were obtained (Table 3). However, dry, large particles had a RMSECV value for Gal that was higher by 1.55% and for Glc that was lower by 1.16% compared with dry, small particles.

Predictions of the individual monosaccharide and lignin contents of the external validation set using ATR spectra of ambient, large or small particles, and dry, large or small particles, and transmission spectra of dry, large or small particles are given in Table 3, and the predicted values are shown plotted against the measured values in Online Resource 4. Overall, particle size mostly had only a small effect on predictions using ATR spectra with most RMSEP and SE values of ambient, small or large particles affected by 0.00–0.45 and 0.03–0.45%, respectively, and dry, small or large particles affected by 0.00–0.70 and 0.01–0.91%, respectively. The largest effects of particle size were found in predictions using ATR spectra where the RMSEP values for Gal and lignin, and the SE value for lignin using ambient, small particles were 2.40, 3.84, and 2.15% higher, respectively, compared with ambient, large particles. Overall, particle size mostly had only a small effect on predictions using transmission spectra with most RMSEP and SE values of dry, small or large particles affected by 0.04–0.64 and 0.02–0.39%, respectively. The largest effects of particle size were found in predictions using transmission spectra where dry, large particles gave a RMSEP value 2.35% higher for lignin, and the SE values for Gal, Glc, and lignin were 2.61, 2.25, and 3.31% higher, respectively, compared with dry, small particles.

Predictions of the individual monosaccharide and lignin contents of the separate test set using ATR spectra of ambient, large or small particles, and dry, large or small particles, and transmission spectra of dry, large or small particles are given in Table 4, and the predicted values are shown plotted against the measured values in Online Resource 5. Overall, particle size mostly had only a small effect on predictions using ATR spectra with most RMSEP and SE values of ambient, small or large particles affected by 0.01–0.50 and 0.01–0.53%, respectively,

and dry, small or large particles affected by 0.01–0.70 and 0.00–0.27%, respectively. The largest effect of particle size was found in the prediction of Gal using ATR spectra where the RMSEP value was 3.66% higher in ambient, large particles when compared with dry, large particles. Overall, particle size mostly had only a small effect on predictions using transmission spectra with most RMSEP and SE values of dry, small or large particles affected by 0.01–0.47 and 0.02–0.72%, respectively. The largest effects of particle size were found in predictions using transmission spectra where dry, large particles had RMSEP values for Gal and lignin that were higher by 3.60 and 2.98%, respectively, and a SE value for lignin that was higher by 2.38% when compared with dry, small particles.

Effect of sampling mode

A comparison was made of PLS models using ATR and transmission spectra of dry, large and small particles of the training set. The R^2 and RMSECV values for models built using both sets of spectra were all similar for both large and small particles (Table 3).

Predictions of the individual monosaccharide and lignin contents of the external validation set using ATR spectra of dry, large or small particles, and transmission spectra of dry, large or small particles, are shown in Table 3, and the predicted values are shown plotted against the measured values in Online Resource 6. ATR spectroscopy of dry, large and small particles better predicted most of the individual monosaccharide and lignin contents of the external validation set than did transmission spectroscopy of dry, large and small particles. Overall, sampling mode mostly had only a small effect on predictions with most RMSEP and SE values of dry, large particles affected by 0.01–0.73 and 0.02–0.88%, respectively, and dry, small particles affected by 0.01–0.83 and 0.00–0.86%, respectively. The largest differences in predictions between the two sampling modes were the RMSEP values for Glc and lignin that were 2.20 and 2.98% higher, respectively, using transmission spectroscopy of dry, large particles than predictions made using ATR spectroscopy. The SE values for Glc and lignin were 3.25 and 3.50% higher, respectively, using transmission spectroscopy of dry large particles than made predictions using ATR spectroscopy. The prediction of Glc using transmission spectroscopy of dry,

small particles had a RMSEP value that was 2.13% higher than predictions using ATR spectroscopy.

Predictions of the individual monosaccharide and lignin contents of the separate test set using ATR spectra of dry, large or small particles, and transmission spectra of dry, large or small particles, are shown in Table 4, and the predicted values are shown plotted against the measured values in Online Resource 7. ATR spectroscopy of dry, large and small particles better predicted most of the individual monosaccharide and lignin contents of the separate test set than did transmission spectroscopy of dry, large and small particles. Overall, sampling mode mostly had only a small effect on predictions with most RMSEP and SE values of dry, large particles affected by 0.02–0.96 and 0.00–0.74%, respectively, and dry, small particles affected by 0.03–0.44 and 0.00–0.88%, respectively. The largest differences in predictions between the two sampling modes were the RMSEP values for Gal and lignin that were 3.76 and 4.99% higher, respectively, and the SE value for lignin was 2.33% higher using transmission spectroscopy of dry, large particles when compared with predictions made using ATR spectra of dry, large particles.

Preferred method

The preferred method to predict the wood cell-wall compositions was ATR spectroscopy of large, ambient samples. These conditions were chosen based on the SE values for the CW and OW samples of the separate test set, which showed the variation in cell-wall composition within the two wood types. This method resulted in the lowest SE values for the five monosaccharides Ara (0.36%), Xyl (1.05%), Gal (1.79%), Glc (6.32%), and 4OMeGlcA (0.20%) (Table 4). Although ATR spectra of large, dry samples gave a lower SE value for Man (1.44%) than ATR spectra of large, ambient samples (1.51%), it was lower by only 0.07% (Table 4). Also, transmission spectra of small, dry particles gave lower SE values for GalA and lignin, 0.21 and 1.49%, respectively, than ATR spectra of large, ambient samples, 0.28 and 1.67%, respectively, but these were lower by only 0.07 and 0.18%, respectively (Table 4). GlcA was not predicted using the preferred method as no model could be built (see PLS analysis). The differences between the mean measured and mean predicted values of individual monosaccharides and lignin contents of CWs and

OWs from the separate test set are shown in Table 5. The predicted values are shown plotted against the measured values in Fig. 3.

Discussion

In the present study, ATR FTIR spectroscopy coupled with PLS regression has been successfully used to predict the monosaccharide compositions and lignin contents of *Pinus radiata* CW and OW. Using our preferred method of ATR spectra obtained from large particles (< 0.422 mm) in ambient conditions, PLS models were built that accurately predicted the monosaccharide compositions and lignin contents of the cell walls of CW and OW, as shown by the high R^2 , and low RMSEP and SE values.

However, the technique has more frequently been used simply to determine the lignin content of plant biomass. Using our preferred method, PLS prediction of lignin achieved a high coefficient of determination ($R^2 = 0.95$). This is higher than the $R^2 = 0.87$ achieved in an earlier study using smaller particles (< 178 μm) (McLean et al. 2014), also on radiata pine wood, in which the lignin content of CW and OW was determined. The coefficient of determination for lignin using our preferred method ($R^2 = 0.95$) was also higher than that obtained in other studies using PLS models built from ATR spectra. For example, the study of the four hardwood species (*Eucalyptus* sp. (*Eucalyptus*), and *Populus* spp. (cotton wood, aspen, and poplar)) (Zhou et al. 2015) using smaller particles (< 178 μm) achieved a correlation coefficient (r) of > 0.95 for lignin content, and a similar study on the bamboo species *Neosinocalamus affinis* using similar sized particles (< 250 μm) obtained the low R^2 value of 0.52 for lignin content (Sun et al. 2011). In another study on poplar (*Populus trichocarpa x deltoides*) wood using much smaller particles (< 20 μm), in which models were built for the lignin content determined by the acetyl bromide method (Dence 1992), the R^2 value achieved was 0.77 (Zhou et al. 2011).

One of the major aims of the present study was to determine the effect of particle size on the accuracy of PLS predictions of the monosaccharide compositions and lignin contents of wood cell walls. Although ATR spectroscopy is a quick analytical technique, producing small particles from wood samples is time-

Table 5 Mean measured^{ab} and mean predicted^b values of wood cell-wall compositions of CW and OW of the separate test set of *Pinus radiata* trees. Mean predicted values from ATR FTIR spectra of large, ambient wood samples

	Ara	Xyl	Gal	Glc	Man	GalA	GlcA	4OMeGlcA	Lignin
Compression wood									
Mean measured	1.5 ± 0.3	4.4 ± 0.8	12.6 ± 2.1	31.3 ± 6.7	6.0 ± 1.5	0.5 ± 0.2	0.1 ± 0.1	0.7 ± 0.1	36.5 ± 1.6
Mean predicted	1.7 ± 0.2	6.0 ± 0.7	13.2 ± 1.8	38.5 ± 2.6	7.3 ± 0.7	0.7 ± 0.1	–	0.7 ± 0.1	32.0 ± 1.6
Opposite wood									
Mean measured	2.3 ± 0.4	7.8 ± 1.4	1.8 ± 0.3	40.1 ± 7.6	8.7 ± 1.8	0.7 ± 0.2	0.1 ± 0.0	1.4 ± 0.3	24.9 ± 1.5
Mean predicted	2.5 ± 0.2	9.0 ± 0.6	2.3 ± 0.8	49.9 ± 2.3	10.2 ± 0.6	1.2 ± 0.1	–	1.4 ± 0.1	22.1 ± 1.6

^aExpressed as mean percentage of oven-dry weight of wood (± standard deviation)

^bMean of measured or predicted values of CW or OW samples (n = 20)

consuming and laborious, involving mechanical milling or grinding in a mortar and pestle. Small particles are usually used for ATR spectroscopy to achieve good contact with the crystal, and obtain spectra of sufficient intensity. Producing these small particles from wood or any other plant biomass rich in cell walls is the most time-consuming part of the technique, and so, the ability to build models and predict cell-wall composition using a technique that involves the least sample preparation will be valuable to a wide range of industries and research areas. As far as we are aware, no thorough study on the effect of particle size on the accuracy of PLS predictions of the monosaccharide compositions and lignin contents of wood has previously been carried out. Our results show that the prediction accuracy of the monosaccharide compositions and lignin contents of the separate test set from ATR spectra of large (< 0.422 mm) particles were mostly better than small (< 0.178 mm) particles, and so reducing the samples to this small particle size is unnecessary. It should be noted that sieving wood particles to achieve a smaller particle size was avoided in the present study, as particles of different sizes may have varying proportions of some cell-wall components (TAPPI T 264 cm-97 1997). However, in the present study, PLS predictions made from transmission spectra were mostly found to be better using the small particles. This was expected as the quality of transmission spectra is affected by the dispersion of the material in the KBr matrix, and so the finer the material, the better the dispersion (Faix and Böttcher 1992).

Although studying the effect of particle size on the spectra obtained was not a major aim, it was also investigated. We observed differences in relative band intensities between the spectra of large and small particles of CW and OW, collected in both ATR and transmission spectroscopies, and it was unknown if this would strongly impact the accuracy of prediction by PLS. A study using transmission spectroscopy on different sized wood particles (< 25 to > 100 µm) of the hardwood tree bubinga (*Guibourtia* sp.) also found differences in the intensities of spectral bands with different particle sizes (Faix and Böttcher 1992). In the present study, the effect of milling and grinding on the chemical structures of wood cell-wall components was not investigated, but a different study has reported changes in the relative band heights in FTIR spectra of Norway spruce (*Picea abies*) after ball milling for different times, and concluded there were changes in chemical structures, which were not due to temperature increases (Schwanninger et al. 2004). A reduction in the degree of polymerization of cellulose due to the breaking of glycosidic linkages has also been found after ball milling plant cell walls (Kim et al. 2008). We also observed differences in relative band intensities between ATR spectra of CW and OW, and made assignments to these bands according to the literature. In the present study, the compositional differences between latewood and earlywood were not investigated, but another study has reported differences in relative band heights of mid-IR spectra between the latewood and earlywood of southern pines (*Pinus* spp.) (Via et al. 2011).

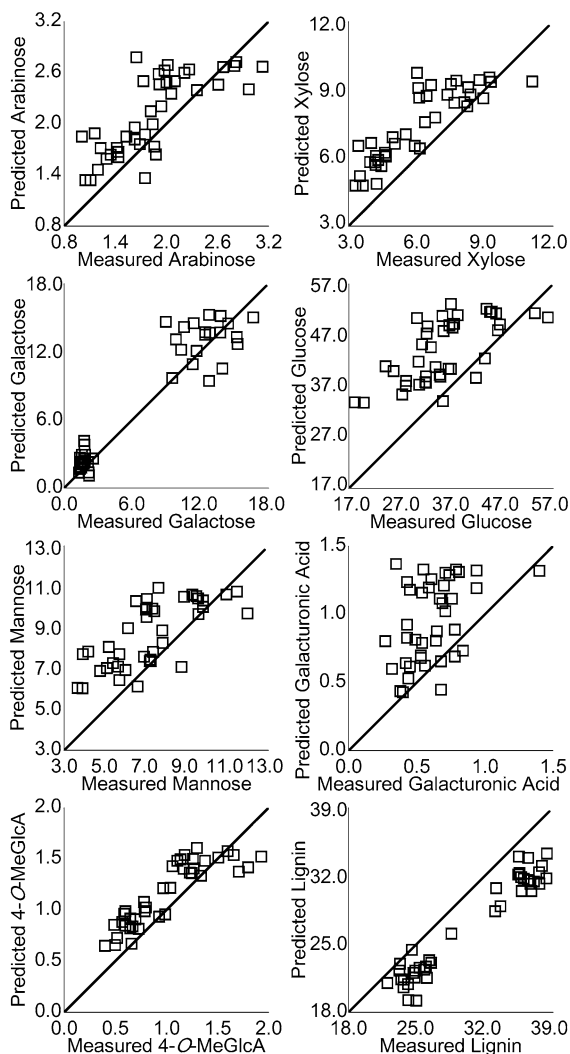


Fig. 3 ATR FTIR predicted values are plotted against the measured values for the chemical composition of compression and opposites woods from field grown, *Pinus radiata*, the separate test set. Values predicted by the preferred method, ATR FTIR of large particles in ambient conditions are plotted

Another aspect examined in the present study was the effect of moisture content of samples on the PLS prediction accuracy. Because water absorbs strongly in the mid-infrared region, this region ($\sim 1639\text{ cm}^{-1}$) of wood IR spectra is highly sensitive to moisture content. However, drying wood to constant weight to reduce the variation in moisture content can be time consuming. Comparison of ATR spectra of ambient and dry wood samples showed differences in the intensities of the whole spectra, but the effect of sample moisture on the accuracy of prediction by PLS

models was unknown. The present study showed that predictions of monosaccharide compositions and lignin contents of the CW and OW were mostly better using samples kept in ambient conditions. This is most likely because during collection of ATR spectra the wood kept under ambient conditions absorbed very little further atmospheric moisture, and also because a background spectrum was collected before every set of three replicate spectra of each sample to factor out any small humidity changes throughout the day that may affect the spectra.

For PLS regression analysis, the development of training sets should ideally comprise samples that provide a wide range of values for the components or attributes being modelled and which covers the range to be predicted. Our training set using mixtures of CW and OW was successfully developed to provide such a range by utilising the differences in cell-wall compositions of CW and OW, and to avoid the clustering of data for the monosaccharide compositions and lignin contents. As the present study showed, this set of wood mixtures provided a linear range of monosaccharide compositions and lignin contents, but it should be noted that its range was not able to cover that of several of the separate test samples. In a preliminary study (data not shown), exclusion of the samples that fell outside the training set range actually resulted in higher RMSEP and SE values, and so, all samples were included. A recent study used a similar strategy to give an extended range of cell-wall compositions by exploiting the different cell-wall compositions of hardwoods and softwoods. Mixtures were used of the hardwoods *Quercus petraea* (sessile oak), *Fagus sylvatica* (beech), and the softwoods *Pinus* spp. (pine), and *Abies* spp. (fir) (Duca et al. 2016). However, these sets were not used to predict monosaccharide compositions and lignin contents, but only to predict the proportions of hardwoods and softwoods in mixed samples. Another study took a different approach and used commercially available beech wood xylan, lignin and cellulose blended in different proportions to provide a range of values for each component (Krasznai et al. 2012).

In conclusion, the known compositional differences between CW and OW were used by mixing the two wood types to provide a training set that had a wide, linear range of monosaccharide compositions and lignin contents. PLS models were built using ATR spectra of the training set, and were used to

successfully predict the monosaccharide compositions and lignin contents of the external validation and separate test sets. ATR spectroscopy of large particles in ambient condition gave the best predictions for most of the monosaccharide compositions of the separate test set. Although some monosaccharide compositions and lignin contents were slightly better predicted by other methods, our preferred method was chosen as it is the least time consuming and laborious sample preparation and spectroscopic method. This method could be applied to determining the monosaccharide compositions and lignin contents of the cell walls of other materials.

Acknowledgments We thank Professor John C. F. Walker (School of Forestry, University of Canterbury) for providing the wood samples, and Professor Christopher M. Triggs and Associate Professor Brian H. McArdle (Department of Statistics, University of Auckland) for statistical advice. This work was supported by the New Zealand Foundation for Research, Science and Technology (now Ministry of Business, Innovation and Employment) [PROJ-12401-PPS-UOC, “Compromised Wood Quality”].

References

- Åkerholm M, Salmén L (2001) Interactions between wood polymers studied by dynamic FT-IR spectroscopy. *Polymer* 42:963–969
- Apiolaza LA, Butterfield B, Chauhan SS, Walker JCF (2011) Characterization of mechanically perturbed young stems: can it be used for wood quality screening? *Ann For Sci* 68:407–414
- Araque E, Parra C, Freer J, Contreras D, Rodríguez J, Mendonça R, Baeza J (2008) Evaluation of organosolv pretreatment for the conversion of *Pinus radiata* D. Don to ethanol. *Enzym Microb Technol* 43:214–219
- Baker MJ et al (2014) Using Fourier transform IR spectroscopy to analyze biological materials. *Nat Protoc* 9:1771–1791
- Blakeney AB, Harris PJ, Henry RJ, Stone BA (1983) A simple and rapid preparation of alditol acetates for monosaccharide analysis. *Carbohydr Res* 113:291–299
- Brennan M, McLean JP, Altaner CM, Ralph J, Harris PJ (2012) Cellulose microfibril angles and cell-wall polymers in different wood types of *Pinus radiata*. *Cellulose* 19:1385–1404
- Chavan RR, Fahey LM, Harris PJ (2015) Quantification of (1 → 4)-β-D-galactans in compression wood using an immuno-dot assay. *Plants* 4:29–43. doi:10.3390/plants4010029
- Currie HA, Perry CC (2006) Resolution of complex monosaccharide mixtures from plant cell wall isolates by high pH anion exchange chromatography. *J Chromatogr* 1128:90–96
- Dence CW (1992) The determination of lignin. In: Lin SY, Dence CW (eds) *Methods in lignin chemistry*. Springer, Berlin, pp 33–61
- Duca D, Pizzi A, Rossini G, Mengarelli C, Foppa Pedretti E, Mancini M (2016) Prediction of hardwood and softwood contents in blends of wood powders using mid-infrared spectroscopy. *Energy Fuels* 30:3038–3044
- Faix O (1991) Classification of lignins from different botanical origins by FT-IR spectroscopy. *Holzforchung* 45:21–27
- Faix O, Böttcher JH (1992) The influence of particle size and concentration in transmission and diffuse reflectance spectroscopy of wood. *Holz Roh Werkst* 50:221–226
- Fengel D, Ludwig M (1991) Möglichkeiten und grenzen der FTIR-spektroskopie bei der charakterisierung von cellulose. Teil 1. Vergleich von verschiedenen cellulose-fasern und bakterien-cellulose. *Das Papier* 45:45–51
- Fukushima RS, Hatfield RD (2004) Comparison of the acetyl bromide spectrophotometric method with other analytical lignin methods for determining lignin concentration in forage samples. *J Agric Food Chem* 52:3713–3720
- Hair JFJ, Black WC, Babin BJ, Anderson RE (2010) *Multivariate data analysis. A global perspective*, 7th edn. Pearson, Upper Saddle River
- Harrington KJ, Higgins HG, Michell AJ (1964) Infrared spectra of *Eucalyptus regnans* F. Muell. and *Pinus radiata* D. Don. *Holzforchung* 18:108–113
- Harris PJ (2005) Diversity in plant cell walls. In: Henry RJ (ed) *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*. CABI Publishing, Wallingford, pp 201–227
- Harris PJ, Stone BA (2008) Chemistry and molecular organization of plant cell walls. In: Himmel ME (ed) *Biomass recalcitrance: deconstructing the plant cell wall for bioenergy*. Blackwell Publishing Ltd., Oxford, pp 61–93
- Hergert HL (1960) Infrared spectra of lignin and related compounds. II. Conifer lignin and model compounds. *J Org Chem* 25:405–413
- Hergert HL (1971) Infrared spectra. In: Sarkanen KV, Ludwig CH (eds) *Lignins. Occurrence, formation, structure and reactions*. John Wiley and Sons, New York, pp 267–297
- Higgins HG, Stewart CM, Harrington KJ (1961) Infrared spectra of cellulose and related polysaccharides. *J Polym Sci* 51:59–84
- Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315:804–807
- John MJ, Thomas S (2008) Biofibres and biocomposites. *Carbohydr Polym* 71:343–364
- Kačuráková M, Capek C, Sasinková V, Wellner N, Ebringerová A (2000) FT-IR study of plant cell wall model compounds: pectic polysaccharides and hemicelluloses. *Carbohydr Polym* 43:195–203
- Kačuráková M, Wilson RH (2001) Developments in mid-infrared FT-IR spectroscopy of selected carbohydrates. *Carbohydr Polym* 44:291–303
- Kato K, Nitta M, Mizuno T (1973) Infrared spectroscopy of some mannans. *Agric Biol Chem* 37:433–435
- Kim H, Ralph J, Akiyama T (2008) Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO-*d*₆. *Bioenergy Res* 1:56–66
- Krasznai DJ, Champagne P, Cunningham MF (2012) Quantitative characterization of lignocellulosic biomass using

- surrogate mixtures and multivariate techniques. *Bioresour Technol* 110:652–661
- Liang CY, Marchessault RH (1959) Infrared spectra of crystalline polysaccharides. II. Native celluloses in the region from 640 to 1700 cm^{-1} . *J Polym Sci* 39:269–278
- Lupoi JS, Singh S, Simmons BA, Henry RJ (2013) Assessment of lignocellulosic biomass using analytical spectroscopy: an evolution to high-throughput techniques. *Bioenergy Res* 7:1–23
- Marchessault RH (1962) Application of infra-red spectroscopy to cellulose and wood polysaccharides. *Pure Appl Chem* 5:107–130
- Marchessault RH, Liang CY (1962) The infrared spectra of crystalline polysaccharides. VIII. Xylans. *J Polym Sci* 59:357–378
- Mathlouthi M, Koenig JL (1987) Vibrational spectra of carbohydrates. *Adv Carbohydr Chem Biochem* 44:7–89
- McLean JP, Jin G, Brennan M, Nieuwoudt MK, Harris PJ (2014) Using NIR and ATR-FTIR spectroscopy to rapidly detect compression wood in *Pinus radiata*. *Can J For Res* 44:820–830
- Müller G, Schöpfer C, Vos H, Kharazipour A, Polle A (2009) FTIR-ATR spectroscopic analyses of changes in wood properties during particle- and fibreboard production of hard- and softwood trees. *BioResources* 4:49–71
- Pauly M, Keegstra K (2008) Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J* 54:559–568
- Saeman JF, Moore WE, Millett MA (1963) Sugar units present. Hydrolysis and quantitative paper chromatography. In: Whistler RL (ed) *Methods in carbohydrate chemistry*, vol 3. Academic Press, New York, pp 54–69
- Scheller HV, Ulvskov P (2010) Hemicelluloses. *Annu Rev Plant Biol* 61:263–289
- Schwanninger M, Rodrigues JC, Pereira H, Hinterstoisser B (2004) Effects of short-time vibratory ball milling on the shape of FT-IR spectra of wood and cellulose. *Vib Spectrosc* 36:23–40
- Smith BG, Harris PJ (1995) Polysaccharide composition of unignified cell walls of pineapple [*Ananas comosus* (L.) Merr.] fruit. *Plant Physiol* 107:1399–1409
- Sun B, Liu J, Liu S, Yang Q (2011) Application of FT-NIR-DR and FT-IR-ATR spectroscopy to estimate the chemical composition of bamboo (*Neosinocalamus affinis* Keng). *Holzforschung* 65:689–696
- TAPPI T 222 om-98 (1998) Acid-insoluble lignin in wood and pulp
- TAPPI T 249 cm-85 (1985) Carbohydrate composition of extractive-free wood and wood pulp by gas-liquid chromatography
- TAPPI T 264 cm-97 (1997) Preparation of wood for chemical analysis
- Timell TE (1986) *Compression wood in gymnosperms*. Springer, Berlin
- Via BK, Fasina O, Pan H (2011) Assessment of pine biomass density through mid-infrared spectroscopy and multivariate modeling. *BioResources* 6:807–822
- Walker J (2013) Wood quality: a perspective from New Zealand. *Forests* 4:234–250. doi:10.3390/f4020234
- Xu F, Yu J, Tesso T, Dowell F, Wang D (2013) Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: a mini-review. *Appl Energy* 104:801–809
- Zhang M, Chavan RR, Smith BG, McArdle BH, Harris PJ (2016) Tracheid cell-wall structures and locations of (1 → 4)- β -D-galactans and (1 → 3)- β -D-glucans in compression woods of radiata pine (*Pinus radiata* D. Don). *BMC Plant Biol* 16:18. doi:10.1186/s12870-016-0884-3
- Zhou G, Taylor G, Polle A (2011) FTIR-ATR-based prediction and modelling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars. *Plant Methods* 7:1–10. doi:10.1186/1746-4811-7-9
- Zhou C, Jiang W, Via BK, Fasina O, Han G (2015) Prediction of mixed hardwood lignin and carbohydrate content using ATR-FTIR and FT-NIR. *Carbohydr Polym* 121:336–341