

Original article

Guillaume Hans* and Bruce Allison

On-line characterization of wood chip brightness and chemical composition by means of visible and near-infrared spectroscopy

<https://doi.org/10.1515/hf-2021-0027>

Received February 12, 2021; accepted May 7, 2021;

published online June 7, 2021

Abstract: Historically, on-line and real-time measurement of wood chip properties in the pulp and paper industry has been a challenge and has hampered the development of advanced process control strategies. In this study, visible and near-infrared (VIS-NIR) spectroscopy is investigated as a means to characterize wood chip brightness and chemical composition (i.e. extractives, lignin and holocellulose content) on-line. The estimated standard error on the holocellulose reference measurement was significantly reduced using data reconciliation. VIS-NIR calibration models were developed using partial least square regression. Derivative and baseline correction were found to be the most appropriate pre-processing methods. Model desensitization to the influence of moisture content and temperature by means of external parameter orthogonalization resulted in more robust models critical for on-line applications under harsh industrial conditions. Wavelength selection improved model accuracy for all properties. A comparison of two different spectrometer and probe combinations demonstrated that, after wavelengths selection, a non-contact measurement of wood chips performs as well as a contact measurement of wood powder for monitoring chemical composition. On-line prediction of wood chip brightness and chemical composition using the developed VIS-NIR models was demonstrated over 7 months in a kraft pulp mill processing both hardwood and softwood chips.

Keywords: analyzer; chemical composition; near-infrared spectroscopy; non-destructive; on-line monitoring; wood chips.

1 Introduction

Wood is a biological raw material and, as such, has an inherent variability in chemical composition and physical properties. Wood chemical composition consists mainly of holocellulose (cellulose and hemicelluloses), lignin, and extractives. The concentrations vary substantially among trees and are a function of wood type (hardwood versus softwood), species, age, growing conditions and season of harvest (see Table 1). Wood properties also vary with anatomy within a tree and after harvest during storage and seasoning. Historically, on-line and real-time measurement of wood chip properties in the pulp and paper industry has been a challenge and has hampered the development of advanced process control strategies (Trung and Allison 2015). MacLeod (2007) summarizes the situation as follows, “Surprisingly for a worldwide industry which has been in business for many decades, there is no simple, fast, and cheap way to determine the gross chemical composition of the wood in use”. According to MacLeod (2007), the chemical composition of wood is the primary variable in determining kraft pulp yield, delignification rate and the alkali charge needed to reach a target kappa number. Brightness is a physical property of the wood that is of particular importance for thermomechanical pulp (TMP) as it influences bleaching agent consumption and product quality (Hu et al. 2011; Schanda 2007). High feedstock variability makes it difficult to obtain uniformly high quality pulp at the desired production rate, yield and target properties. Frequent and low-cost measurement of wood chip properties would be beneficial to the industry in many ways. First, it could be used to develop new and advanced process control strategies for optimizing operations (digester, refiner and bleach plant) in order to maximize the value extracted from the feedstock, minimize waste, and keep production volumes and product quality on target. Second, it would facilitate process troubleshooting when issues arise in a mill (e.g. accelerated corrosion, pitch depositions, kappa variability, etc.). Third, it would help to gain a better understanding of the chip seasoning process and, in turn, to better

*Corresponding author: Guillaume Hans, Smart Manufacturing, FPInnovations, 2665 East Mall, Vancouver, BC V6T 1Z4, Canada, E-mail: guillaume.hans@fpinnovations.ca. <https://orcid.org/0000-0002-6503-5760>

Bruce Allison, Smart Manufacturing, FPInnovations, 2665 East Mall, Vancouver, BC V6T 1Z4, Canada

Table 1: Chemical composition of hardwoods and softwoods (in % dry mass).

Chemical	Hardwoods	Softwoods	References
Cellulose	38–51	33–42	Tarasov et al. (2018)
Hemicellulose	17–38	22–40	
Lignin	21–31	27–32	
Extractives	2–4	2–5	Sjöström and Alén (1999)
Ash	0.2–0.8	0.2–0.4	Rowell (2005)

regulate the duration of seasoning and its impact on the pulping process. Fourth, it could help to better assess the quality and value of purchased chips.

The literature shows that visible and near-infrared (VIS-NIR) spectroscopy combined with chemometrics can be used to measure a multitude of wood properties (Tsuchikawa 2007; Tsuchikawa and Kobori 2015). VIS-NIR spectroscopy is a promising technology for on-line real-time characterization of wood and biomass properties because it is non-contacting, non-destructive, fast and safe with no requirement for sample preparation and, most importantly, it can measure multiple properties simultaneously. Supplementary Table S1 summarizes results obtained from literature studies on the measurement of brightness and wood chemical compounds. Most of these publications were either feasibility studies or studies that aimed at developing a portable or bench-top spectroscopic system for solid or ground wood in a conditioned or dry state. The measurement of fresh wood chips entering a mill is a much harder problem. First, fresh chips are expected to exhibit wide variations in moisture content and temperature that influence the spectra and, thereby, interfere in the prediction of the targeted properties (Hans and Allison 2019). Second, random chip orientation and conveyer movement result in more spectral variability than from a stationary sample (Axrup et al. 2000). Off-line predictions of TMP brightness from VIS-NIR spectra collected on solid and ground wood samples was reported in Nilsson et al. (2005). However, this study, like others using RGB cameras (Achiche et al. 2005, 2007; Benaoudia et al. 2005), inferred pulp brightness and did not measure wood brightness directly. A direct method to measure wood chip ISO brightness off-line was proposed in Hu et al. (2011) and a linear correlation to TMP brightness was demonstrated. On-line measurements of fresh wood chip extractives and Klason lignin were reported in Axrup et al. (2000). A model for total (soluble and insoluble) lignin content was subsequently presented in Skvaril et al. (2015a) and used in simulation of model based control and diagnostic strategies for a continuous pulp digester (Moksadur et al. 2018; Skvaril et al. 2015b). To the best of our knowledge, these four studies are the only ones that discuss the

on-line measurement of fresh wood chip chemical composition under industrial conditions. However, the problem of calibration model robustness to confounding variables such as moisture content and temperature was not discussed in their work. In addition, it appears that the measurement of fresh chip holocellulose content has not been presented.

Recently, an on-line analyzer based on NIR spectroscopy capable of measuring wood chip and biomass moisture content, temperature and calorific value was developed (Hans and Allison 2019; Hans et al. 2017) and subsequently employed in a dry chip feed-rate control strategy for continuous digesters (Gilbert et al. 2018). The novelty of this study is the extension of this work to the on-line characterization of wood chip brightness and chemical composition, i.e. extractives, lignin and holocellulose content while accounting for the influence of moisture content and temperature on the spectra as well as changes in chip properties occurring during seasoning. For this purpose, wood chips from four western Canadian species (hardwoods and softwoods) were seasoned for one year and samples collected every month. Including different species provided variability in holocellulose and lignin content while variability in brightness and extractives content was generated by seasoning (Allen et al. 1991; Gutiérrez et al. 1998, 1999). This variability is especially useful for developing robust calibration models based on VIS-NIR spectroscopy. Potential errors in the reference measurements were mitigated using data reconciliation (Crowe et al. 1983). VIS-NIR measurements were done on frozen and unfrozen fresh chips as well as on dry chips and wood powder by means of two different spectrometers. This experimental protocol enabled studying the effect of the instrumentation design and material type and condition on the spectra and prediction models. The choice of spectra pre-processing method and wavelength selection are discussed and the models are desensitized for the influence of moisture content and temperature by means of external parameter orthogonalization (EPO) (Roger et al. 2003). Finally, the calibration models obtained are qualitatively validated on-line over 7 months of data acquired in a western Canadian kraft mill that produces both hardwood and softwood pulp grades.

2 Materials and methods

2.1 Sample preparation

In October 2018, logs were freshly harvested in the vicinity of the town of 100 Mile House, in the province of British Columbia, Canada. Two logs of 2 m long and with an average diameter of about 15 cm in diameter were harvested from the base of two separate trees for four

different species common in the area: aspen (*Populus tremuloides* Michx.), black cottonwood (*Populus trichocarpa* Torr. & A.Gray ex. Hook.), black spruce (*Picea mariana* (Mill.) Britton, Sterns & Poggenburg) and Douglas fir (*Pseudotsuga menziesii* (Mirbel) Franco). In the remainder of this paper, we will refer to these as A, BC, BS and DF, respectively. The logs were sealed in plastics bags immediately after harvesting and shipped to FPInnovations' facilities in Vancouver. Logs were debarked manually and chipped in a CM&E Model 36 disc chipper. Chips from each species were sealed in separate plastics bags and the chipper was cleaned between species to avoid mixing. In November 2018, four small chip piles of approximately 1 m in diameter and 50 cm in height, were constructed in FPInnovations' yard and left to age for one year under natural conditions. Each month, a sample of chips of approximately 3 kg was collected from the surface of each pile, sealed in a plastic bag and immediately frozen. In November 2019, all samples collected during the previous 12 months were processed through a Williams classifier for 1.5 min. Chips smaller than 3/8" (0.95 cm) and larger than 1 1/8" (3.49 cm) were discarded.

2.2 Near-infrared spectroscopy

Fresh samples were scanned frozen (−20 °C) and unfrozen (20 °C) with a Polytec PSS-2121 (Polytec GmbH, Waldbronn, Germany) near-infrared (NIR) spectrometer connected to a Polytec PSS-H-A03 sensor head. This spectrometer operates in the wavelength range of 1100–2100 nm. Further details on this system are given in Hans et al. (2017). The sensor head was installed over a turntable, at a distance of 25 cm from the surface of the sample that was placed in a 40 cm diameter pan. The Polytec spectrometer's integration time was set to 25 ms and 400 spectra were averaged per scan. The samples were then dried overnight at 35 °C in a ventilated drying chamber to a moisture content of 5.5 ± 0.8% (wet basis). In the chamber, the chips were spread on a metallic lattice to ensure even drying. Samples were then re-scanned under the Polytec system and with an ASD LabSpec 4 VIS-NIR spectrometer connected to an ASD turntable. The ASD LabSpec 4 operates in the wavelength range of 350–2500 nm with a resolution of 1 nm. The light source of the ASD turntable was positioned 10 cm above the surface of the sample that was placed in a 14 cm diameter pan. The ASD spectrometer's integration time was set to 34 ms and 50 spectra were averaged per scan. Finally, following standard T257 (Tappi standard 2014), the chip samples were ground in a Wiley mill (model ED-5; Thomas Scientific, Swedesboro, New Jersey) equipped with a 0.50 mm sieve and then further ground in a second Wiley mill (model 8-338; Thomas Scientific) equipped with a 40 mesh (0.38 mm). Dry ice was added with the chips in the Wiley mills to avoid over-heating (Hu et al. 2011). A subsample of the final wood powder produced was collected from each sample and placed in an aluminum pan with a diameter of 5 cm. This subsample was scanned with the ASD LabSpec 4 connected to an ASD contact probe with an illumination diameter of 1 cm. The

contact probe was placed directly in contact with the powder but no pressure was applied (other than that generated by the probe's own weight).

This succession of spectroscopic measurements generated four distinct spectral datasets acquired from various types of wood material, with different instruments and under diverse conditions of moisture and temperature. The characteristics of these datasets along with their labels are summarized in Table 2.

2.3 Brightness and chemical analysis

For each sample, another subsample of wood powder was collected and ISO brightness was measured as the directional reflectance of light at 457 nm following Hu et al. (2011). A final subsample of approximately 1 g of wood power was extracted with acetone for 20 min on an ASE (Accelerated Solvent Extraction) system (Thermo Fischer Scientific) at 1600 psi and 90 °C. The extractives were evaporated, oven dried for 1 h and weighed to determine the percentage of extractives. The extracted wood powder was weighed in a beaker, hydrolyzed with 72% sulfuric acid and diluted to 3% by adding 84 mL of milli-Q water before autoclaving for 1 h at 121 °C. After cooldown, the resulting samples were filtered in a medium grit glass filter (Corning Pyrex 30 ml Gooch Crucible Medium Fritted 30 mm Disc). The filtered solids were dried and ashed at 525 °C to determine the percentage of Klason or acid insoluble lignin. The filtrate of the hydrolyzed sample was diluted to 100 ml and analyzed by UV spectroscopy (at 205 nm using a quartz cell) to determine the percentage of acid soluble lignin. Total lignin was computed as the sum of acid soluble and insoluble lignin. The filtrate was then further diluted to 100 mL and analyzed by ion chromatography with the ICS-5000+ (Thermo Fischer Scientific) configured for HPAE-PAD (high-performance anion-exchange chromatography with pulsed amperometric detection) and equipped with a CarboPac PA-1 analytical column and a post-column addition. Six standards were used to constitute the calibration curve and an internal standard of deoxy-D-glucose was added to all injections to ensure the stability of the results. Following Acquah et al. (2015), the percentages of arabinan, galactan, glucan, xylan and mannan were determined and summed to obtain holocellulose.

2.4 Data reconciliation

Raw data is subject to errors. Data reconciliation is a technique used to denoise data through the enforcement of equality constraints that typically arise from conservation laws, such as mass and energy balances. It aims at making the minimum adjustments to data to satisfy conservation laws (Crowe et al. 1983). In this study, the measurement of extractives, lignin and holocellulose content was performed on a percent of total mass basis. The sum of these constituents must equal

Table 2: Spectral database characteristics summary.

Dataset label	Material type	Spectrometer	Moisture	Temperature
Polytec-Fresh	Chips	Polytec	Fresh (variable)	Frozen and unfrozen
Polytec-Dry	Chips	Polytec	Air-dry	Room
ASD-Chips	Chips	ASD	Air-dry	Room
ASD-Powder	Powder	ASD	Air-dry	Room

99.5% for hardwoods and 99.7% for softwoods to meet conservation laws, assuming average ash contents of 0.5 and 0.3, respectively (Table 1). According to this, and following Crowe et al. (1983), the raw data (X) were reconciled (X_{cor}) by matrix projection as follows:

$$X_{\text{cor}} = X - PB^T(BPB^T)^{-1}BX$$

where P is a diagonal matrix of the squared standard error of measurements (SEMs) and B is a balance matrix (+1 for inputs and -1 for outputs). For extractives, because a linear regression model could be fit to the original (unreconciled) trends with p -value <0.001 for each species; the root mean square error (RMSE) of this model was used as an estimate of the SEMs. For lignin and holocellulose, the standard deviation of the data for each species was used to estimate the SEMs. The estimated SEMs used for the data reconciliation are given in Table 3 for each species.

3 Results and discussion

3.1 Age and species effects

In order to mitigate variability in feedstock properties, it is common practice in the pulp and paper industry to season the wood. Seasoning is the practice of storing the wood in piles (either of logs or chips) for a certain amount of time. This process typically leads to a reduction of extractives content and to changes in the wood chemical composition (Allen et al. 1991; Gutiérrez et al. 1998, 1999). When chips are stored in piles, only a small layer at the surface is exposed to abiotic factors. This fraction is referred to as the “outer weather belt” in Heinek et al. (2013) and has a depth of 20–40 cm. In large chip piles, the rest of the chips generally offer an aerobic atmosphere protected from most abiotic factors. Under these conditions, several processes and chemical reactions take place including exothermic oxidation reactions, aerobic respiration (from the microorganisms), enzymatic reactions, heat transfer, continuation of life functions (remaining living cells) and evaporation of volatile compounds (Ferrero et al. 2009; Jirjis and Theander 1990; Nugent et al. 1976; Ramnath et al. 2018). A review of the microorganism communities present in large-scale

wood chip and log piles is provided in Noll and Jirjis (2012). However, the chip piles created for the purpose of this study, because of the small size, were only representative of the “outer weather belt” and it can be safely assumed that the inner part was not protected. As a result, the chips were mostly affected by abiotic factors such as rain, temperature, and UV radiation (Oberhofnerová et al. 2017) as well as by non-decaying fungi (mould, yeast and staining fungi) that grow at the surface or within cell lumens (Stirling 2005).

Table 3 shows the estimated SEMs after data reconciliation. In comparison with the ones before, it can be seen the estimated SEMs of extractives and lignin both remain essentially unchanged. However, the SEMs of holocellulose were reduced by a factor of 3–8 depending on species. This reflects the relatively higher measurement error for this compound obtained with the reference method (ion chromatography).

Reconciled data illustrating changes in chip brightness and chemical composition with age are shown in Figure 1. It can be seen that brightness and extractives both decreased with age. The statistical significance of these trends was determined by means of an F -test ($\alpha = 0.05$) over a linear regression model. All trends were significant for both variables and all species with only one exception: the brightness trend of DF had a p -value of 0.056. Loss in brightness and extractives is explained by the abiotic degradation of the wood chips. Brightness loss is attributed mostly to the growth of non-decaying fungi at the surface of the chips and the leaching of water soluble compounds. A 7% drop in ISO brightness of lodgepole pine chips was reported in Stirling (2005) after 115 days of storage and was attributed to the grow of mould. To a lesser extent, brightness loss may also have been caused by the deposition of air dust particles (Oberhofnerová et al. 2017). Loss of extractives was caused by oxidation, leaching and evaporation processes as well as degradation by non-decaying fungi (Gutiérrez et al. 2009; Stirling 2005; Weigel et al. 2002).

No noticeable changes in lignin were observed while only two species, BC and DF, showed a significant increase in holocellulose with age. The small pile size did not create the right conditions for wood decaying fungi such as white-rot, brown-rot and soft-rot to develop. This explains the constant percentage of lignin content observed since it is well accepted that losses in lignin and holocellulose are mainly caused by the activity of decaying fungi. The slight increase of holocellulose for BC and DF observed in this study is only relative as the amount of each chemical compound is expressed as a percentage of the total mass, which decreases with time, i.e. the percentage of

Table 3: Estimates of the standard errors of measurement before and after data reconciliation.

Data reconciliation	Property/species	A	BC	BS	DF
Before	Extractives	0.206	0.150	0.120	0.195
	Lignin	0.599	0.638	0.622	0.653
	Holocellulose	2.158	2.616	5.362	3.490
After	Extractives	0.214	0.149	0.120	0.197
	Lignin	0.572	0.540	0.589	0.611
	Holocellulose	0.757	0.841	0.666	0.660

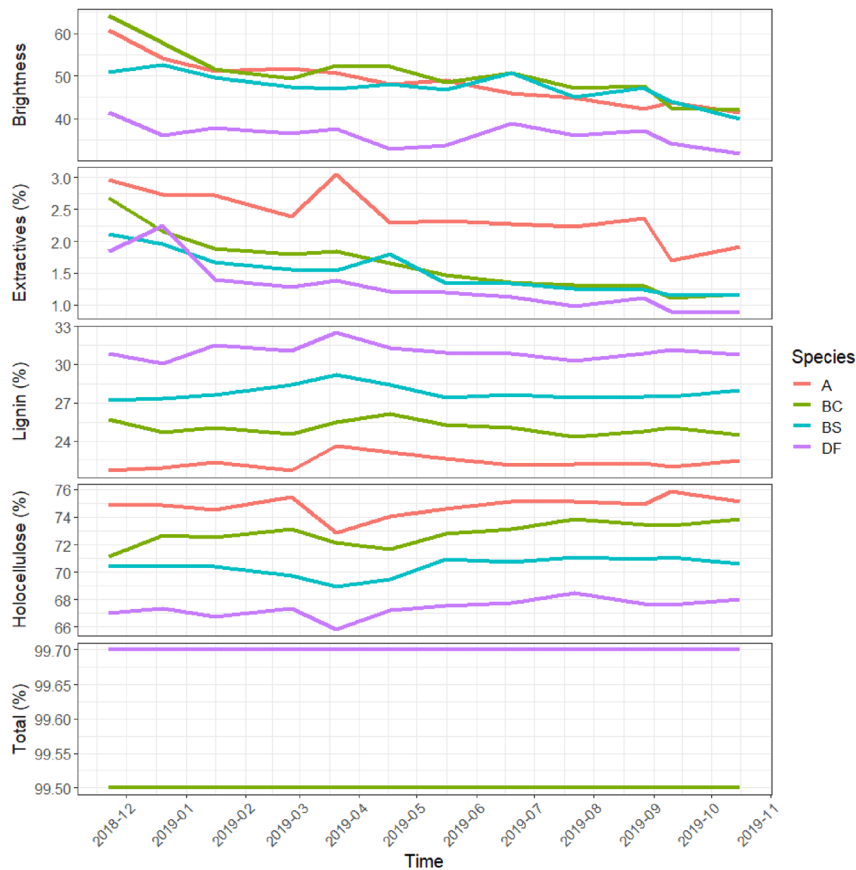


Figure 1: Change in wood chip brightness, extractives, lignin and holocellulose content as well as the sum of the three latter (total) with age for four different species: aspen (A), black cottonwood (BC), black spruce (BS) and Douglas-fir (DF).

extractives dropped by an amount similar to the increase in holocellulose. The fact that the relative percentage of holocellulose increased and not lignin is an artifact of measurement error and data reconciliation. Under ideal circumstances one would expect both holocellulose and lignin percentages to increase proportionally to offset the loss of extractives.

Differences in brightness and chemical composition between species are shown in Figure 2. Highly significant differences (p -value < 0.0001) between all species were observed for lignin and holocellulose. Differences in brightness were mostly observed between the hardwood and softwood groups, with DF having very low brightness. This difference is visible with the naked eye since DF has a characteristic dark red tint due to its heartwood extractives composition.

Extractives were significantly higher for A while no noticeable differences were observed between the three other species. Species A is well known for its high extractives content among Canadian species (Allen 1987). The chemical composition of all four species agrees with the literature (Table 1). Similar levels of extractives and differences between A, BS and DF were reported in Allen (1987). For these, the results obtained for total lignin and

holocellulose content are in accordance with the data reported in Kaar and Brink (1991) and Rowell (2005). No reported values were found for BC.

3.2 VIS-NIR calibration model development

Calibration models capable of predicting brightness, extractives, lignin and holocellulose (target variables) from VIS-NIR spectra were developed for each dataset presented in Table 2. The significant reduction in the estimated SEMs obtained for holocellulose after data reconciliation (Table 3) directly translated into more accurate calibration models. Calibrations obtained for extractives and lignin before and after data reconciliation showed similar performance. For conciseness, only the results obtained from the reconciled data will be presented. The calibration models were developed and refined in three steps as described below.

3.2.1 Preprocessing optimization

For each dataset, the following spectra pre-processing methods were evaluated: standard normal variates

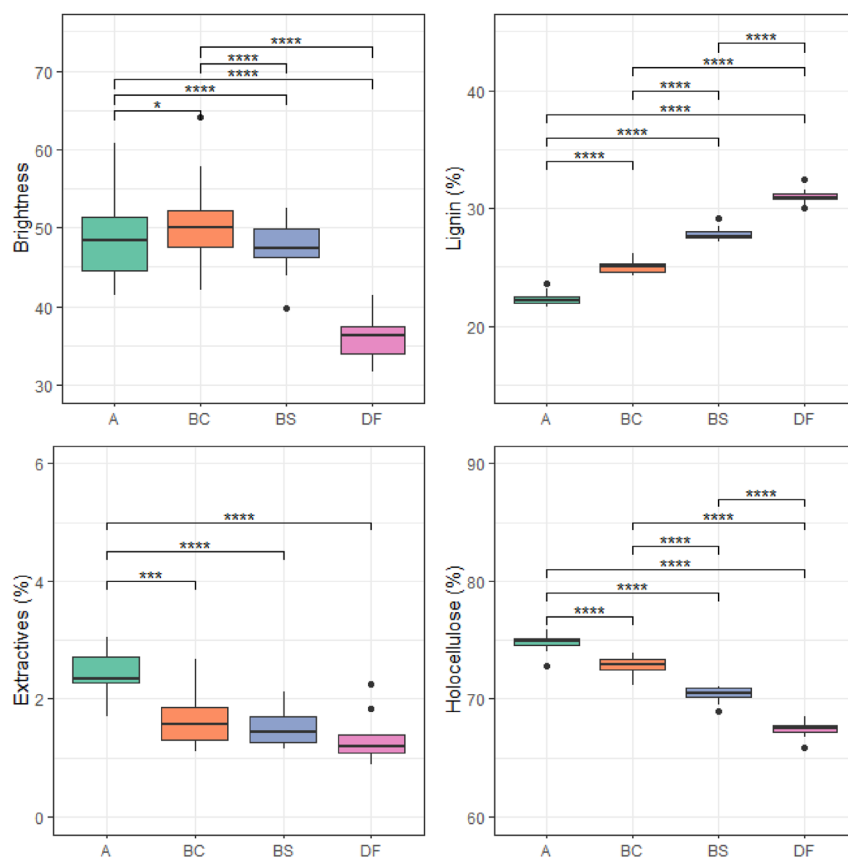


Figure 2: Boxplots showing differences in brightness, extractives, lignin and holocellulose content between species. Differences were determined regardless of age influence by ANOVA followed by Tukey's test with the following coding for p -value: **** < 0.0001, *** < 0.001, ** < 0.01, * < 0.05.

(SNV), multiplicative scatter correction (MSC), derivative, smoothing, detrending, baseline correction and VSN (variable sorting for normalization). Where applicable, parameters of the preprocessing method were tuned using a wide range of parameters. Partial least squares (PLS) regression models were built for each set of pre-processed data and the combination that led to the best model performance was selected. The best model performance is defined here by the highest ratio of performance to interquartile of the 10-fold cross-validation model (RPIQcv). The optimal number of latent variables in the PLS models was automatically selected according to Wold's criterion (Li et al. 2002) with an initial threshold of 0.99. When this threshold was not reached, it was automatically decreased in increments of 0.01 until the algorithm converged. For further details about PLS regression and related statistics employed in this paper, the reader is referred to Hans et al. (2017). This routine was repeated for all target variables. Again, for conciseness purposes, only the best preprocessing method and corresponding PLS model performance are reported.

The performance of the PLS cross-validation models based on the full spectral range are presented in Table 4 along with the pre-processing method used. Baseline correction and derivative pre-processing led to the best

models and these two methods often provided similar results. VIS-NIR spectra often exhibit baseline offset and curvilinear trend, likely caused by changes in illumination angle or optical path length (Rinnan et al. 2009). This is the case here since the height of the chips in the pan was not measured and may have varied between samples. This reflects the conditions at the mill as the height of the chips on the conveyor belt is not controlled. Derivative can be seen as a special form of baseline correction as first derivative removes a constant offset and second derivative removes the offset plus a linear term (i.e. linear tilting of the spectrum) (Maesschalck et al. 1999). Supplementary Table S1 also reports derivative as the preferred spectra pre-processing method for predicting wood extractives, lignin and holocellulose content. However, in contrast with Supplementary Table S1, where MSC and SNV are commonly used, these methods were not the most appropriate for the datasets of Table 4.

Table 4 also shows that the Polytec-Dry dataset always gave better results than the Polytec-Fresh and ASD-Chips datasets, even though the latter covers a larger spectral range. This demonstrates the importance of the optics in the reflectance measurement of wood chips. As mentioned earlier, the ASD-Chips spectra were acquired using the ASD turntable with simply a fiber optic and a halogen light

Table 4: PLS model performance for the prediction of wood chip brightness and chemical components according to the best pre-processing method.

Property	Dataset	Method	N	LVs	RMSE _{CV}	R _{2CV}	RPIQ _{CV}
Brightness	Polytec-Fresh	B	65	7	4.44	0.60	2.13
	Polytec-Dry	B	48	9	3.00	0.82	3.23
	ASD-Chips	B	48	4	5.09	0.50	1.91
	ASD-Powder	B	48	8	1.14	0.97	8.53
Extractives	Polytec-Fresh	FD	65	9	0.29	0.74	3.30
	Polytec-Dry	B	48	8	0.25	0.81	3.76
	ASD-Chips	SD	48	6	0.45	0.36	2.05
	ASD-Powder	SD	48	7	0.26	0.79	3.58
Lignin	Polytec-Fresh	SD	65	9	0.90	0.93	6.21
	Polytec-Dry	B	48	9	0.53	0.97	10.00
	ASD-Chips	B	48	8	1.61	0.76	3.26
	ASD-Powder	SD	48	8	0.56	0.97	9.47
Holocellulose	Polytec-Fresh	SD	65	8	0.88	0.91	5.57
	Polytec-Dry	B	48	9	0.52	0.97	9.58
	ASD-Chips	B	48	8	1.48	0.73	3.40
	ASD-Powder	B	48	7	0.56	0.96	8.96

The following pre-processing abbreviations are used: FD, first derivative, SD, second derivative, B, baseline correction.

source. In comparison, the Polytec PSS-H-A03 probe head is an optical system that provides a collimated illumination beam with the diffuse reflected light collected at a small angle intersecting the incident beam. This setup makes a significant difference by limiting variations in path length and scattering effects.

Finally, it can also be observed from Table 4 that models obtained with the Polytec-Dry dataset for wood components are nearly equivalent to those obtained with the ASD-Powder dataset. This suggests that the information contained in the 350–1100 nm and 2100–2500 nm regions is not critical for the prediction of these components. In contrast, the ASD-Powder dataset performed best for brightness because of the information provided by the VIS region.

3.2.2 Model desensitization

The Polytec-Fresh dataset is of particular interest because it includes samples with variable moisture content scanned under frozen and unfrozen conditions. However, in an industrial environment, chips could be subject to even larger changes in moisture content and temperature. In order for a calibration model to perform well in this type of environment, it is necessary to further desensitize it to moisture content and temperature disturbances (the confounding variables). Here, this was achieved using EPO (Roger et al. 2003). Briefly, EPO assumes that a spectra matrix can be decomposed into two subspaces: a useful subspace that represents the influence of the target variable and an external (parasitic) subspace that represent the

influence of the confounding variables. The goal of EPO is to remove spectral distortion caused by the confounding variable so the new (corrected) spectra contain essentially useful information. The approach and dataset used to perform this task and build the EPO matrix are described in Hans and Allison (2019). The resulting dataset was called Polytec-Fresh-EPO and the models obtained are presented in Table 5.

In comparison to the models summarized in Table 4, desensitization reduced the number of LVs for most models and decreased the RMSE_{CV} of all models except extractives. A large number of EPO components was required for each model (between 2 and 6), which reflects the strong influence of temperature and moisture content on wood NIR spectra. A similar number of EPO components was used in Hans and Allison (2019) to correct biomass spectra for the effect of temperature and moisture content on the calorific value of hog fuel. While model desensitization generally does not make much difference in the calibration process (as shown in Table 5), it is critical to the robustness of model predictions in an industrial environment.

3.2.3 Variable selection

To further simplify the calibration models developed above and enhance their robustness, three common wavelength selection methods were evaluated: variable importance in projection (VIP), genetic algorithm (GA) and interval PLS (iPLS). The reader is referred to Mehmood et al. (2012) for a description of these methods. The VIPs were selected using

Table 5: PLS model performance for the prediction of wood chip brightness and chemical components after desensitization of the Polytec-Fresh dataset against moisture and temperature.

Property	Preprocessing	N	g	LVs	RMSE _{CV}	R ₂ _{CV}	RPIQ _{CV}
Brightness	B + EPO	65	5	4	4.30	0.62	2.20
Extractives	FD + EPO	65	6	7	0.29	0.74	3.30
Lignin	SD + EPO	65	5	6	0.85	0.94	6.57
Holocellulose	SD + EPO	65	2	9	0.81	0.92	6.00

The number of EPO components used is represented by g.

successive iterations until no further improvement in model performance was observed. The GA and iPLS methods were tuned using a wide range of parameters and the ones yielding the best model were selected. The number of latent variables and model performance were evaluated as described earlier. Among the three wavelength selection methods, the genetic algorithm (GA) generally gave the best results, being outperformed by iPLS in only two cases as shown in Table 6. It can be seen that after wavelength selection, model performance improved in every case, leading to higher accuracies and also, in most cases, to simpler models (i.e. requiring fewer LVs). This agrees with Supplementary Table S1 as a large number of studies showed the usefulness of wavelength selection for improving model performance. The wavelength regions selected for each model are shown in Supplementary Figure S1.

For brightness prediction, the wood powder based model achieved a performance comparable to the models

presented in Supplementary Table S1. Most of the wavelengths selected for brightness in the ASD-Powder dataset are located in the 350–1100 nm region (Supplementary Figure S1) and the corresponding model gave by far the best accuracy. Contrastingly, a large number of wavelengths were selected in the other spectral regions for the ASD-Chips datasets. This could be explained by high light scattering from the chips which requires selecting more wavelengths across the spectrum to compensate. This may also apply to the measurement of the wood chemical compounds.

The models for wood chemical compounds showed excellent performance in comparison with the literature. Indeed, the fresh chips and wood powder based models for extractives content outperformed those reported in Supplementary Table S1. The models for lignin and holocellulose measurements were outperformed by only a few studies. Among these, one built their lignin calibration model based on extracted wood samples (Schwanninger

Table 6: PLS model performance for the prediction of wood chip brightness and chemical components after wavelength selection with GA (except for datasets marked with * where iPLS was used).

Property	Dataset	N	LVs	RMSE _{CV}	R ₂ _{CV}	RPIQ _{CV}
Brightness	Polytec-Fresh	65	4	3.49	0.75	2.71
	Polytec-Fresh-EPO	65	5	3.47	0.75	2.73
	Polytec-Dry*	48	9	2.52	0.88	3.85
	ASD-Chips	48	4	3.53	0.76	2.75
	ASD-Powder	48	6	0.90	0.98	10.80
Extractives	Polytec-Fresh	65	5	0.23	0.83	4.10
	Polytec-Fresh-EPO	65	4	0.22	0.85	4.25
	Polytec-Dry	48	4	0.19	0.89	4.95
	ASD-Chips	48	6	0.26	0.80	3.63
	ASD-Powder	48	5	0.17	0.91	5.47
Lignin	Polytec-Fresh	65	4	0.76	0.95	7.34
	Polytec-Fresh-EPO	65	4	0.55	0.97	10.14
	Polytec-Dry	48	4	0.43	0.98	12.15
	ASD-Chips	48	4	1.24	0.86	4.26
	ASD-Powder	48	5	0.40	0.99	13.15
Holocellulose	Polytec-Fresh	65	4	0.65	0.95	7.47
	Polytec-Fresh-EPO	65	5	0.61	0.95	7.96
	Polytec-Dry	48	5	0.40	0.98	12.58
	ASD-Chips	48	4	1.16	0.83	4.32
	ASD-Powder*	48	8	0.46	0.97	10.8

et al. 2011b). Prediction of lignin from extractives free wood is known to improve model accuracy because these two compounds have similar phenolic structures (Jiang et al. 2014). Two other studies measured lignin and holocellulose for species of a single genus (*Pinus*) and utilized a very similar approach to select samples (Hodge and Woodbridge 2004; Jiang et al. 2014). This consisted of sampling from different regions within the same stem, including samples of juvenile and mature wood and samples with different amounts of earlywood and latewood. This sampling approach likely explains the outstanding performance of their models because it generated high variability in lignin and holocellulose content between samples while keeping other wood characteristics homogenous.

The wavelengths selected by GA and iPLS were also in agreement with the literature (Supplementary Table S1). For extractives content, many wavelengths were selected within the 1300–1800 nm range. Most studies in Supplementary Table S1 indicate this wavelength range as the most relevant for extractives prediction, with study of He and Hu (2013) being the only exception recommending wavelengths up to 2355 nm. However, this latter study investigated a very large number of exotic species and the diversity of their extractive compounds may explain the requirement of a larger spectral range. Most of the wavelengths selected by the GA for the Polytec datasets actually fall in the 1600–1800 nm region (Supplementary Figure S1), which was associated by Park et al. (2018) with the 1st overtone of C–H bonds found in aromatic compounds such as extractives and lignin. In addition, according to Schwanninger et al. (2011a), the 1670–1690 nm region is associated with vibrations of carbonyl groups, which may be related to extractives. For lignin and holocellulose, many wavelengths across the entire spectrum were selected by the GA algorithm for the Polytec and ASD datasets. This agrees with Supplementary Table S1, which reported various regions covering the 1100–2500 nm range to be of importance for lignin and holocellulose prediction. It is worth pointing out that the 1640–1730 nm region was used in Schwanninger et al. (2011b) to predict lignin from extractives free wood. Several bands were also selected in this region for lignin (Supplementary Figure S1). According to Schwanninger et al. (2011a), many bands between 1100 and 1800 nm may be assigned to lignin, which are related to the 1st and 2nd overtone of C–H stretching, 1st overtone of O–H stretching and 1st overtone of C–H stretching in aromatic groups. In addition, bands in the 2100–2500 nm regions associated with O–H, C–H, C–O, C–C and C=O stretching were also assigned to lignin (Schwanninger et al. 2011a). Finally, it is interesting to note that, for holocellulose, several bands

were selected in the 1900–2000 nm region of the Polytec-Dry dataset while these are absent from the Polytec-Fresh selection. Water absorption bands (OH groups) can overlap those of holocellulose in this region, which can hamper the measurement of the latter. Supplementary Figure S1 illustrates this phenomenon and explains, to some extent, the better accuracies obtained with the Polytec-Dry dataset in comparison with the Polytec-Fresh.

3.3 Qualitative validation over historical data

An on-line analyzer (FITNIR MC, FITNIR Analyzers Inc., Vancouver) using a Polytec spectrometer was installed on the feed to the digester in a kraft pulp mill located in Western Canada processing both softwood and hardwood grades. Hardwood chips are a mix of A and balsam poplar (*Populus balsamifera* L.), a species closely related to BC as the two species have similar chemistry and can both hybridize (Nesom 2002). Softwood chips are a mix of spruce, pine and fir (including BS). The analyzer scans chips reclaimed from the bottom of the chip pile, after outdoor seasoning. The chips are typically subject to high variability in moisture content and temperature. Spectra acquired between October 23rd, 2019, and May 26th, 2020, were retrieved for validation purposes. The models developed based on the Polytec-Fresh-EPO dataset (Table 6) were employed for the prediction of each property using the corresponding pre-processing, desensitizing and variable selection methods, leading to the time series presented in Supplementary Figure S2. To enhance visualization, time series were first-order filtered using a time constant of 100 min. As discussed previously, the spectral range used in the Polytec spectrometer is not the most appropriate for brightness prediction, nevertheless it is shown for completeness.

In comparison with hardwood, softwood runs are characterized by an increase in lignin content and decrease in brightness, extractives and holocellulose. This agrees with the literature (Table 1) and the data reported earlier (Figure 2). In addition, within each grade, variations in wood properties can also be observed, which is consistent with the observation made in Skvaril et al. (2015b) for lignin. The sum of the chemical compounds (total in Supplementary Figure S2) are found within a narrow range ($100.01 \pm 0.22\%$) but, considering ash content, are slightly overestimated. When transferring calibration models onto an on-line analyzer, it is common that an offset needs to be adjusted to compensate for height variations between the probe head and the chip surface (which are different

between the laboratory and the mill). No attempt at correcting this offset was performed here. This shows that, while this preliminary analysis supports the quality of the measurements, a more formal quantitative validation will be required to fully validate the prediction models.

However, this qualitative validation demonstrates that each property presents variability between runs (Supplementary Figure S2). Between-run variability is quantified here as the difference between the highest and lowest average value for each property of the same grade. Hardwood generally showed less variability than softwood (with the exception of brightness), which was expected as the hardwood chip mixture is composed of only two species while softwood is a poorly defined mixture of several species. Brightness showed the highest between-run variability (2.9% for softwood and 4.7% for hardwood), followed by lignin (1.6% for softwood and 0.8% for hardwood) and holocellulose (1.5% for softwood and 1.0% for hardwood). Finally, extractives presented a smaller between-run variability with 0.3% for softwood and 0.1% for hardwood. However, it is worth pointing out that this measurement does not take into account the extractives content of residual bark that could be present in the chip furnish in small but variable amounts depending on debarking efficiency.

4 Conclusions

On-line characterization of wood chip properties is needed in the pulp and paper industry to develop advanced process control strategies, troubleshoot operations, assess the quality and value of the supply and gain a better understanding of the seasoning process. The aim of this study was to examine the use of VIS-NIR spectroscopy for on-line measurement of fresh wood chip brightness, extractives, lignin and holocellulose content while accounting for the influence of moisture content and temperature on the spectra as well as changes in chip properties occurring during seasoning.

For the purpose of generating an information-rich dataset for calibration model development, variability in wood chip properties was generated by selecting different species and through seasoning. The seasoning was mostly affected by abiotic factors and the activity of non-decaying fungi which contributed to decreased chip brightness and extractives content. Significant differences between wood species were observed for all the properties studied, but particularly for lignin and holocellulose content. Data reconciliation between the wood chemical compounds dramatically reduce the estimated SEMs of holocellulose

(by a factor 3 to 8 depending on species) while leaving extractives and lignin essentially unchanged. Two similar spectra pre-processing methods, derivative and baseline correction, were found to lead to the most accurate models for non-contact measurements of wood chips as well as for contact measurements of wood powder. Model desensitization by means of EPO resulted in more robust models (generally requiring less LVs) with little impact on prediction accuracy. Wavelength selection showed that the VIS spectral region is critical for wood brightness prediction, which is not surprising since the ISO standard uses the 457 nm wavelength as reference. The 1300–1800 nm spectral region was found to be most suitable for extractives prediction and could probably even be narrowed down to 1600–1800 nm. Regions associated with lignin and holocellulose are found across the entire 1000–2500 nm spectral range, although limiting the range to 1100–2100 nm did not appear to significantly impact model performance. The comparison of two different spectrometer and probe combinations demonstrated that, provided the most relevant spectral region is selected in calibration, a non-contact measurement of wood chips performs as well as a contact measurement of wood powder for monitoring chemical composition. This is of particular importance as a non-contact measurement greatly decreases cost and sample preparation time and, thereby, leads to frequent, low cost and accurate on-line measurement.

On-line predictions over 7 months using an NIR analyzer installed in a kraft pulp mill processing both hardwood and softwood chips revealed variability within and between runs for a given grade. Developing process control strategies able to compensate for this variability would help mills to improve product quality and reduce production costs. Taking on-line spectroscopy a step further, future experiments could investigate the potential to determine the amount of saponifiable and unsaponifiable compounds in wood as the ratio is critical to avoid problems in the pulping process. Along the same lines, determining the amount syringyl versus guaiacyl lignin (S/G or S/(S + G) ratio) might also be beneficial as syringyl-rich lignins are less recalcitrant to chemical pulping (Hodge et al. 2018). Finally, measuring independently cellulose and hemicellulose content may be of interest for bioraffineries and for optimizing the pre-hydrolysis step in dissolving pulp mills.

Acknowledgements: The authors would like to thank Stuart Spencer, John Hoffmann, Nelson Uy, David Osmond, and Jérémie Pinsonneault for helping with the experimental work, Rod Stirling for providing valuable comments, Wesley Gilbert and the mill for sharing the historical data.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: This research was financially supported by the Canadian Forest Service under the contribution agreement existing between the Government of Canada and FPInnovations.

Conflict of interest statement: The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- Achiche, S., Baron, L., Balazinski, M., and Benaoudia, M. (2005). Chips image processing to predict pulp brightness using fuzzy logic techniques. In: *Proceedings of the 91st annual meeting of the pulp and paper technical association of Canada*, February 8-10, 2005, Montréal.
- Achiche, S., Baron, L., Balazinski, M., and Benaoudia, M. (2007). Online prediction of pulp brightness using fuzzy logic models. *Eng. Appl. Artif. Intell.* 20: 25–36.
- Acquah, G.E., Via, B.K., Fasina, O.O., and Eckhardt, L.G. (2015). Non-destructive prediction of the properties of forest biomass for chemical and bioenergy applications using near infrared spectroscopy. *J. Near Infrared Spectrosc.* 23: 93–102.
- Allen, L.H. (1987). *Pitch control during the production of aspen kraft pulp*, Report MR116. FPInnovations.
- Allen, L.H., Sithole, B.B., MacLeod, J.M., Lapointe, C.L., and McPhee, F.J. (1991). The importance of seasoning and barking in the kraft pulping of aspen. *J. Pulp Pap. Sci.* 17: 85–91.
- Axrup, L., Markides, K., and Nilsson, T. (2000). Using miniature diode array NIR spectrometers for analyzing wood chips and bark samples in motion. *J. Chemom.* 14: 561–572.
- Benaoudia, M., Laperriere, L., Bedard, P., Leduc, C., and Daneault, C. (2005). *Method for estimating an optimal dosage of bleaching agent to be mixed with wood chips*, application no. CA2447098A1.
- Crowe, C.M., Campos, Y.A.G., and Hrymak, A. (1983). Reconciliation of process flow rates by matrix projection. Part I: linear case. *AIChE J.* 29: 881–888.
- Ferrero, F., Lohrer, C., Schmidt, B.M., Noll, M., and Malow, M. (2009). A mathematical model to predict the heating-up of large-scale wood piles. *J. Loss Prev. Process. Ind.* 22: 439–448.
- Gilbert, W., Trush, I., Allison, B., Reimer, R., and Mason, H. (2018). Dry chip feedrate control using online chip moisture. *Tappi J.* 17: 295–305.
- Gutiérrez, A., del Río, J.C., González-Vila, F.J., and Romero, J. (1998). Variation in the composition of wood extractives from *Eucalyptus globulus* during seasoning. *J. Wood Chem. Technol.* 18: 439–446.
- Gutiérrez, A., Río, J.C.d., González-Vila, F.J., and Martín, F. (1999). Chemical composition of lipophilic extractives from *Eucalyptus globulus* Labill. *Wood. Holzforschung* 53: 481–486.
- Gutiérrez, A., del Río, J.C., and Martínez, A.T. (2009). Microbial and enzymatic control of pitch in the pulp and paper industry. *Appl. Microbiol. Biotechnol.* 82: 1005–1018.
- Hans, G. and Allison, B. (2019). Temperature and moisture insensitive prediction of biomass calorific value from near infrared spectra using external parameter orthogonalization. *J. Near Infrared Spectrosc.* 27: 259–269.
- Hans, G., Allison, B., and Gilbert, W. (2017). On-line biomass moisture content measurement using near-infrared spectroscopy. *J-FOR* 6: 6–14.
- He, W. and Hu, H. (2013). Rapid prediction of different wood species extractives and lignin content using near infrared spectroscopy. *J. Wood Chem. Technol.* 33: 52–64.
- Heinek, S., Huber, M.B., Hofmann, A., Monthaler, G., Fuchs, H.P., Larch, C., and Giovannini, A. (2013). Degradation of biomass during the storage of woodchips. Presented at the 21th European Biomass Conference, Copenhagen: 9 p.
- Hodge, G.R. and Woodbridge, W.C. (2004). Use of near infrared spectroscopy to predict lignin content in tropical and subtropical pines. *J. Near Infrared Spectrosc.* 12: 381–390.
- Hodge, G.R., Acosta, J.J., Unda, F., Woodbridge, W.C., and Mansfield, S.D. (2018). Global near infrared spectroscopy models to predict wood chemical properties of *Eucalyptus*. *J. Near Infrared Spectrosc.* 26: 117–132.
- Hu, T.Q., Zhao, M., Bicho, P., and Losier, P. (2011). A method for estimating wood chip brightness and its applications. *Can. J. For. Res.* 41: 2114–2119.
- Jiang, W., Han, G., Via, B.K., Tu, M., Liu, W., and Fasina, O. (2014). Rapid assessment of coniferous biomass lignin-carbohydrates with near-infrared spectroscopy. *Wood Sci. Technol.* 48: 109–122.
- Jirjis, R. and Theander, O. (1990). The effect of seasonal storage on the chemical composition of forest residue chips. *Scand. J. For. Res.* 5: 437–448.
- Kaar, W.E. and Brink, D.L. (1991). Summative analysis of nine common North American woods. *J. Wood Chem. Technol.* 11: 479–494.
- Li, B., Morris, J., and Martin, E.B. (2002). Model selection for partial least squares regression. *Chemometr. Intell. Lab. Syst.* 64: 79–89.
- MacLeod, M. (2007). The top ten factors in kraft pulp yield. *Pap. Puu* 89: 1–7.
- Maesschalck, R.D., Estienne, F., Verdú-andrés, J., Centner, V., Despagne, F., Walczak, B., and Massart, D.L. (1999). The development of calibration models for spectroscopic data using principal component regression. *Internet J. Chem.* 2: 19.
- Mehmood, T., Liland, K.H., Snipen, L., and Sæbø, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometr. Intell. Lab. Syst.* 118: 62–69.
- Moksadur, R., Avelin, A., Kyprianidis, K., Jansson, J., and Dahlquist, E. (2018). Model based control and diagnostics strategies for a continuous pulp digester. In: *Proceedings of papercon, 2018, April 15–18, Charlotte, USA*.
- Nesom, G. (2002). *Black cottonwood*. Baton Rouge, LA, USA: USDA, NRCS, National Plant Data Center.
- Nilsson, D., Edlund, U., Sjöström, M., and Agnemo, R. (2005). Prediction of thermo mechanical pulp brightness using NIR spectroscopy on wood raw material. *Pap. Puu* 87: 102–109.
- Noll, M., and Jirjis, R. (2012). Microbial communities in large-scale wood piles and their effects on wood quality and the environment. *Appl. Microbiol. Biotechnol.* 95: 551–563.
- Nugent, H.M., Allen, L.H., and Bolker, H.I. (1976). *Effect of seasoning on the extractives composition of the wood from black spruce, jack pine and trembling aspen*, Report PPR152. FPInnovations.

- Oberhofnerová, E., Pánek, M., and García-Cimarras, A. (2017). The effect of natural weathering on untreated wood surface. *Maderas Cienc. Tecnol.* 19: 173–184.
- Park, S.-Y., Kim, J.-C., Yeon, S., Yang, S.-Y., Yeo, H., and Choi, I.-G. (2018). Rapid prediction of the chemical information of wood powder from softwood species using near-infrared spectroscopy. *BioResources* 13: 2440–2451.
- Ramnath, L., Sithole, B., and Govinden, R. (2018). The effects of wood storage on the chemical composition and indigenous microflora of Eucalyptus species used in the pulping industry. *BioResources* 13: 86–103.
- Rinnan, Å., Berg, F.v.d., and Engelsen, S.B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem. (Reference Ed.)* 28: 1201–1222.
- Roger, J.-M., Chauchard, F., and Bellon-Maurel, V. (2003). EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemometr. Intell. Lab. Syst.* 66: 191–204.
- Rowell, R. (2005). *Handbook of wood chemistry and wood composites*. Boca Raton: CRC Press.
- Schanda, J. (2007). *Colorimetry: understanding the CIE system*. Hoboken: John Wiley & Sons.
- Schwanninger, M., Rodrigues, J.C., and Fackler, K. (2011a). A review of band assignments in near infrared spectra of wood and wood components. *J. Near Infrared Spectrosc.* 19: 287–308.
- Schwanninger, M., Rodrigues, J.C., Gierlinger, N., and Hinterstoisser, B. (2011b). Determination of lignin content in Norway spruce wood by Fourier transformed near infrared spectroscopy and partial least squares regression. Part 1: wavenumber selection and evaluation of the selected range. *J. Near Infrared Spectrosc.* 19: 319–329.
- Sjöström, E., and Alén, R. (Eds.) (1999). *Analytical methods in wood chemistry, pulping, and papermaking*. Berlin: Springer-Verlag.
- Skvaril, J., Kyprianidis, K., Avelin, A., Odlare, M., and Dahlquist, E. (2015a). Fast determination of lignin content in feedstock material for pulping process monitoring and optimization. In: *Proceedings of the 8th International Conference on Advanced Vibrational Spectroscopy*, July 12–17, 2015, Vienna.
- Skvaril, J., Kyprianidis, K., Avelin, A., Odlare, M., and Dahlquist, E. (2015b). Multivariate analysis models for wood properties combined with Open Modelica model for process performance monitoring. In: *Proceedings of the 8th Vienna International Conference on Mathematical Modelling*, February 18–20, 2015, Vienna.
- Stirling, R.A. (2005). *Detection and evaluation of decay in pulp and paper fibre supplies*, Ph.D. thesis. University of British Columbia.
- Tappi standard (2014). *T 257. Sampling and preparing wood for analysis*. Tappi.
- Tarasov, D., Leitch, M., and Fatehi, P. (2018). Lignin-carbohydrate complexes: properties, applications, analyses, and methods of extraction: a review. *Biotechnol. Biofuels* 11: 269.
- Trung, T., and Allison, B. (2015). Advanced online process analyzer for chemical recovery and pulp mill control. *O Papel*. 76: 47–56.
- Tsuchikawa, S. (2007). A review of recent near infrared research for wood and paper. *Appl. Spectrosc. Rev.* 42: 43–71.
- Tsuchikawa, S., and Kobori, H. (2015). A review of recent application of near infrared spectroscopy to wood science and technology. *J. Wood Sci.* 61: 213–220.
- Weigel, G., Bicho, P., and Watson, P. (2002). *Seasonal variations in the extractives contents of four British Columbia softwoods*, Report PRR1590. FPInnovations.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/hf-2021-0027>).