Results in Chemistry 4 (2022) 100496

Contents lists available at ScienceDirect

Results in Chemistry

journal homepage: www.sciencedirect.com/journal/results-in-chemistry

Comparison of calibration models for rapid prediction of lignin content in lignocellulosic biomass based on infrared and near-infrared spectroscopy

Kristoffer Mega Herdlevær *, Camilla Løhre, Egil Nodland, Tanja Barth

Department of Chemistry, University of Bergen, Allégaten 41, N-5007 Bergen, Norway

ARTICLE INFO

Keywords:
Biomass
Lignin
FT-IR
NIR
PLS regression
PCA

ABSTRACT

Lignocellulosic biomass has great potential as a renewable energy source due to abundance both as vegetation and in residues from agriculture and forestry. It can be utilized for biofuels and for value-added chemicals and materials. The challenge concerning lignocellulosic biomass lies in its heterogenous nature. The chemical composition may vary according to species, location, harvest season and botanic fractions. Therefore, it is crucial to assess the composition prior to any biofuel conversion method. Vibrational spectroscopy has been used extensively for rapid predictions of the chemical properties of biomass. In this study, calibration models based on infrared and near-infrared spectra has been compared using the same calibration set, 36 samples comprising wood, bark, needles, and twigs of three different wood species. PLS was performed to correlate EMSC pretreated infrared and near-infrared spectra to the chemical contents of the samples. R^2 for the standard curve of the infrared model is 0,896 and for the near-infrared standard curve 0,921, using 3 PLS components. The effect of heterogeneity was tested by comparing calibration models based on finely and coarsely ground sample, where the R^2 of the coarsely ground sample was 0,825, lower than 0,896, but still significant. The results also show that there is more variation in the lignin content between the fractions from a single tree than between similar fractions from different species. The calibration models that have been developed will be useful for frequent, rapid determination of lignin content in wood biorefining feedstocks.

Introduction

Biomass is emerging as the primary renewable source of organic carbon compounds to replace petroleum-based products in a future circular economyBiomass can be converted to liquid fuels, and into bulk and fine chemicals to replace products that at present are produced from fossil carbon sources. However, even with a great potential for future sustainable use, the economics of production and refining at present is a barrier for the development of biomass use in the overall economy [1]. Thus, research is needed to provide economically viable and sustainable biomass valorization concepts.

Lignocellulosic biomass has great potential to be a renewable energy source due to a wide abundance both as standing vegetation and as residues from agriculture and forestry [2]. It can be utilized for both biofuels and for value-added chemicals and materials. The challenge concerning lignocellulosic biomass lies in its heterogenous nature. The chemical compositions may vary according to species, location, harvest season and botanic fractions [3]. Therefore, it is crucial to assess the composition prior to any biofuel conversion method. The conventional assessment procedure of lignocellulosic biomass includes decomposing the sample, which is time consuming, labor-intensive, expensive and destroys the sample. Vibrational spectroscopy has proven effective to predict the composition of lignocellulosic biomass using models based on a calibration set [4].

Vibrational spectroscopy is a useful analytical technique that is sensitive to functional groups present in organic matter. These techniques have been utilized to determine the lignin content in wood, pulp, paper and plants [5]. Early studies were limited by the classical dispersive methods which produced low resolution spectra and a low signal-to-noise ratio. Fourier transform spectrometers were later developed and provides wide range spectra with high resolution that are rapidly obtained [6]. This is done by measuring all frequencies simultaneously, as opposed to dispersive methods which only provides individual single-frequency scans. Using Fourier transform spectroscopy, there is no degradation of optical throughput, which provides higher resolution without compromising on signal-to-noise ratio. The vibrational spectroscopic techniques included in this paper is Fourier transformed infrared (FT-IR) and near-infrared (NIR). For FT-IR, an

* Corresponding author.
E-mail address: kristoffer.herdlever@uib.no (K.M. Herdlevær).

https://doi.org/10.1016/j.rechem.2022.100496
Received 28 April 2022; Accepted 24 August 2022
Available online 28 August 2022

attenuated total reflectance (ATR) attachment built into the instrument allows for measurements without complex sample preparation. ATR allows for enhanced band intensities at the lower wave numbers.

A few applications where FT-IR and NIR spectroscopy have been used for lignin quantification have already been described. Sanderson et al, (1996) used NIR spectroscopy for compositional analysis of biomass feedstocks. They investigated a sample set comprising 121 samples of woody and herbaceous feedstocks and generated a calibration model using PLS modeling. The accuracy of their predictions of lignin content is reported to be very promising, with an R^2 of 0,98 and a standard error of performance (SEP) of 0,70. Tamaki et al, (2010) [7] used FT-IR to determine lignin content in triticale (67 samples) and wheat (47 samples). Their model based on PLS yielded an R^2 of 0,985 and a root-mean-square error of prediction (RMSEP) of 0,163.

Thus, though FT-IR and NIR have been used for biomass characterization, they have not previously been compared based on the same sample set. In this work, the predictive power of two different models based on infrared and near-infrared spectroscopy and their combinations are evaluated and considered with respect to lignin content, a large-scale scenario, tolerance to heterogeneity, and effectiveness. Here, we establish a calibration model for predicting the lignin content of woody biomass based on both IR and NIR spectra of powdered samples. FT-IR is included because the spectral elements can be interpreted in terms of chemical functional groups, compared to NIR which may give more precise models but where interpretation in terms of chemical structures is not an option. To cover a sufficient range of variability in the calibration model, samples from three different species of trees; spruce, birch and pine, three different trees for each specie, growing naturally on the coast of Western Norway are further fractionated into heartwood, bark, twigs, and leaves/needles, yielding a total of 36 different samples. The fractionation step is to secure a wide range of variability in the calibration set, as it is assumed that the variability can be as large between these fractions as between tree species [8]. Another advantage of fractionating the samples is an easier comparison with industrial applications. The wood samples will be comparable to waste from sawmills to produce high-value products and bark samples will be comparable to bark waste from forestry. Next, chemical degradation of each sample is performed to determine the lignin content. The spectra are then used as dependent variables to predict the different ratios of constituents in the biomass.

Materials and methods

To cover a large range of variability in the calibration model, samples from three different trees; spruce, birch and pine growing naturally on the coast of Western Norway (Radøy) were collected in September 2020. These species were chosen as they are most relevant in terms of abundancy and upscaling. The samples were taken from young, small trees, and manually separated into wood, bark, twigs, and leaves/needles as seen in Fig. 1, yielding a total of 36 different samples.

The samples were oven-dried at 105 °C until the change in weight did not exceed 1 %, measured every 24 h [9]. The air-dried samples were then milled and sieved with a 500 mesh to achieve a homogenous particle size. A high lignin content was expected for all samples as lignin content in trees tend to be anticorrelated with age [10].

Fig. 1. From left to right; ground heartwood, bark, needles, and twigs.

NIRSystems^{TM} Holographic Grating model 6500 were used to collect NIR diffuse reflectance spectra, averaging 32 scans within 1100–2500 nm with a resolution of 2 nm. FT-IR spectra were collected with a Nicolet iS 50R FTIR Spectrometer equipped with an ATR diamond. The spectral range for the spectrometer is 4000–400 cm^{-1} and a spectral resolution better than 2 cm^{-1}.

Reference analysis

After spectral acquisition, the samples were subjected to the wet chemistry procedure used to determine the content of extractives, ash, lignin, and carbohydrates. The ash content was measured in accordance with the NREL procedure "Determination of Ash in Biomass, NREL/TP-510-4262" [11], which involved subjecting a weighted oven-dry sample to a 575 °C furnace overnight. Extractives was measured by a procedure involving Soxhlet extraction "Determination of Extractives in Biomass, NREL/TP-510-42619" [12] with ethanol as the solvent to extract ethanol soluble material. Lignin content was determined by a two-step acid hydrolysis process based on the NREL procedure "Determination of Structural Carbohydrates and Lignin in Biomass, NREL/TP-510-42618" [13]. Total lignin content consists of acid soluble and insoluble lignin. The whole procedure is summarized in Fig. 2. Two replicates were made for all samples to determine reliability of the procedure.

Data treatment

Extended multiplicative signal correction (EMSC) is a pretreatment method used to remove spectral variation due to physical interferences [14]. This method was chosen for both infrared and near-infrared spectra to yield better predictions. Without any correction for light scattering variation, a highly complex calibration model would have been needed. The effects of this pretreatment procedure are illustrated in Fig. 3.

The spectral information was further investigated by performing a principal component analysis (PCA) to reduce the dimensionality for better interpretation of the data with minimal information loss [15]. This is done by creating new, uncorrelated variables, termed principal components (PC), that successively explain maximum variance to avoid statistical information loss. PC1 is the principal component explaining the largest amount of variation, PC2 second largest etc. In the case of spectral data, spectra are easily comparable by plotting PCs against each other, creating 2-dimensional planes named score plots. Score plots reveal information like similarities between spectra and potential outliers.

Partial least squares regression (PLS) was then used to correlate the spectral data with the chemical constituents (w/w%) of the tree samples. PLS is a method for relating a data matrix of independent variables to a response vector or matrix. Relating the independent variables to a response vector often result in a better model in comparison to decomposing a response matrix. The drawback by doing this is the requirement of one model per response vector. It differs from the more traditional multiple linear regression as PLS can handle collinearity in the independent variables [16]. Infrared spectroscopy produces spectra with highly collinear variables where PLS is a suitable regression method.

The resulting PLS models were validated by evaluating the correlation coefficient, R^2, and the root mean square error of leave-one-out cross validation (RMSECV). For further evaluation, a validation set was made from replicates of 9 of the samples. In addition, The American Association of Cereal Chemists (AACC) has established a score for evaluating a model's performance, termed R/SEP. R is the range of the validation set and SEP is the standard error of performance [17]. Any model with R/SEP $>$ 4 is qualified for screening calibration, between 4

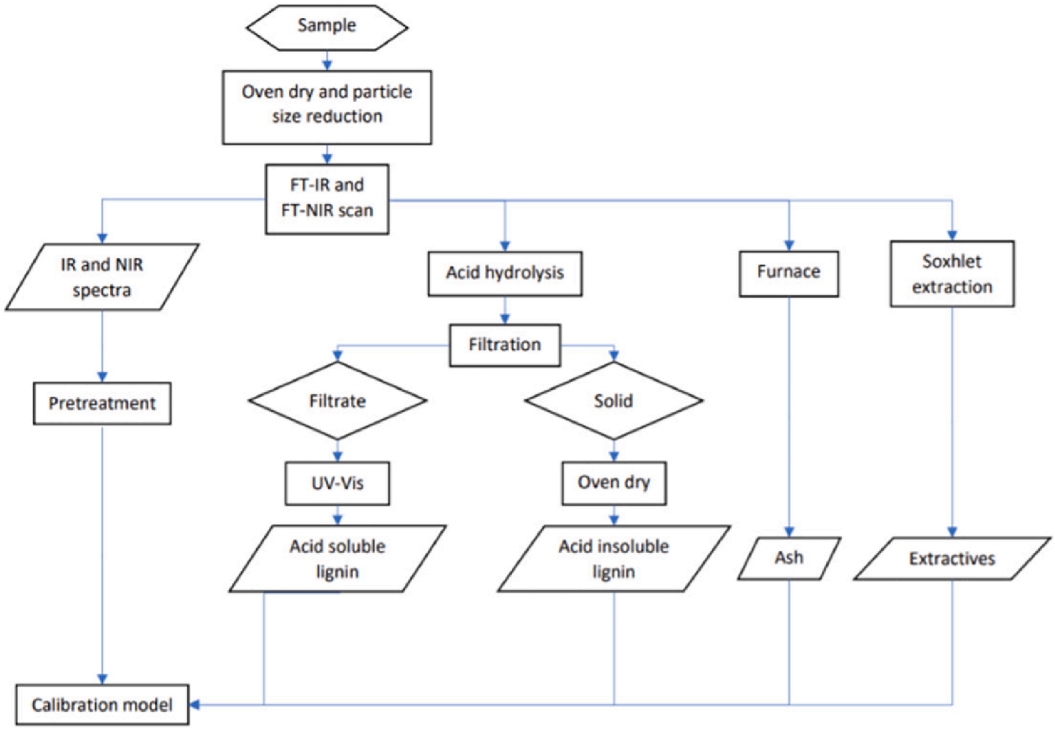

Fig. 2. Flowchart of the whole procedure, including wet chemistry analysis and spectral acquisition.

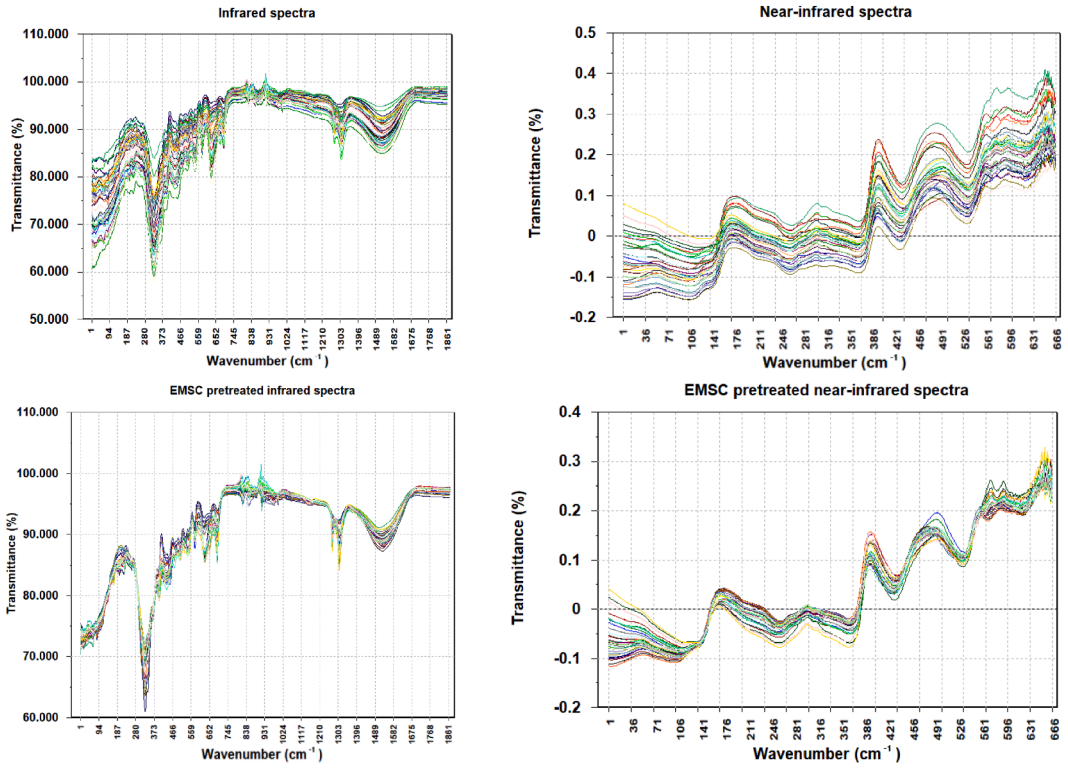

Fig. 3. Infrared and near-infrared spectra before and after EMSC pretreatment.

and 10 is acceptable for quality control, while between 10 and 15 is suitable for research quantification.

Results and discussion

The results from the wet chemistry procedures are presented in Table 1. The result shows the contents of the different constituents of every sample. In terms of lignin content, it is clear that the largest

Table 1
Chemical composition of tree species and fractions. The first character in the label code represents tree species. P = Pine, S = Spruce and B = Birch. The second character represents fraction, where B = bark, W = wood, N = needles/leaves and T = twigs. The numbers represent individual trees. Two replicates for each sample were used to calculate the uncertainly. No replicates were taken for ash, and the carbohydrate fraction is just calculated based on the other fractions.

	Ash (wt%)	Extractives (wt%)	Lignin (wt%)	Carbohydrates (wt%)
PB1	2,34	11,33 ± 0,72	41,01 ± 4,59	45,32
PB2	1,97	10,28 ± 0,09	52,33 ± 3,23	35,42
PB3	3,58	5,53 ± 0,15	51,27 ± 4,01	39,62
SB1	3,09	15,47 ± 0,47	37,82 ± 4,58	43,61
SB2	3,11	12,98 ± 0,40	39,89 ± 3,50	44,02
SB3	3,22	13,52 ± 0,61	43,02 ± 0,92	40,24
BB1	1,63	10,65 ± 0,52	50,81 ± 0,70	36,92
BB2	1,57	12,02 ± 0,02	49,65 ± 4,89	36,76
BB3	1,74	16,58 ± 0,74	50,73 ± 3,91	30,95
PW1	0,54	2,33 ± 1,22	31,80 ± 2,89	65,33
PW2	0,39	3,46 ± 1,32	30,58 ± 4,93	65,57
PW3	0,44	1,64 ± 0,36	32,08 ± 3,80	65,83
SW1	0,24	6,25 ± 0,95	30,24 ± 1,40	63,27
SW2	0,65	4,53 ± 1,71	33,93 ± 0,89	60,89
SW3	0,70	4,75 ± 1,52	34,41 ± 0,70	60,14
BW1	0,65	3,71 ± 0,67	27,37 ± 1,77	68,26
BW2	0,76	4,20 ± 0,54	29,04 ± 3,21	66,00
BW3	0,58	5,34 ± 0,46	26,93 ± 4,86	67,14
PN1	2,55	8,23 ± 1,85	37,89 ± 1,76	51,33
PN2	2,85	8,04 ± 0,40	40,92 ± 3,96	48,19
PN3	2,50	11,22 ± 1,29	39,95 ± 2,28	46,33
SN1	5,20	9,89 ± 1,33	47,69 ± 2,11	37,23
SN2	3,43	7,21 ± 1,95	43,61 ± 3,32	45,74
SN3	5,14	6,50 ± 0,12	38,11 ± 2,45	50,25
BN1	1,98	18,65 ± 0,26	44,75 ± 2,07	34,62
BN2	5,64	11,55 ± 1,35	49,97 ± 2,17	32,84
BN3	7,34	10,14 ± 1,60	52,06 ± 2,52	30,46
PT1	1,98	10,94 ± 4,76	43,02 ± 1,55	44,06
PT2	1,91	9,57 ± 2,97	47,73 ± 0,57	40,79
PT3	1,90	12,69 ± 5,36	42,60 ± 0,42	42,82
ST1	1,16	7,64 ± 2,78	41,14 ± 0,66	50,06
ST2	1,43	4,87 ± 0,98	41,85 ± 1,48	51,85
ST3	2,34	4,67 ± 1,48	44,35 ± 0,60	48,64
BT1	1,59	8,67 ± 4,78	49,93 ± 0,31	39,80
BT2	2,33	18,71 ± 6,53	44,57 ± 2,36	34,39
BT3	1,65	15,35 ± 4,25	39,74 ± 3,00	43,26

variation in lignin content lies between the different fractions, with the largest variation of 33,6% between the average of bark and the average of wood. The largest variation regarding individual trees lies between spruce and birch with 11,3 %. The smallest variation is observed between species with 6,7%, indicating that there is a larger variation between individual trees, independent of species. This is significant in a biorefinery concept, as feedstock screening should focus on fractions more than tree species.

Spectral analysis

Fig. 4 shows infrared spectra of pine samples. The different fractions differentiate the most at the peaks of the different functional groups. The spectra clearly support that wood contains the highest amount of carbohydrates, as the peaks of hydrogen bonded OH groups and particularly the peak for C—O—C is stronger than the others. The peaks for CH groups are less defined in comparison to the other fractions, which shows two distinct peaks. CH_2 is more prominent in carbohydrates, whereas CH_2 and CH_3 groups are more equally present in other chemical constituents, yielding two defined peaks. C═O is another peak that differs in the fractions. This can also be tied to carbohydrates, which does not contain any C═O groups, while other chemical constituents do. The near-infrared spectra are not as easy to interpret without the use of multivariate methods. A selection of typical spectra shown in Fig. 3 for a general overview.

PCA of EMSC pretreated spectral data is performed. The transformed

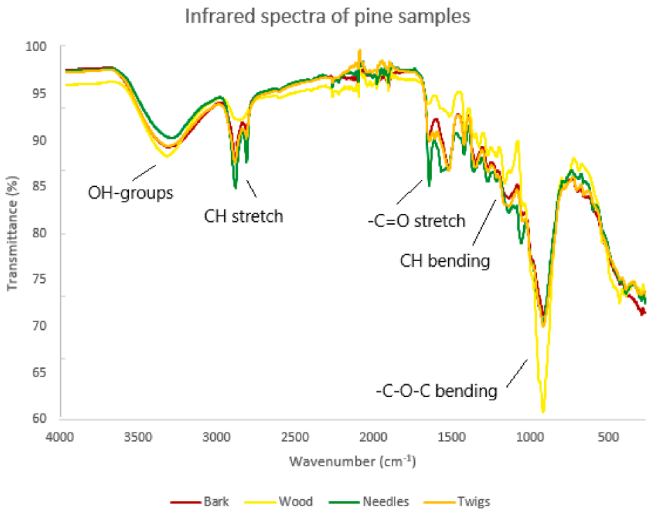

Fig. 4. Infrared spectra of samples of pine bark, wood, needles, and twigs from sample PB1, PW1, PN1 and PT1 in Table 1.

information is presented as score plots in Fig. 5, plotting principal component 1 (PC1) against PC2. These score plots can explain the maximum variance of the original data, as PC1 and PC2 are the principal components explaining most of the variance. PC1 and PC2 for the infrared model explains 69,3% and 9,5% of the variance, summed to 78,8%. PC1 and PC2 for the near-infrared model covers 75,5% and 14,0% of the variance, summed to 89,5%. As seen in Fig. 5, the infrared spectra reveals that the samples of the same biomass fraction tend to group together. This supports the observed variation in the chemical components, which indicates that there is larger variation between fractions than tree species. The score plot also reveals that the wood samples have the largest leverage on the model, being furthest away from the origin of the plot. In general, well defined and separate groupings indicates chemical differences between the group. Very different groups may not be suitable for incorporation into the same model.

As for the near-infrared spectra, there are no recognizable groupings. This indicates that near-infrared spectra alone are not capable of differentiating between fractions but is more suitable for a calibration model. The first principal components of the near-infrared score plot also covers more of the variance.

Relative standard deviation (RSD) vs leverage is a useful plot to identify statistical outliers in the model. RSD alone can be an indicator of an outlier, but that does not necessarily entail a bad model. If some spectra have a high RSD and a high level of leverage on the model, exclusion must be considered. Fig. 6 shows that no particular spectrum has a very high leverage when considering PC1, PC2 and PC3. This indicates that the model does not contain any outliers.

Plotting loadings and against wavenumber yields information of which parts of the spectra that are significantly different. Generally, loadings can be interpreted as the coefficients of the linear combination of the initial variables from which the principal components are constructed. As shown in Fig. 4, the peaks for C—O—C, OH groups, CH groups and C═O groups differed between the fractions. By plotting loadings in the same spectra, it is clearly shown where the largest variations lie. In the case of infrared spectra, loadings 1, 2 and 3 all show that the largest variations lie within the peaks of these functional groups, shown in Fig. 7. Information of functional groups is an advantage IR holds over NIR, as NIR-spectra are harder to interpret qualitatively.

Establishing a calibration model

PLS is performed to correlate EMSC pretreated infrared and near-infrared spectra to the chemical contents of the samples. The

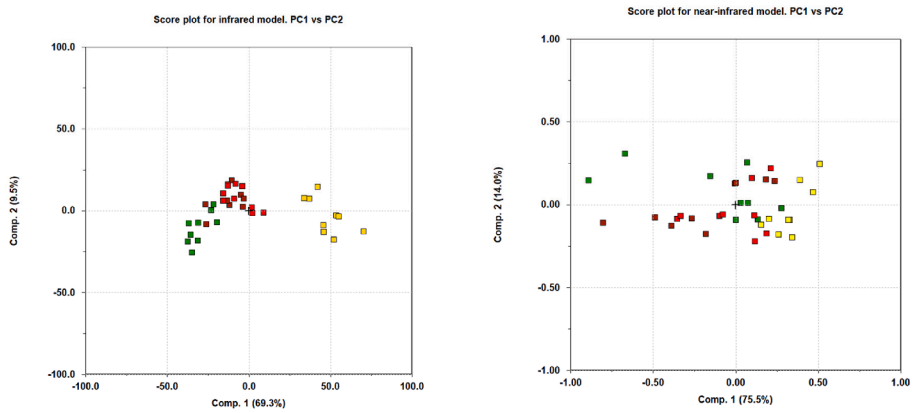

Fig. 5. Score plot of PC1 and PC2 from PCA on infrared (left) and near-infrared spectra (right). Yellow = wood, brown = bark, red = twigs and green = needles/leaves. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

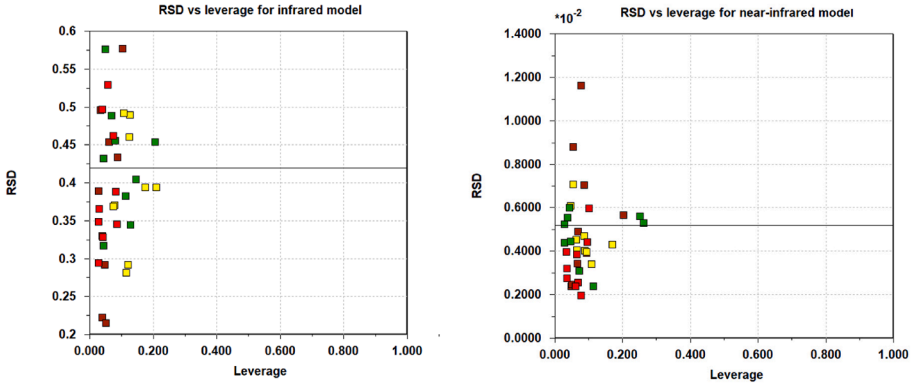

Fig. 6. RSD vs leverage for infrared (left) and near-infrared (right) model. Yellow = wood, brown = bark, red = twigs and green = needles/leaves. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

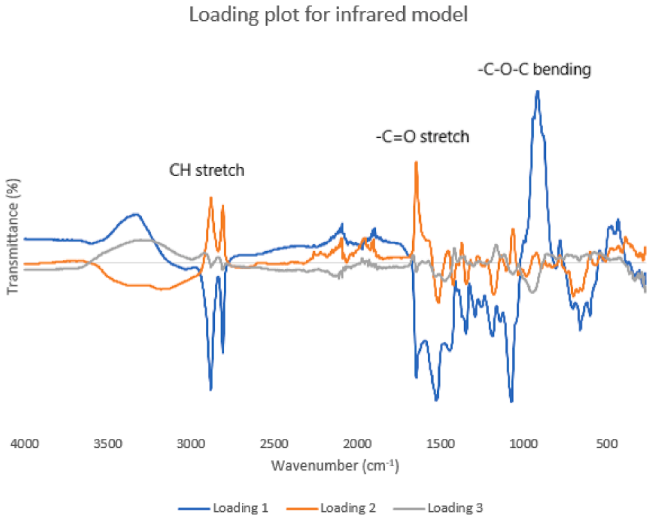

Fig. 7. Loading 1, 2 and 3 plotted against wavenumber to visualize which peaks yield the most important information for the infrared model.

calibration curves are shown in Fig. 8. The coefficient of determination, R^2 for the standard curve of the infrared model is 0,896 and for the near-infrared standard curve 0,921, using 3 PLS components. This shows that a high correlation is achieved with relatively few samples, and both spectroscopic techniques show good predictive abilities. The near-infrared model seems to have a slight edge, but more validation is needed.

Response residuals shows how each object deviates from the calibration curve presented in Fig. 9. Only one object has a response residual greater than 5 % for the infrared model and none for the near-infrared model.

The root mean square error of cross validation (RMSECV) is an estimate of the model performance based on how well the model predicts the values obtained from cross-validation. RMSECV for the infrared model is 3,29 wt% and 2,21 wt% for the near-infrared model. This shows that there is a lower mean error associated with the near-infrared model. A validation set consisting of 9 random replicate samples was incorporated into the near-infrared model. The square error of performance (SEP) is 3,73 % and the root mean square error of performance (RMSEP) is 3,95. With R/SEP = 6,38, the model falls in the range that is considered suitable for quality control according to AACC. In the work of Tamaki et al, (2010) and Sanderson et al, (1996), the reported R/SEP was 10,67 and 11,86, respectively.

Prediction of other chemical constituents

Calibration models were made to predict other constituents than lignin as well. Most of these fell under an R^2 of 0,7 with 3 PLS components, deeming the model unviable for comparison purposes. This contradicts the results reported for "Determination of Extractives in Biomass, NREL/ TP-510–42619" [11], where calibration models based on near-infrared spectra are shown to be suitable for predicting the contents of extractives in lignocellulosic biomass. However, these studies include more samples, resulting in more robust models. For the samples investigated here, a noteworthy model was the prediction of ash

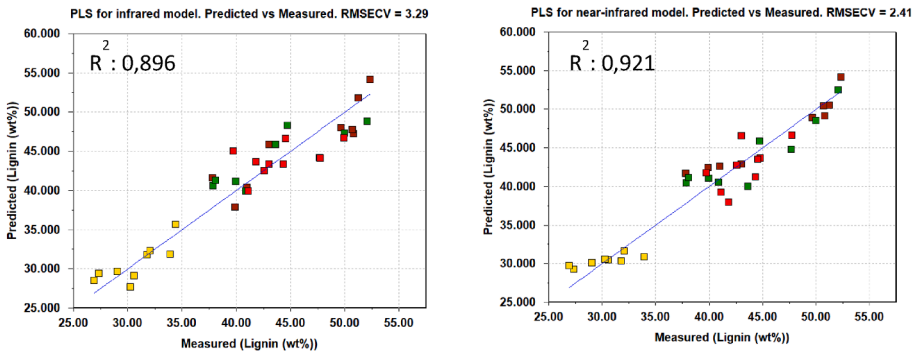

Fig. 8. Predicted values by the PLS model plotted against measured values for infrared (left) and near-infrared (right) model. Yellow = wood, brown = bark, red = twigs and green = needles/leaves. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

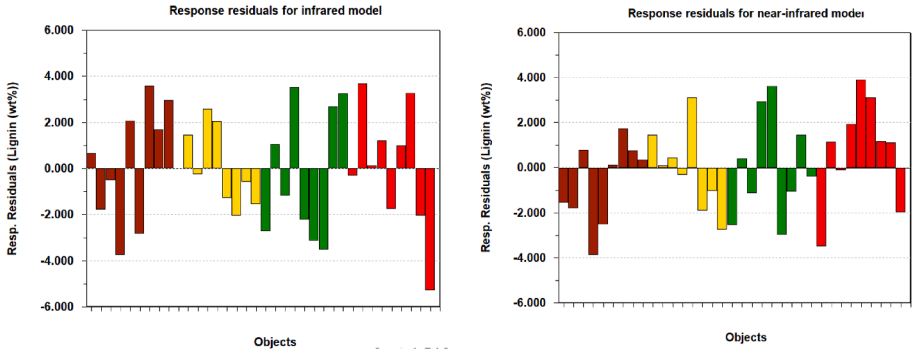

Fig. 9. Response residuals for the PLS-model for infrared (left) and near-infrared (right). Yellow = wood, brown = bark, red = twigs and green = needles/leaves. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

content based on the infrared spectra, showed in Fig. 10. This calibration model achieved an R^2 of 0,815, with RMSECV of 1,16.

Heterogeneity

Particle size of the samples has been shown to have great impact of results regarding the wet chemistry analysis chemical degradation and spectroscopic information. By using spectra from coarsely ground sample (approximately 0,3 cm^3) to build the model shown in Fig. 11, the R^2 for the calibration curve was 0,825, notably lower than 0,896 for the model using more finely ground (0,5 mm^3 sieved with a 500 mesh) sample. Although it yields lower prediction power, an R^2 of 0,825 is still significant. According to NRELs procedure "Determination of Structural Carbohydrates and Lignin in Biomass, NREL/TP-510–42618" [12] particles sized too big and too small yields a higher content of lignin, which

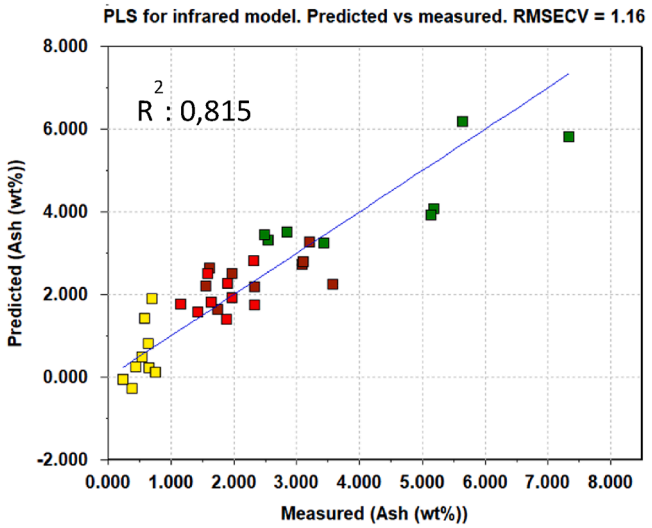

Fig. 10. PLS model that predicts ash content based on infrared spectra. Yellow = wood, brown = bark, red = twigs and green = needles/leaves. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

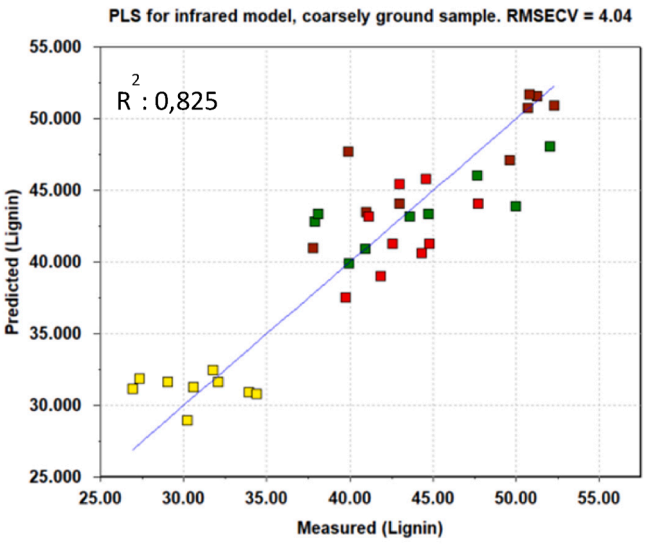

Fig. 11. PLS model of coarsely ground sample that predicts lignin content based on infrared spectra. Yellow = wood, brown = bark, red = twigs and green = needles/leaves. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

provides uncertainty within the procedures. As for spectroscopy, particle size greatly affects the spectra of the samples, and the differences is directly measurable by acquiring both IR and NIR spectra of the same sample set. As near-infrared spectra are measured at the surface of the sample, different particles meet the near-infrared probe. This was handled by averaging 6 spectra measured after stirring the sample vial. The same problem occurred for infrared measurements. As the ATR-diamond has a very small surface area, only a small portion of the sample was measured, and was not necessarily representative for the whole sample. This was handled by averaging 6 spectra. Fig. 12, a score plot containing spectra of both finely and coarsely ground samples, reveals no clustering. This indicates that particle sizes within these limits might be included in the same model and a reasonable calibration may be performed. Precise monitoring of particle size in a large-scale scenario may not be necessary, but an evaluation of spectroscopic techniques and their resilience to heterogeneity in the samples is indicated.

Overview and model evaluation

Table 2 presents a summary of the statistical values for every model. As the model based on NIR shows the best R^2 and RMSECV, the statistical values produced from the validation set, SEP and RMSEP, shows that the model based on IR might be a more robust model. The R/SEP of the IR model is above 10, and thus meets the criteria for research quantification. As for the models which predicts the ash content and lignin from coarsely ground samples, no validation set was made and needs to be investigated further.

Conclusion

Results from the wet chemistry procedure showed that there is more variation between individual trees than it is for tree species in terms of lignin content. This is significant in a biorefinery concept, as feedstock screening should focus on fractions rather than tree species. Infrared spectra showed that differences between the fractions were clearly observable, as OH and C—O—C groups were most prominent in wood samples and CH_2, C═O and the lack of C—O—C were observed in the other fractions. This is due to higher carbohydrate content in the wood samples. PCA showed that scores from infrared spectra were more grouped together in their respective fractions, while scores from near-infrared were more convoluted. R^2 for the standard curve made by PLS of the infrared model is 0,896 and for the near-infrared standard curve 0,921, using 3 PLS components. RMSECV also substantiated that the near-infrared model yielded the best results in terms of prediction power. It is worth noting however, that prediction of the wood samples seems to be less precise compared to the rest of the model. The effect of heterogeneity was tested by comparing calibration models based on finely and coarsy ground sample, where the R^2 of the coarsely ground sample was 0,825, lower than 0,896. For use as a rapid measurement of feedstock in a biorefinery concept, this information will be relevant for choosing the right instrumentation and calibration model.

A user friendly, open access tool for lignin predictions based on the sample set presented in this paper will be published for easy quality control predictions of lignin content in lignocellulosic biomass.

CRediT authorship contribution statement

Kristoffer Mega Herdlevær: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Camilla Løhre:** Supervision, Writing – review & editing, Conceptualization, Validation. **Egil Nodland:** Supervision, Writing – review & editing, Conceptualization, Validation. **Tanja Barth:** Supervision, Writing – review & editing.

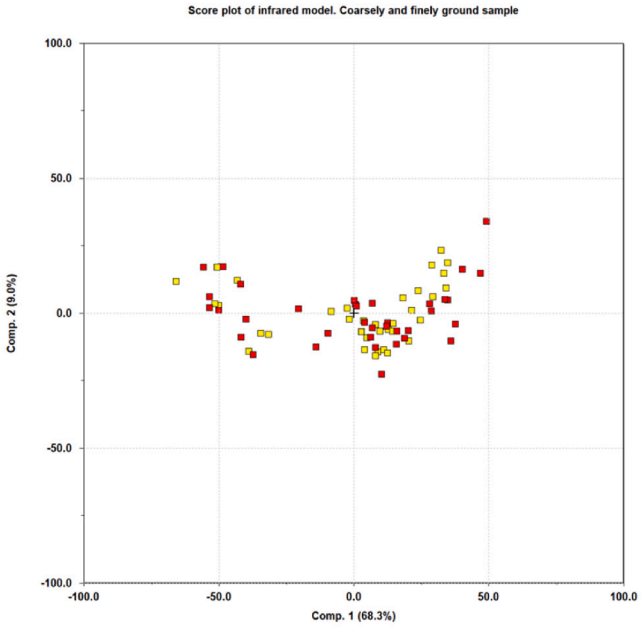

Fig. 12. Score plot of PC1 and PC2 based on infrared spectra from both finely and coarsely ground sample. Yellow = finely ground sample and red = coarsely ground sample. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Statistics for model evaluation and cross validation for every model covered. NIR covers lignin predictions based on NIR spectra, IR covers lignin predictions based on IR spectra, Ash covers ash predictions based on IR spectra and Heterogenous covers lignin predictions on coarsely ground samples based on IR spectra.

Models	R^2	RMSECV (%)	St. dev	SEP (%)	RMSEP (%)	R/ SEP
NIR	0,921	2,41	2,41	3,73	3,95	6,38
IR	0,896	3,29	3,41	2,37	3,19	10,72
Ash	0,815	1,16	1,34	–	–	–
Heterogenous	0,825	4,04	4,10	–	–	–

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank AT skog for providing help inquiring raw material for this work. We would also like to thank the research group for providing valuable input and Climate and Energy Transition at UiB for funding the project.

References

[1] J. Sherwood, The significance of biomass in a circular economy, Bioresour. Technol. 300 (2020).
[2] A.J. Ragauskas, et al., Lignin valorization: improving lignin processing in the biorefinery, Science 344 (6185) (2014) 1246843.
[3] M.R. Roger, et al., Cell Wall Chemistry. Handbook of Wood Chemistry and Wood Composites, CRC Press, 2012.

[4] M.A. Sanderson, et al., Compositional analysis of biomass feedstocks by near infrared reflectance spectroscopy, Biomass Bioenergy 11 (5) (1996) 365–370.
[5] Near infrared spectroscopy for the analysis of wood pulp: quantifying hardwood - softwood mixtures and estimating lignin content, TAPPI J., October 1990, Vol. 73 (10).
[6] P. Griffiths, J.A. de Hasseth, Fourier Transform Infrared Spectrometry, 2nd ed., Wiley-Blackwell, 2007. ISBN 978-0-471-19404-0.
[7] Y. Tamaki, et al., Rapid determination of lignin content of straw using Fourier transform mid-infrared spectroscopy, J. Agric. Food Chem. 59 (2) (2010) 504–512.
[8] R. Rowell, Handbook Of Wood Chemistry And Wood Composites, CRC Press, 2005.
[9] Preparation of Samples for Compositional Analysis. NREL/TP-510-42620. National Renewable Energy Laboratory. 2008.
[10] A.L. Healey, et al., Effect of aging on lignin content, composition and enzymatic saccharification in Corymbia hybrids and parental taxa between years 9 and 12, Biomass Bioenergy 93 (2016) 50–59.
[11] Determination of Ash in Biomass, Technical Report NREL/TP-510-42622. National Renewable Energy Laboratory. 2008.
[12] Determination of Extractives in Biomass, NREL/TP-510-42619. National Renewable Energy Laboratory. 2008.
[13] Determination of Structural Carbohydrates and Lignin in Biomass, NREL/TP-510-42618. National Renewable Energy Laboratory. 2012.
[14] H. Martens, et al., Light scattering and light absorbance separated by extended multiplicative signal correction. application to near-infrared transmission analysis of powder mixtures, Anal. Chem. 75 (3) (2003) 394–404.
[15] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 374 (2065) (2016) 20150202.
[16] S. Wold, et al., PLS-regression: a basic tool of chemometrics, Chemometrics and Intelligent Laboratory Systems 58 (2) (2001) 109–130.
[17] AACC. (1999). Approved methods of the American Association of Cereal Chemists, Method 39-00. Near-infrared methods—Guidelines for model development and maintenance. St. Paul, MN: AACC Press.