Talanta 138 (2015) 263–272

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Multivariate analysis of coconut residues by near infrared spectroscopy

M.K.D. Rambo [a,b], A.R. Alves [c], W.T. Garcia [c], M.M.C. Ferreira [a,*]

[a] Institute of Chemistry, University of Campinas – UNICAMP, CEP 13083-970 Campinas, Brazil
[b] Department of Chemistry, University of Tocantins – UFT, CEP 77838-824 Araguaína, Brazil
[c] Sugarcane Research Center – CTC, CEP 13418-900 Piracicaba, Brazil

ARTICLE INFO

Article history:
Received 2 December 2014
Received in revised form
6 March 2015
Accepted 7 March 2015
Available online 12 March 2015

Keywords:
Biomass residue
Chemical composition
Chemometrics
Figures of merit
Monosaccharide

ABSTRACT

Near infrared (NIR) spectroscopy was used to determine the content of Klason lignin, acid-soluble lignin, total lignin, extractives, ash, acid-insoluble residue, glucose, xylose, rhamnose, galactose, arabinose, mannose and total sugars in coconut residues. The samples were analyzed at several processing stages: wet unground (WU), dried unground (DU) and dried and sieved (DS). Partial least squares models were built, and the models for the analytes exhibited $R^2 > 0.80$, with the exceptions of rhamnose, arabinose, galactose, mannose and ash from all fractions, and the lignin content from the WU fraction, which were predicted poorly ($R^2 < 0.70$). There were some significant differences between the models for the main lignocellulosic components at the various stages of biomass. These results proved that NIR spectroscopy is useful for analysis at biorefineries, and it can be used as a faster and more economical alternative to the standard methods.

1. Introduction

Brazil generates a substantial amount of lignocellulose agricultural waste [1], which includes coconut biomass. Coconut (*Cocos nucifera L.*) is grown in approximately 93 countries [2], with a world production of 60.7 million tons in an area of 11.8 million hectares [3]. Brazil is ranked fourth in the world in coconut production, producing 2.8 million tons of coconuts in an area of 287 thousand hectares [4].

However, coconut production is an important contributor to the nation's pollution problems because 80–85% of the coconut's raw weight is treated as solid waste residue in the form of husks [5], resulting in an annual production of approximately 2.3 million tons of coconut husks in the country.

Coconut husk is the mesocarp, composed of coir fibers [6]. The world production of coir fibers ranges between 5 and 6 million tons per year. However, less than 10% of coir fiber is commercialized [7], and most of the husks are abandoned in nature, wasting natural resources and causing environmental pollution [8].

Coconut husk is attractive because of its high proportions of well-defined polymeric structures of cellulose (35.00–47.00%), hemicellulose (15.00–28.00%) and lignin (16.00–45.00%); its low amounts of ash (2.70–10.00%); and, depending on the coconut variety, its high extractives content, ranging from 3.40% to 30.00% [9–13]. Efforts have been made to enhance the value of this residue as a precursor for biorefinery technologies. As a result, coconut husks have been considered a renewable resource in lignocellulosic biorefining for the production of biofuels [13,14], polymer composites [6,8,15], adsorbents [7] and chemicals [2].

The yields of these processes depend on the chemical composition of the coconut samples [16]. The biomass composition can be determined by traditional methods [17], although they are frequently time-consuming and expensive. Therefore, accurate and robust methods of analysis are of great value, particularly if integrated online in a biorefinery. In this context, NIR spectroscopy is fast, simple to apply, and non-destructive, so it is a suitable alternative to the existing reference methods. However, the applications of NIR spectroscopy are almost entirely dependent on chemometric tools. Partial least squares regression (PLS) can be directly applied to the NIR spectra, resulting in calibration models to predict the properties of interest [18].

Studies using NIR spectroscopy, coupled with chemometric tools, have shown the utility of NIR spectroscopy for the characterization of different biomasses. However, these models were developed for samples that have undergone extensive biomass preparation, including cutting, drying, comminution, sieving and removing extractives. As a result, significant amounts of time and money are spent on these analyses [16].

In this study, NIR spectra and chemometric methods were applied to minimally processed (wet unground (WU) and dried unground (DU)) coconut samples, as well as to dried and sieved

* Corresponding author.
E-mail address: marcia@iqm.unicamp.br (M.M.C. Ferreira).

http://dx.doi.org/10.1016/j.talanta.2015.03.014
0039-9140/

Table 1
Descriptive statistics for the chemical constituents (%) of 28 samples of coconut husks.

Statistic parameters/constituents[a]	Median value	Minimum value	Maximum value	Mean value	Standard deviation	Coefficient of variation
TL	25.130	17.001	35.807	24.802	4.305	17.305
KL	23.830	16.003	34.504	23.405	4.034	17.204
ASL	1.370	0.702	2.102	1.421	0.407	28.207
AIR	24.178	16.252	34.955	24.002	4.066	17.005
Extrac	20.175	1.404	41.601	20.467	11.451	55.708
Ash	1.070	0.336	3.099	1.404	0.969	68.601
Glu	25.005	17.642	32.406	24.938	3.233	12.964
Xyl	11.220	6.400	16.516	11.464	0.536	4.676
Arab	2.390	1.828	4.302	3.043	0.606	19.702
Gal	1.041	0.591	1.669	1.054	0.581	55.167
Rha	0.293	0.281	0.515	0.389	0.060	15.800
Man	0.505	0.362	2.033	0.752	0.403	53.001
TS	40.614	28.797	50.998	40.825	3.868	9.476

[a] TL: total lignin; KL: Klason lignin; ASL: acid-soluble lignin; AIR: acid-insoluble residue; Extrac: extractives; Glu: glucose; Xyl: xylose; Arab: arabinose; Gal: galactose; Rha: rhamnose; Man: manose; TS: total sugars.

(DS) biomasses, in order to determine their chemical compositions with respect to extractives (Extrac), ash, acid-soluble lignin (ASL), Klason lignin (KL), acid-insoluble residue (AIR), total lignin (TL), glucose (Glu), xylose (Xyl), galactose (Gal), mannose (Man), arabinose (Arab), rhamnose (Rha) and total sugars (TS). The quality of the final models was evaluated by determining the figures of merit.

2. Material and methods

2.1. Samples

Twenty-eight samples of coconut residues were analyzed with respect to the DS fraction, and 26 samples were analyzed with respect to the DU and WU fractions. All of the samples were collected during the period of 2010–2012 in Brazil; most of them originated from the North and Northeast regions, whereas the others came from the Southeast. Approximately 500 g of each biomass was collected during different processing stages (WU and DU). The WU and DU fractions were initially separated and stored in a freezer.

An additional 500 g of each biomass was cut into small pieces (20 mm sieve aperture), dried at 105 °C (until they reached a constant weight), ground using a Romer micro mill (Romer Labs, São Paulo, Brazil) and sieved for 20 min (180–850 μm). This biomass fraction was designated as DS (dried and sieved), and it was the fraction used for the reference analysis.

2.2. Reference analyses of biomasses

The reference analyses were carried out using standard NREL methods [19,20]. The moisture level was determined as the loss of mass after drying at 105 °C in an oven overnight, and the ash content was determined as the residue after the combustion of a sample with known dry-matter content. The muffle furnace Naber therm L-240H1SN was used at a temperature of 575 °C for 4 h.

Each fraction of the sample was then extracted with 95% ethanol using accelerated solvent extraction in a Dionex ASE 200 system (Thermo Fisher Scientific, Waltham, MA, USA), and these extractive-free materials were used for subsequent analyses. The extracted samples were then subjected to a two-stage acid hydrolysis, with 72% sulfuric acid (3 mL) in a water bath in the first step, followed by hydrolysis in an autoclave for 1 h at 120 °C with an acid concentration of 4%.

The ASL extract consisted of low-molecular-weight lignin solubilized in the acidic hydrolysis solution. The ASL concentration was measured in the diluted hydrolyzate (with a low-concentration acid solution) by UV spectroscopy in a Shimadzu UV-1700 spectrometer (Shimadzu, Kyoto, Japan) measuring the absorbance at 240 nm. The AIR, i.e., the dried solid residues (at 105 °C overnight) after the acid hydrolysis, was ashed to determine the acid-insoluble ash (AIA). The difference between the AIR and AIA levels gave the KL content. Finally, the TL content was calculated as the sum of the soluble and insoluble lignin during the acid hydrolysis, ASL and KL, respectively.

Structural carbohydrates were hydrolyzed into monomeric sugars, releasing monosaccharides into the acid hydrolysis solution (arabinose (Arab), galactose (Gal), rhamnose (Rham), glucose (Glu), xylose (Xyl) and mannose (Man)). They were quantified by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) using an ED 40 electrochemical detector, along with a CarboPac-PA 10 column and precolumn. Standard monomeric sugar solutions were hydrolyzed concurrently during the secondary hydrolysis step, and this degradation was used to account for the loss of carbohydrates during the acid hydrolysis.

All of the analyses were performed in duplicate, with a standard deviation between duplicates of less than 1% for all parameters. All of the results were presented as percentages (%).

2.3. Visible–near infrared spectroscopy

A FOSS XDS instrument (FOSS, Hillerød, Denmark), equipped with the associated rapid content analyzer (RCA) module and a diffuse reflectance detector, was utilized to record the near-infrared spectra. The spectra (1100–2500 nm) were obtained in 0.5 nm increments and were generated by averaging 32 successive scans. WU and DU samples were scanned in a large rectangular cell because of their heterogeneous particles and larger particle sizes, and the DS samples were scanned in a small circular cell because of their small particle sizes. Each sample was analyzed in triplicate, and the average spectrum was used for further data treatment. One small circular cell with a ceramic standard as reference material was used over the scanning window to register the blank spectrum.

2.4. Multivariate data analysis

Statistical and multivariate data analyses were conducted using the UNSCRAMBLER 10.3 software package (Camo Software, Oslo, Norway), and one PLS routine from the PLS-toolbox 6.7 was used to calculate the figures of merit (Eigenvector Research, Wenatchee, WA, USA) for Matlab 7.2 software (MathWorks, South Natick, MA, USA).

PLS-1 (one dependent variable) was used for constituent quantification [21]. The original data set was randomly divided into two

sub-sets, one used for calibration (consisting of 20 samples) and the other used for external validation (consisting of the remaining samples). The number of latent variables (LV) in the calibration models was determined based on the minimum root-means-square error of cross-validation (RMSECV) [22].

Several pre-processing methods were tested, and the best results were obtained by combining the standard normal variate (SNV) with first (1D)/second (2D) derivative transformations [23,24] and by combining the SNV with Detrend (DT) [23]. In addition, the second derivative pretreatment method was individually used [24]. Variables were selected through an automatic uncertainty test (Martens' uncertainty test), to select the most significant variables in the models [25].

The performance of each model was evaluated based on the external validation data set and the calibration data set by calculating the coefficients of determination (R^2_{ext} and R^2_{cal}, respectively); the standard error of calibration (SEC); the standard error of prediction (SEP); bias; the relative standard deviation (RSD%= SEP*100/mean), where the mean was taken from the reference values in the validation set; the number of LVs used; the fraction of outliers excluded, identified by analyzing the plot of leverage versus Student residuals [22]; and the range error ratio (RER= $Range_y$ of validation data/SEP) [26,27]. The last parameter (RER) provides good results when the RER value is ≥ 4 and very good results for research quantification when RER ≥ 15 [26].

To provide reliable multivariate calibration models, the figures of merit were also calculated [28]. For that, the net analyte signal (NAS) was used, according Bro and Andersen [29]. The analytical sensitivity (γ), its inverse (γ^{-1}), the limit of detection (LOD) and the limit of quantification (LOQ) were calculated using the instrumental noise (δx) estimated from the standard deviation in each wavelength from 15 blank spectra pretreated according to each specific PLS model. Then, the sensitivity (SEN) and the selectivity (SEL) were calculated based on the NAS [18]. The accuracy was expressed as the SEC and the SEP. The linearity was graphically depicted by the plots of residuals and by the plot of the reference versus the predicted values. Additionally, the standard deviations (SEC, SEP, SECV) were statistically compared by the *F*-test.

Finally, the regression coefficient vector was interpreted for each constituent [30].

3. Results and discussion

The statistical results regarding the chemical constituents are summarized in Table 1. The highest range was observed for ash and extractives (Extrac), with a wide variation in the coefficient of variation (68.60 and 55.70, respectively), followed by the sugars mannose (Man) and galactose (Gal). The most important sugar, glucose (Glu), and the total sugars (TS) presented a small range of variation, with low coefficients of variation (12.96 and 9.47, respectively). All of the lignins showed intermediate variations between 17.00 and 28.20. This wide compositional variability for some parameters was obtained because highly heterogeneous populations, such as samples from different regions, soils, years, cultivars and species were sampled.

The NIR raw spectra of the 28 coconut samples from the DS fraction (Fig. 1A), and the 25 samples from the DU (Fig. 1B) and WU (Fig. 1C) fractions, are shown in Fig. 1.

The main absorption bands are located at 1450/1470 nm, 2090 nm, and 1920 nm, and weak bands appear as shoulders at 1170/1270 nm and 2274/2336 nm. These bands were mainly related to the O–H stretch from water (1920 nm), the O–H combination from polysaccharides (2090 nm), and the O–H stretch and the 1st overtone of the OH groups with H-bonds of intermediate strength (1450–1470 nm) [31,32]. The less intense bands (1170/1270 nm) are associated with the C–H stretch 2nd overtone from lignin, and the bands at 2274/2336 nm are assigned to the O–H stretch/C–H stretch and/or the C–O stretch combination/C–H deform combination of the polysaccharides [31]. The clear

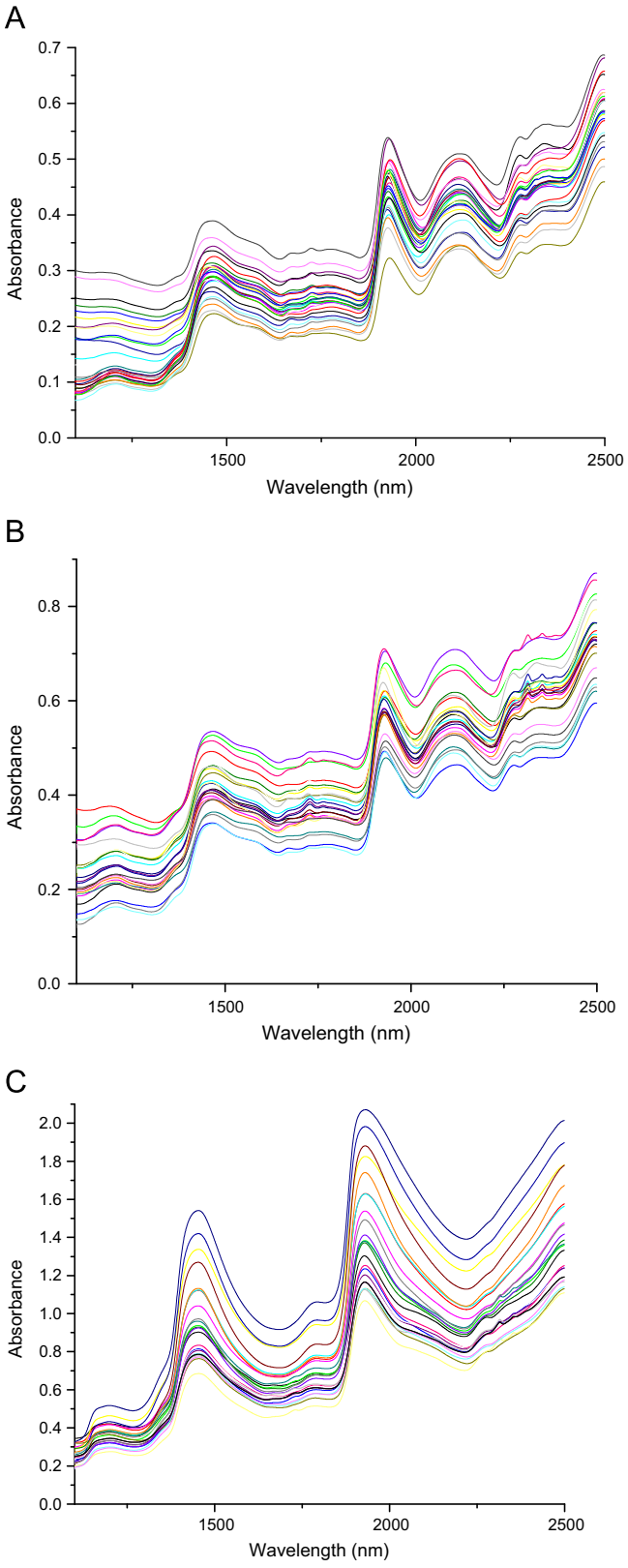

Fig. 1. NIR raw spectra for all coconut samples of (A) the DS fraction, (B) the DU fraction and (C) the WU fraction.

Table 2
Parameters and statistics for validation of the best PLS models obtained for lignin and extractive contents.

Y	Sample set	Pre-treatment	Samples of calibration sets	Samples of prediction sets	LV	Out.	Inl.	R^2		Pred	RM SE		Pred	RSD	RER
								Cal	CV		Cal	CV			
TL	DS*	2D(15)+SNV	20	7	3	–	1	0.86	0.78	0.83	1.593	2.019	1.740	7.00	6.0
	DU*	SNV+1D(3)	19	4	3	1	1	0.86	0.79	0.79	1.499	1.946	1.600	4.2	4.0
	WU	–	–	–	–	–	–	–	–	–	–	–	–	–	–
KL	DS*	2D(15)+SNV	19	7	7	1	1	0.98	0.83	0.86	0.491	1.724[a]	1.291[b]	5.50	6.0
	DU*	SNV+1D(3)	20	4	4	–	–	0.91	0.77	0.87	1.019	1.784	1.089	5.00	5.7
	WU	–	–	–	–	–	–	–	–	–	–	–	–	–	–
ASL	DS*	2D(15)	20	7	6	–	1	0.95	0.73	0.83	0.085	0.212	0.106	6.40	10.0
	DU*	SNV+1D(3)	20	5	3	–	–	0.89	0.82	0.88	0.131	0.175	0.166	11.6	4.0
	WU	–	–	–	–	–	–	–	–	–	–	–	–	–	–
AIR	DS*	2D(15)+SNV	20	7	5	–	1	0.95	0.87	0.77	0.882[a,b]	1.510[a]	1.844[b]	8.50	5.5
	DU*	SNV+1D(3)	19	5	5	1	–	0.86	0.80	0.88	1.377	1.764	1.193	5.2	4.70
	WU	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Extrac	DS*	2D(15)+SNV	18	7	6	2	1	0.99	0.92	0.90	0.607[a,b]	2.979[a]	3.019[b]	14.0	11.0
	DU	SNV+DT	19	4	4	1	1	0.89	0.84	0.88	3.540	4.647	3.717	13.5	6.0
	WU*	2D(15)	18	5	7	2	–	0.92	0.75	0.90	3.115[a]	6.081[a]	3.414	15.0	8.8

* Variable selection; LV: latent variables; Out: outliers removed from the calibration set; Inl: outliers removed from the validation set.
[a] Significantly different calibration and cross validation set errors.
[b] Significantly different calibration and external validation set errors.

Table 3
Parameters and statistics for validation of the best PLS models obtained for sugar composition.

Y	Sample set	Pre-treatment	Samples of calibration sets	Samples of prediction sets	LV	Out.	Inl.	R^2		Pred	RM SE		Pred	RSD	RER
								Cal	CV		Cal	CV			
Glu	DS*	SNV+1D(3)	19	6	4	1	2	0.90	0.83	0.83	1.075[b]	1.508	1.770[b]	6.7	7.0
	DU*	SNV+DT	19	4	3	1	1	0.88	0.84	0.87	1.332	1.646	1.486	6.0	5.0
	WU*	2D (25)	18	5	4	2	–	0.92	0.83	0.78	1.059[a,b]	1.619[a]	2.332[b]	10.0	4.0
Xyl	DS*	SNV+1D(3)	19	7	3	1	1	0.89	0.78	0.80	0.954[a]	1.444[a]	1.252	10.0	10.0
	DU	SNV+DT	20	5	3	–	–	0.82	0.74	0.78	1.280	1.621	1.333	11.6	5.4
	WU*	–	–	–	–	–	–	–	–	–	–	–	–	–	–
TS	DS*	SNV+1D(3)	19	7	5	1	1	0.91	0.80	0.79	1.922[a,b]	2.934[a]	3.541[b]	8.0	5.5
	DU*	SNV+DT	19	5	4	1	–	0.92	0.86	0.91	1.954	2.738	2.077	4.6	12.0
	WU*	2D(25)	19	5	4	1	–	0.85	0.71	0.78	2.544[a,b]	3.918[a]	3.829[b]	9.0	5.4

* Variable selection; LV: latent variables; Out: outliers removed from the calibration set; Inl: outliers removed from the validation set.
[a] Significantly different calibration and cross validation set errors.
[b] Significantly different calibration and external validation set errors.

differences among the spectra from different fractions (Fig. 1A–C) occur in the regions of approximately 1450/1470 nm (associated with the first overtone of the symmetric vibration of OH) and 1920 nm (associated with the OH stretch), mainly related to water [32,33], where the WU fraction presented more intense bands.

3.1. Partial least-square regression modeling

Unique PLS models were built for different fractions of the samples and for different analytes, which are summarized in Table 2, and for individual monosaccharides, as well as the total sugar content (Table 3). The NIR spectra were analyzed in the region between 1100 and 2500 nm with variable selection.

The NIR spectra presented non-linear baseline deviations due to multiplicative light scattering caused by the non-homogeneous particle size distribution. The deviations were more intense for the unground samples from the WU fraction (Fig. 1C). These instrumental deviations are not related to the chemical composition and can be removed by preprocessing methods. In this work, the most common preprocessing techniques used for the spectra were the second or first derivative [24] and SNV combined with Detrend [34] or derivatives [23].

In most of the cases, larger amounts of variability unrelated to the analyte were removed when applying two pretreatments for correcting the spectral data, such as SNV followed by derivatives, or otherwise. The SNV and derivatives remove different types of effects, and there might well be some improvement in the models when

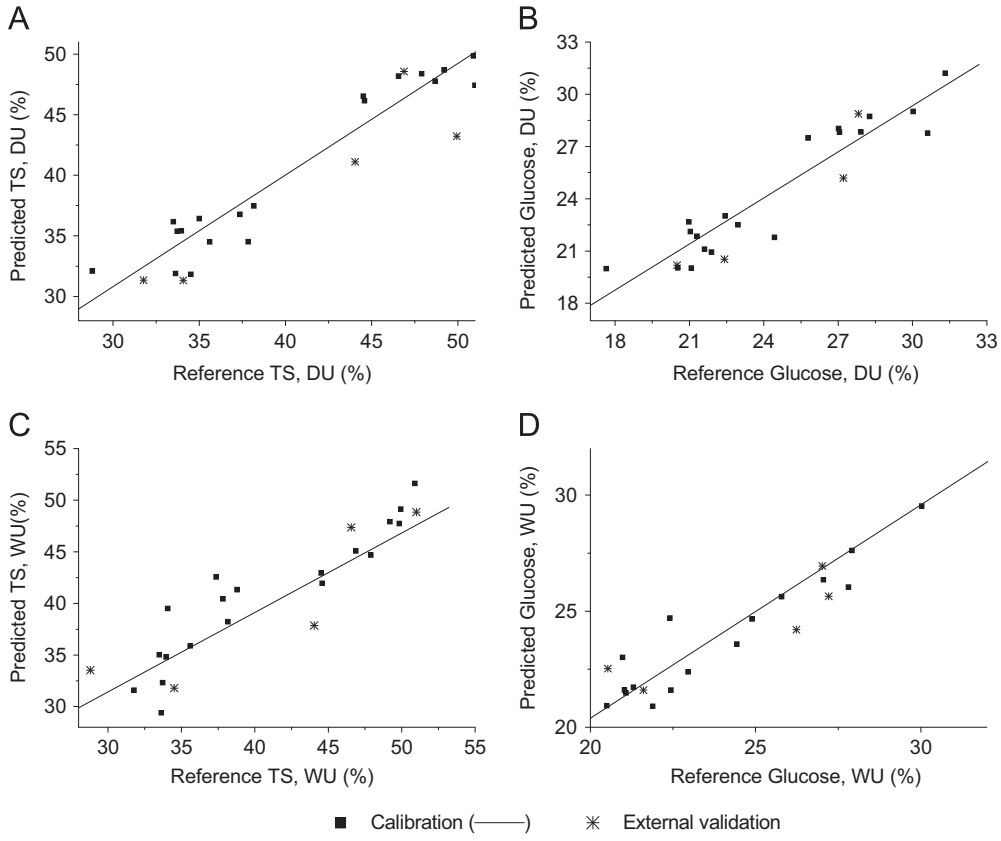

Fig. 2. Plot of reference *versus* predicted values from the calibration and external validation models from the DU and WU fractions for (A) the total sugar content with 4 VL; (B) the glucose content with 3 VL; (C) the total sugar content with 4 VL; (D) the glucose content with 4 VL.

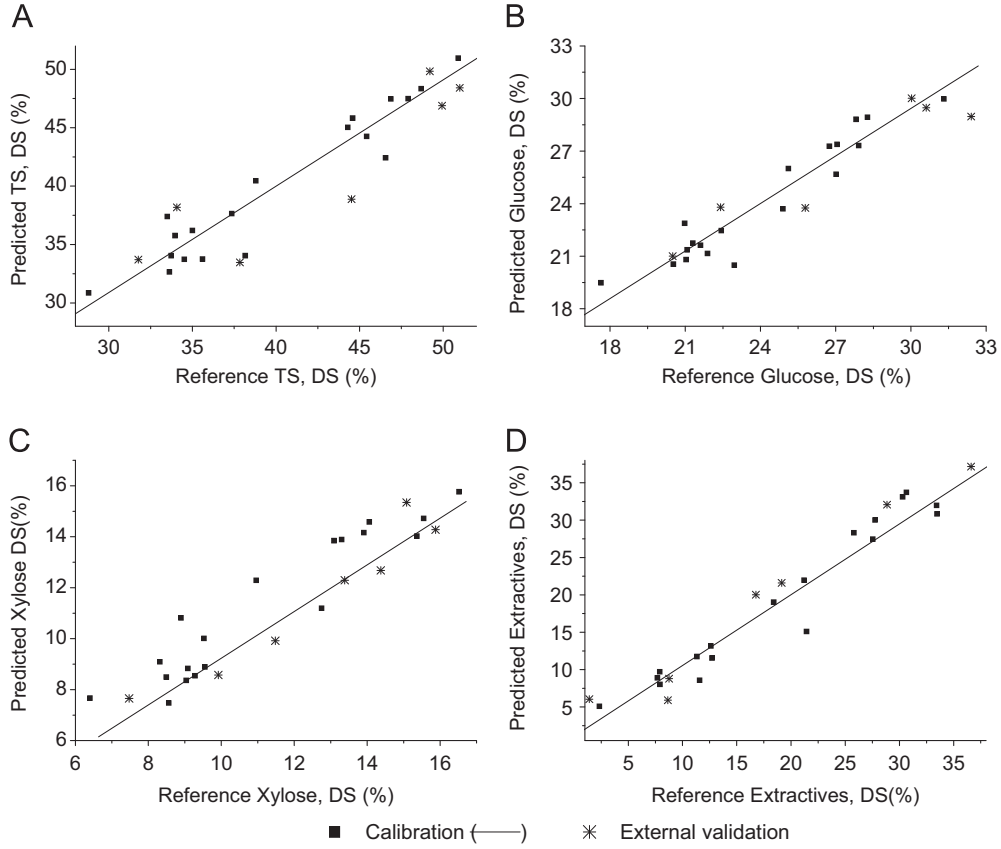

Fig. 3. Plot of reference *versus* predicted values from the calibration and external validation models from the DS fractions for (A) the total sugar content with 5 VL; (B) the glucose content with 4 VL; (C) the xylose content with 3 VL; (D) the extractives content with 6 VL.

applying both pretreatments. What matters most in such cases is the order in which the pretreatments are applied [23]. Except for ASL, the best models for all the analytes from the DS fraction were obtained when using one of the combinations mentioned above. Improvements were shown for lignins and extractives when the derivatives method was applied first, whereas sugar models provided better results when SNV pretreatment was applied first. When SNV is applied after the derivative, more variability is removed, i.e., more irrelevant information in the spectra is removed. However, applying SNV before taking the derivatives (1D or 2D) removes multiplicative effects. Thus, this combination removes both multiplicative effects and sloping baselines [23].

Detrend is a baseline correction technique that utilizes polynomials and can be used to remove nonlinear trends and reduce multicollinearity, baseline shift and curvature [34], such as those found in the spectra from DU fractions. The procedure is generally used in conjunction with pretreatment SNV [35] because the data corrected by SNV may still be affected by baseline curvature. However, the DU fraction also shows favorable models for lignins using combined pretreatments such as SNV+first derivative.

Regarding all the models for the WU fraction, the relative errors when using SNV were significantly higher than those observed when the second derivative was applied. Apparently some important spectral information in WU spectra is removed when scatter correction techniques are applied, and the derivatives appear to be a better alternative.

For lignins (TL, KL, ASL and AIR) and extractives (Table 2), all of the PLS models, independent from the fraction, can be considered parsimonious, with an LV < 7 for most models. Low relative errors were obtained for lignins, in the range of 4.2–11.6, whereas higher values of RSD for extractives (> 13%) were obtained. The RER was ≥ 10.00 for ASL (in the DS fraction) and extractives (in the DS fraction), indicating that very good models were obtained at the level of quality control. All of the other models, with RER ≥ 4.0, were qualified for screening calibration [25]. The lignin models for the WU, DU and DS fractions presented significant differences; in general, the heterogeneity of the samples influenced the performances of the models, with better results obtained for the models from the DS fractions.

The presented models for the lignins (Table 2) were similar to and, in most cases, better than the previously reported models in the literature for other biomasses [16,18,36–39]. The studies that reported higher values of $R^2_{cal,pred}$ [16,38] were modeled with an excessive number of factors (≥ 10). In addition, compared with the studies cited above that used the same number of factors as in the present paper (3–7 VLs), the RER and R^2 values of the cited studies were always lower than those reported in this work.

The extractives (Table 2) were determined by accelerated solvent extraction, and this method was chosen to increase the extraction speed and efficiency as it requires lower volumes of solvent than Soxhlet methods. Despite being a complex mixture with a high degree of superposition of absorption bands from the many compounds found in the extractives [18,40], very good models were obtained for the extractives, with RER ≥ 6, $R^2_{cal,val}$ ≥ 0.89 and 0.88. Suchat et al. [40] developed calibration models for acetone extracts and obtained a similar $R^2_{cal,val}$ of 0.96 and 0.96, respectively, for homogeneous samples, but without reporting the number of LVs in the model. Hayes [16] obtained $R^2_{cal} > 0.91$ for 95% ethanol-soluble extractives content but with high numbers of factors 13, 15 and 17 LVs for the DU, DS and WU fractions, respectively, compared with 4, 6, and 7 LVs (for the DU, DS and WU fractions, respectively) in this study.

The ash calibration models are not acceptable for screening calibration or for routine quality control because all of the RER values obtained were lower than 4.00, and the RSD values were higher than 30.00%. Some authors [39] obtained good results for ash, with few LVs.

With respect to the sugar content, the RSD for the major constituents, glucose, xylose and total sugars (TS), were less than 10.0%, 11.6% and 9.0%, respectively, showing that the models have good predictability regardless of the fraction used. The performances of the models presented in this work (Table 3) are similar to those previously reported in the literature for other biomasses [16,41,42].

Arabinose and galactose, monosaccharides with intermediate concentrations in coconut husks, were not well-modeled independent of the fraction, WU, DU or DS. The coefficients of determination $R^2_{cal,\ val}$ were below 0.70, and the models were not reliable (data not shown in Table 3). Compared with the major sugars, the predicted results for the minority sugars, mannose and rhamnose from the fractions DS, DU and WU were also relatively poor, and the results are not reported in Table 3. This poor model performance is most likely due to the smaller percentages of these sugars by weight, resulting in less intense signals that could be masked by others. Another factor that can justify the worse fitting is the fact that small proportions result in low concentrations in the hydrolyzate, so the measurement by chromatography might not be as accurate as for glucose and xylose, for example [43].

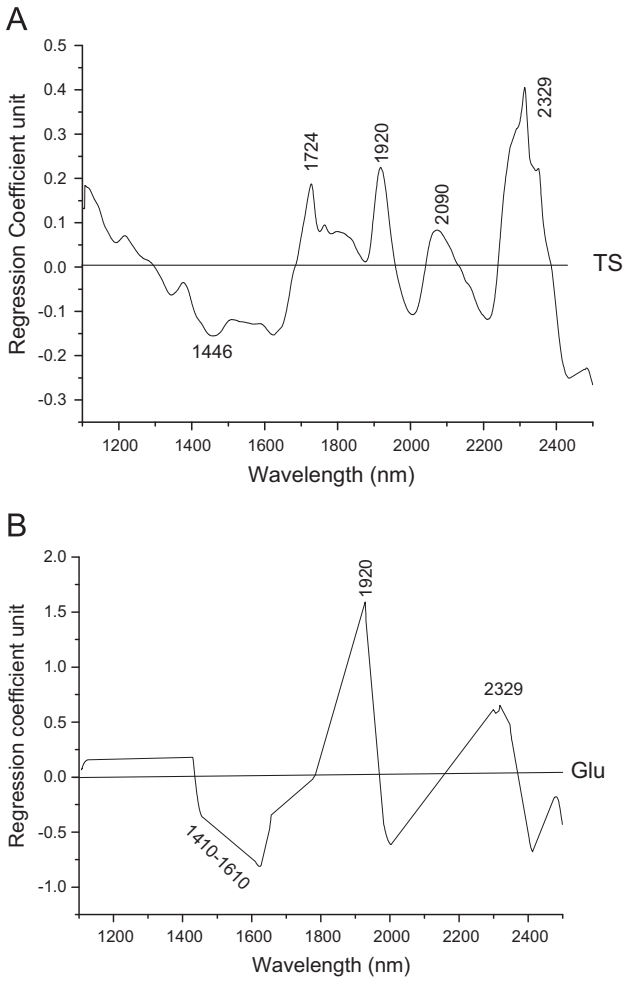

Fig. 4. Plot of reference *versus* predicted values from the calibration and external validation models from DU and DS fractions for (A) the total lignin content with 3 VL; (B) the acid lignin soluble content with 3 VL; (C) the total lignin content with 3 VL; (D) the acid lignin soluble content with 3 VL.

Fig. 2 presents the plots of the reference *vs*. the predicted values from the sugar calibration models (total sugar and glucose) and the external validation for both unground samples, wet (WU) and dried (DU).

Fig. 3 shows the plots (calibrations and external validation) of the reference *vs*. the predicted values from the sugar calibration (total sugar, glucose and xylose) and extractives models for the sieved and dried samples (DS). Comparing the three fractions, DS, DU and WU, in Figs. 2 and 3, the DS and the DU fractions presented models with lower RSD and higher RER values in most of the cases, and a lower number of LV and outliers were detected for the sugar models.

The regression coefficients of the reference and the predicted values for the lignins (total lignin and acid-soluble lignin) models from the dried fractions (DU and DS) are shown in Fig. 4.

The external validation sets were used to check the predictability of the models. They were of different sizes, and the number of samples for each analyte is listed in the third column of Table 4. The extractive models had the most deviations, whereas the sugar models (Glu and TS) presented the least deviations (Table 4). Although the number of samples is not large, all of the models obtained can be considered reliable for screening calibration and, in some cases, for quality control.

3.2. Regression coefficients interpretation

Examining the regression coefficients for the lignin models (TL, KL and AIR), from the unground (DU) as well as the ground samples (DS) sets, it can be observed that the regression coefficients are quite similar and present signals that can be assigned to the lignin structures. Fig. 5 shows the regression vectors for KL and ASL, both from the DS fraction, which displays the third overtone of the Aryl C–H stretch at 1114–1122 nm, C–H stretching in the aromatic structure at 1672 nm [44], and the lignin combination bands in the region of 2150–2460 nm. The significant bands at 1900 nm and 1940 nm are assigned to water. The significant bands at 1920 nm and 2090 nm confirm the contribution of polysaccharides (which have a negative correlation with the lignin content) to the insoluble lignins [32,33]. This contribution can indicate the presence of sugars that have not been fully hydrolyzed in the insoluble residue of the lignins. The ASL also shows a polysaccharide contribution at 1430 nm with a negative correlation (assigned to the O–H stretch, which is the 1st overtone of amorphous polysaccharides). Additionally, one band at 1672 nm assigned to lignin (C–H stretch of 1st overtone) with negative correlation was found. Because these regression coefficients were obtained from the second derivative spectra, the negative coefficients correspond to a direct relationship.

Other differences observed in the correlation coefficients from ASL occur at 1114–1122 nm with poor bands compared with the other lignins and in the second overtone of the carboxyl and ester stretching (1943 nm). None of the lignin fractions, with the exception of the ASL (i.e., the TL, KL and AIR fractions), presented this band at 1943 nm, indicating that lignin, in general, was not correlated with carboxyl groups. However, the positive correlation (i.e., the negative regression coefficient) of the signal at 1943 nm with the ASL fraction can be attributed to the inability of the ASL determination method to differentiate ASL from other acid-soluble compounds that also absorb in the UV region, such as uronic acids (which present carboxyl groups), furfural, and possibly the acid degradation products of the extractives that may not have been fully removed in the extraction step [45]. These possibilities can indicate a possible overestimation of the ASL content by the reference method. Nevertheless, good calibration models were obtained for ASL. A typical lignin band was found at 1724 nm (the 1st overtone of the C–H stretch,) for the ASL coefficient.

The regression coefficients from the PLS models for the total sugars and the glucose from the DU fraction (Fig. 6) show typical carbohydrate bands. The absorption at 1724 nm was attributed to the 1st overtone of the C–H stretching in hemicellulosic sugars [45]. The band at 1920 nm was attributed to the O–H stretch from polysaccharides, and it overlapped with H_2O [32]. The combination bands at 2090 and 2329 nm were attributed to polysaccharides. A

Table 4
Predicted and reference values and corresponding relative residues for some PLS models.

y	Fraction	Sample number	Pred. value	Ref. value	Rel. res.	Fraction	Sample number	Pred. value	Ref. value	Rel. res.
TS	WU	1	34.516	31.7957	−8.556	DU	1	44.042	41.970	−4.903
		2	46.560	47.3694	1.707		2	34.074	32.558	−4.656
		3	44.042	37.8538	16.348		3	46.875	45.728	−2.508
		4	50.994	48.851	−4.385		4	31.771	29.964	−6.028
		5	28.797	33.5288	14.113		5	49.935	47.321	−5.524
ASL	DU	1	1.06	1.09	3.13	DS	1	1.431	1.58303	9.667
		2	1.65	1.73	4.42		2	1.082	0.89629	−20.496
		3	1.31	1.48	11.29		3	1.151	0.89093	−29.078
		4	0.96	1.11	13.50		4	1.820	1.65498	−9.971
		5	1.43	1.70	15.72		5	1.068	0.91996	−16.091
			–	–	–		6	1.655	1.38404	−19.215
							7	1.9660	1.61256	−21.921
Extrac	WU	1	30.300	24.6116	23.112	DU	1	18.41	22.1429	16.858
		2	27.760	32.5689	14.765		2	33.54	37.4737	10.497
		3	11.330	11.3087	−0.188		3	7.69	12.7107	39.500
		4	18.410	18.3169	−0.508		4	36.61	37.4200	2.165
		5	8.760	11.0151	20.473		–	–	–	–
		–	–	–	–		–	–	–	–
Gluc	DS	1	20.500	21.0011	2.385	DU	1	27.2102	25.1961	−7.994
		2	32.406	28.9720	11.854		2	22.4103	20.5293	−9.162
		3	22.410	23.8003	5.840		3	27.813	28.8821	3.702
		4	30.026	30.0156	−0.037		4	20.5003	20.1852	−1.561
		5	30.611	29.4722	−3.864		–			
		6	25.787	23.7575	−8.545		–			

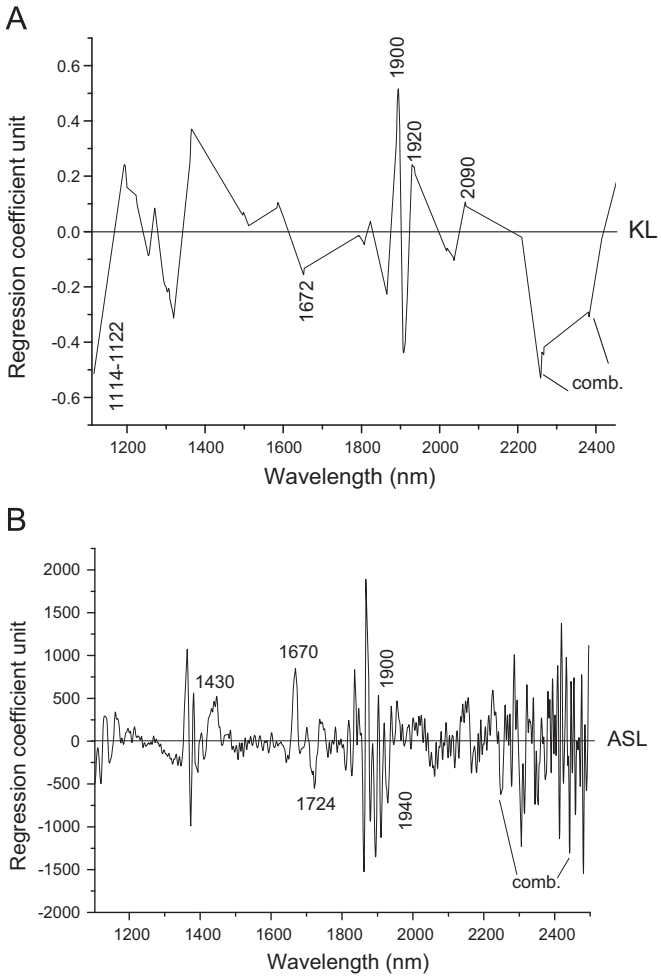

Fig. 5. Regression coefficients of the lignin (KL and ASL) PLS models for the DS fraction. The comb. refers to combination bands.

negative band at 1446 nm was assigned to the phenolic hydroxyl groups, most likely of lignins [46], which demonstrated an inverse relationship of the TS with the lignin content. However, for the glucose coefficient regression (Fig. 6B), one broad band at 1410–1610 nm was attributed to the O–H stretch, specifically the 1st overtone of alcoholic/sugar structures. Other sugars most likely presented a negative correlation with glucose. Here, the presented coefficients were obtained from a SNV+DT pretreatment, thus positive coefficients corresponded to a direct relationship.

The signals that were correlated to the extractive contents (data not shown) were assigned to waxes RCO_2R (1938 nm), proteins, and/or polyalcohols $CONH_2R$ (2056 nm), which are typical components present in extractives [47]. The second overtone of the carboxylic acid (RCO_2H) stretching vibration (1910 nm) is negatively correlated with the extractives, which most likely results from the long-chain fatty acids insoluble in ethanol, whose content is negatively correlated with the ethanol-soluble fraction.

The first overtone of the alkyl region's stretching (1625–1775 nm) and combination bands (2200–2450 nm) are complex, presenting positive and negative regression coefficients. This complex structure of the PLS model, indicated by the regression coefficient features and the high number of LVs, is due to the multicomponent mixture character of the extractives fraction.

3.3. Analytical validation by figures of merit

Table 5 summarizes the parameters estimated for evaluating the main figures of merit of the developed models.

The linearity of the method was evaluated by examining the residuals of the PLS models (data not shown), checking for the absence of systematic trends in the distribution of residuals and suggesting their random behavior. Additionally, the points should be linearly distributed around a diagonal line in the plots (Figs. 2–4).

The SEL provides an estimate of the amount of instrumental signal that was used by the calibration model for determining the analyte. The models showed SEL varying from 0.53 to 0.006 (55.00–0.60%) (Table 5), i.e., less than 0.006 of the original signal was orthogonal to the space of the interferents, carrying less than 1.00% of the analyte information modeled. However, this behavior occured for a single model (galactose from the WU fraction). The other models showed higher SEL, above 4.00%.

The SEN value is not appropriate for comparing models because it depends on the pretreatment applied and the analyzed matrix. Lower values of sensitivity, on the order of 10^{-5}, were obtained for models that employed derivatives, mainly 2D, as observed for the models of the sugars (glucose and total sugars for WU fractions) and the extractives (WU). Moreover, models using SNV in combination with other pretreatments such as derivatives or Detrend, as in the models for lignins, among others, showed high sensitivity values. This is because the derivative spectrum has very small intensities, requiring large regression coefficients for conversion to the analyte concentration [18]. Consequently, the increase in the regression coefficients results in decreased sensitivity [48].

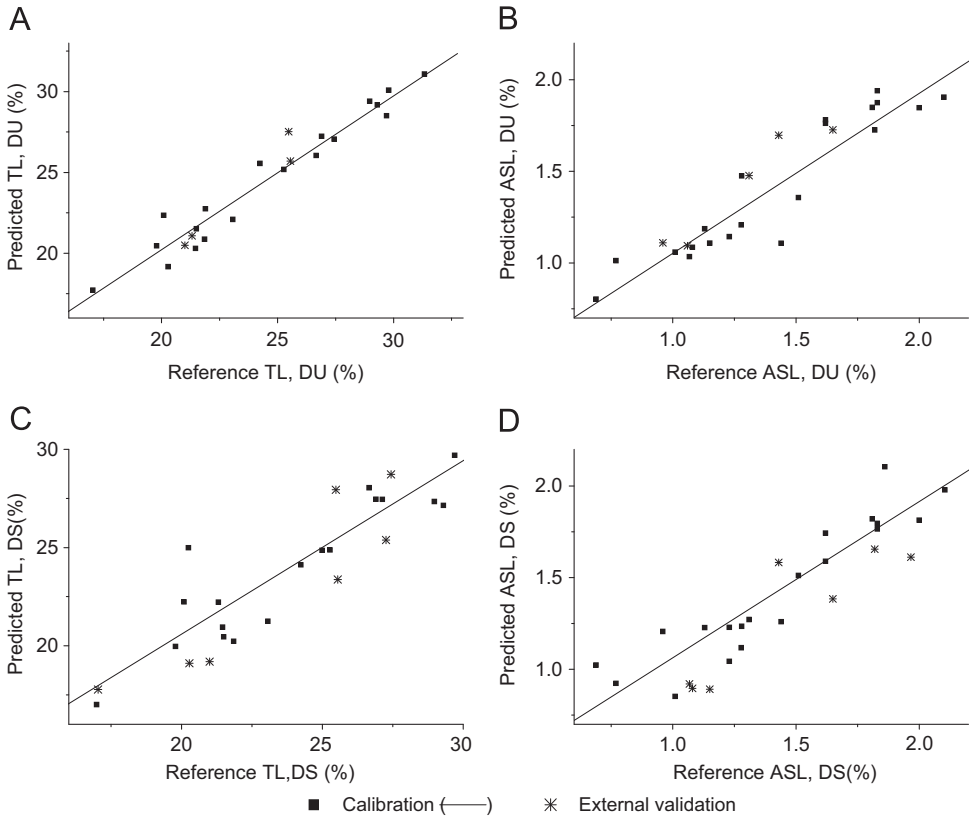

Fig. 6. Regression coefficients of the total sugar and glucose PLS models for the DU fractions.

Table 5
Results of the obtained figures of merit for the PLS models.

y	Sample set	*SEL*	*SEN*	γ	γ^{-1}	*LOD*	*LOQ*
TL	DS	0.483	51.387	413.24	0.002	0.007	0.024
	DU	0.476	0.113	41.524	0.0241	0.072	0.240
KL	DS	0.006	2.087	87.233	0.011	0.034	0.114
	DU	0.208	0.104	455.517	0.0022	0.006	0.022
ASL	DS	0.129	355.99	7.225×10^{4}	1.383×10^{-5}	4.151×10^{-5}	1.383×10^{-4}
	DU	0.536	1.627	4.099×10^{3}	2.439×10^{-4}	7.318×10^{-4}	0.002
AIR	DS	0.324	31.996	71.620	0.014	0.041	0.139
	DU	0.033	0.044	5.692	0.175	0.527	1.756
Extractives	DS	0.036	6.275	220.127	0.004	0.013	0.045
	DU	0.093	21.660	653.479	0.0015	0.005	0.015
	WU	0.011	8.459×10^{-5}	29.638	0.033	0.101	0.337
Glu	DS	0.075	0.058	880.383	0.001	0.003	0.011
	DU	0.061	30.806	251.522	0.004	0.011	0.039
	WU	0.277	5.648×10^{-5}	23.558	0.042	0.127	0.424
Xyl	DS	0.092	0.019	72.306	0.013	0.041	0.138
	DU	0.270	12.843	165.737	0.006	0.018	0.060
TS	DS	0.116	0.007	29.771	0.033	0.100	0.336
	DU	0.202	3.839	74.899	0.013	0.040	0.133
	WU	0.105	1.542×10^{-5}	45.430	0.022	0.066	0.220

The inverse of the analytical sensitivity (γ^{-1}), with values all smaller than 0.17%, provides an estimation of the minimum concentration difference that is discernible by the analytical method.

The minimal observed values were all larger than the LOQ, with a maximum value of 1.76. The results for the LOD indicated that the NIR–PLS method was able to detect concentrations in the coconut husks samples for all of the models once the minimal observed values for all of the parameters were larger than the LOQ.

4. Conclusions

This study demonstrated that it was possible to predict the major lignocellulosic constituents (TS, glucose and TL) of coconut husks samples based on their dried spectra. The wet spectra provided reasonable results for sugar models but not for lignins. The intermediate and minor constituents such as sugars (arabinose, rhamnose, galactose and mannose) and ash cannot be used for screening

or quantitative predictions. In general, small differences among the models from different biomass stages were observed, with the DS fraction showing slightly better results for the extractives and the DU fraction showing better results for lignins, whereas for sugar models, no significant differences were observed between the fractions. However, considering that sample preparation, as in DS fraction, requires extensive time and labor, the multivariate analysis techniques applied to the NIR wet and dried spectra (WU and DU) with minimally processed samples have demonstrated the potential for on-line use because good RER values were obtained, appropriate for screening calibrations for all models and, in some cases, suitable for research quantification.

Acknowledgments

The authors thank the Sugarcane Research Center – CTC, for kindly providing their facilities for this study. One of the authors (MMCF) acknowledges CNPq, Brazil. The DIBANET research leading to these results has received funding from the European Union Seventh Framework Programme [FP7/2007-2013] under Grant agreement no. [227248].

References

[1] J.C. Bergmann, D.D. Tupinambá, O.Y.A. Costa, J.R.M. Almeida, C.C. Barreto, B. F. Quirino, Renew. Sustain. Energy Rev. 21 (2013) 411–420.
[2] A. Bouaid, M. Martínez, J. Aracil, Bioresour. Technol. 101 (2010) 4006–4012.
[3] FAO, Food and Agriculture Organization ⟨http://www.fao.org⟩.
[4] Embrapa, Brazilian Agricultural Research Corporation ⟨https://www.embrapa.br/⟩.
[5] M.F. Rosa, F.J.S. Santos, A.A.T. Montenegro, F.A.P. Abreu, D. Correia, F.B.S. Araújo, E.R.V. Norões, Comunidade Téc. 54 (2001) 1–6.
[6] J. Sarkia, S.B. Hassana, V.S. Aigbodiona, J.E. Oghenevweta, J. Alloy. Compd. 509 (2011) 2381–2385.
[7] A. Bhatnagar, V.J.P. Vilar, C.M.S. Botelho, R.A.R. Boaventura, Adv. Colloid Interface Sci. 160 (2010) 1–15.
[8] W. Wei, H. Gu, Mater. Des. 30 (2009) 2741–2744.
[9] S.J.T. Pollard, G.D. Fowler, C.J. Sollars, R. Perry, Sci.Total Environ. 116 (1992) 31–52.
[10] J.E.G. Van Dam, M.J.A. Van Dem Oever, E.R.P. Keijers, J.C. Van Der Putten, C. Anayron, F. Josol, A. Peralta, Ind. Crop. Prod. 24 (2006) 96–104.
[11] K. Bilba, M.A. Arsene, A. Ouensanga, Bioresour. Technol. 98 (2007) 58–68.
[12] A.I.S. Brígida, V.M.A. Calado, L.R.B. Gonçalves, M.A.Z. Coelho, Carbohydr. Polym. 79 (2010) 832–838.
[13] P. Vaithanomsat, W. Apiwatanapiwat, N. Chumchuent, W. Kongtud, S. Sundhrarajun, Kasetsart J. (Nat. Sci.) 45 (2011) 159–164.
[14] F.A. Gonçalves, H.A. Ruiz, C.C. Nogueira, E.S. Santos, J.A. Teixeira, G.R. Macedo, Fuel 131 (2014) 66–76.
[15] F. Tomczak, T.H.D. Sydenstricker, K.G. Satyanarayana, Compos. Part A 38 (2007) 1710–1721.
[16] D.J.M. Hayes, Bioresour. Technol. 119 (2012) 393–405.
[17] D.B. Bevilaqua, M.K.D. Rambo, T.M. Rizzetti, A.L. Cardoso, A.F. Martins, J. Clean. Prod. 47 (2013) 96–101.
[18] M.K.D. Rambo, E.P. Amorim, M.M.C. Ferreira, Anal. Chim. Acta 775 (2013) 41–49.
[19] NREL, Determination of Extractives in Biomass 2008 Biomass Analysis Technology Team Laboratory Analytical Procedure, National Renewable Energy Laboratory.
[20] NREL, Determination of Structural .rbohydrates and Lignin in Biomass 2008 Biomass Analysis Technology Team Laboratory Analytical Procedure, National Renewable Energy Laboratory.
[21] H. Martens, T. Naes, Multivariate Calibration, first ed., John Wiley & Sons, New York, 1996.
[22] M.M.C. Ferreira, A.M. Antunes, M.S. Melgo, P.L.O. Volpe, Quím. Nova 22 (1999) 724–731.
[23] N. Fearn, NIR News 14 (2003) 9–11.
[24] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627–1638.
[25] F. Westad, H. Martens, J. Near Infrared Spectrosc. 8 (2000) 117–124.
[26] L. Liu, X. Ye, A. Saxton, A.J. Womac, J.Near Infrared Spectrosc. 18 (2010) 317–331.
[27] A. Versari, G.P. Parpinello, A.U. Mattioli, S. Galassi, Am. J. Enol. Vitic. 5 (2008) 317–322.
[28] A.C. Olivieri, N.K.M. Faber, J. Ferré, R. Boqué, J.H. Kalivas, H. Mark, Pure Appl. Chem. 78 (2006) 633–661.
[29] R. Bro, C.M. Andersen, J. Chemom. 17 (2003) 646–652.
[30] K. Kjeldahla, R.J. Bro, J. Chemom. 24 (2010) 558–564.
[31] A. Watanabe, S. Morita, Y. Ozaki, Appl. Spectrosc. 60 (2006) 611–618.
[32] J.S. Shenk, J.J. Workman, M.O. Westerhaus, in: D.A. Burns, E.W. Ciurczak (Eds.), Handbook of Near-Infrared Analysis, CRC Press, Florida, 2008, third ed..
[33] S.A. Arnold, L.M. Harvey, B. McNeil, J.W. Hall, BioPharm Int., 26–34.
[34] P.R. Mobley, B.R. Kowalski, J.J. Workman, R. Bro, Appl. Spectrosc. Rev. 34 (1996) 347–368.
[35] R.J. Barnes, M.S. Dhanoa, S. Lister, J. Appl. Spectrosc. 43 (1989) 772–777.
[36] F.S.J. Poke, J. Wood Chem. Technol. 26 (2006) 187–199.
[37] P.D. Jones, L.R. Schimleck, G.F. Peter, R.F. Daniels, A. Clark, Wood Sci. Technol. 40 (2006) 709–720.
[38] A. Vergnoux, M. Guiliano, D.Y. Le, J. Kister, N. Dupuy, P. Doumeng, Sci. Total Environ. 407 (2009) 2390–2403.
[39] K. Nkansah, A.B. Dawson, J. Slahor, Bioresour. Technol. 101 (2010) 4570–4576.
[40] S. Suchat, D. Pioch, S. Palu, E. Tardan, E.N. Van Loo, F. Davrieux, Ind. Crop Prod. 45 (2013) 44–51.
[41] S.S. Kelley, T.G. Rials, R. Snell, L.H. Groom, A. Sluiter, Wood Sci. Technol. 38 (2004) 257–276.
[42] A. Baum, J. Agger, A.S. Meyer, M. Egebo, J.D. Mikkelsen, New Biotechnol. 29 (2012) 293–301.
[43] L. Liu, X. Ye, A. Womac, S. Sokhansanj, Carbohydr. Polym. 81 (2010) 820–829.
[44] C. Krongtaew, K. Messner, T. Ters, K. Fackler, Bioresources 5 (2010) 2063–2080.
[45] J.B. Sluiter, R.O. Ruiz, C.J. Scarlata, A.D. Sluiter, D.W. Templeton, J. Agric. Food Chem. 58 (2010) 9043–9053.
[46] K. Mitsui, T. Inagaki, S. Tsuchikawa, Biomacromolecules 9 (2008) 286–288.
[47] R.C. Kuhad, A. Singh, Crit. Rev. Biotechnol. 13 (1993) 151–173.
[48] N.K.M. Faber, Anal. Chem. 71 (1999) 557–565.