Multi-product calibration models using NIR spectroscopy

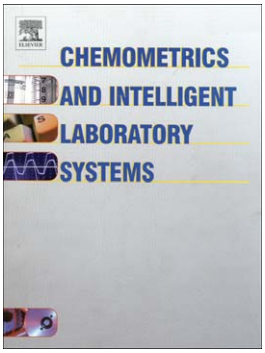
Article *in* Chemometrics and Intelligent Laboratory Systems · December 2015

DOI: 10.1016/j.chemolab.2015.12.013

3 authors, including:

Márcia M. C. Ferreira
University of Campinas
187 PUBLICATIONS **3,058** CITATIONS
SEE PROFILE

Edson Perito Amorim
Brazilian Agricultural Research Corporation (E…
95 PUBLICATIONS **422** CITATIONS
SEE PROFILE



Multi-product calibration models using NIR spectroscopy

M.K.D. Rambo, M.M.C. Ferreira, E.P. Amorim

PII: S0169-7439(15)00328-7
DOI: doi: 10.1016/j.chemolab.2015.12.013
Reference: CHEMOM 3152

To appear in: *Chemometrics and Intelligent Laboratory Systems*

Received date: 30 September 2015
Revised date: 11 December 2015
Accepted date: 16 December 2015

Please cite this article as: M.K.D. Rambo, M.M.C. Ferreira, E.P. Amorim, Multi-product calibration models using NIR spectroscopy, *Chemometrics and Intelligent Laboratory Systems* (2015), doi: 10.1016/j.chemolab.2015.12.013

Multi-product calibration models using NIR spectroscopy

M.K.D.Rambo [1]*; M.M.C.Ferreira[2;] E. P. Amorim[3]

[1] Department of Chemistry, University of Tocantins –UFT, CEP 77838-824, Araguaína, Brazil

[2] Institute of Chemistry, University of Campinas -UNICAMP, CEP 13083-970, Campinas, Brazil

[3] Embrapa Cassava & Fruits, 44380-000, Cruz das Almas, Brazil

**magalerambo@mail.uft.edu.br*

Abstract:

The physical-chemical composition of multiple biomasses can be predicted from one single calibration model instead of compositional prediction conducted by individual models. In this work, multi-product models, involving banana, coffee and coconut samples were built by partial least square regression (PLS) for ten different chemical constituents (total lignin, klason lignin, acid insoluble lignin, acid soluble lignin, extractives, moisture, ash, glucose, xylose and total sugars). The developed PLS models show satisfactory results, with relative error (RE%) less than 20.00, except for ash and xylose models; ratio performance deviation (RPD) values above than 4.4 and range error ratio (RER) values above 4.00. This means that all models are qualified for screening calibration. Principal component analysis (PCA) was useful to demonstrate the possibility and the rationale for combining three biomass residues into one calibration model. The results have shown the potential of NIR in combination with chemometrics to quantify the chemical composition of feedstocks.

Keywords: PCA; PLS; Chemical composition; Banana, Coffee, Coconut.

1. Introduction

Near infrared spectroscopy (NIR) has received considerable attention in the last years, as a tool for rapid, non-destructive, non-expensive (1–5% of the wet chemistry procedure cost), of simple application and that allows simultaneous assessment of multiple parameters of biomass composition [1,2]. The combination of NIR with chemometric tools allowed the development of multivariate calibration models for the rapid analysis of the chemical composition of feedstocks [3-7].

To ensure reliable prediction using the correlation of NIR spectra with the reference data from biomass composition, the NIR methods must be calibrated to an accurate primary reference analytical method. For this initial calibration, advanced multivariate models are developed, and although the process cost is slightly increased (30% of the wet chemistry procedure), they are still lower than the wet analysis [1]. Besides, another question raised when building calibration models, is the necessity to have a large variability of the calibration population and of the chemical characteristic of the samples [8].

Most frequently this variability is reached by sampling over different times and locations, what increases the process costs. To avoid such additional costs, some authors have used different botanical fractions from biomass to increase the variability in calibration models [8-10]. One promising alternative for increasing sample variability would be to use various feedstocks. However, literature [3,8] is scarce on the use of multi-biomass calibration models in which one single model combining different biomasses is developed.

According to Liu et al. [8] the main difficulty in building such models is associated to the dissimilarity among biomasses (different NIR spectra). It is s not practical to develop a NIR calibration model with species showing large dissimilarity.

So, to ensure a good prediction and reliable result, principal component analysis (PCA) was performed [11] to justify the development of a single calibration model containing three different biomasses. Besides, the usual statistical parameters (calibration and validation plots, calibration and validation errors, among others) were used to ensure the confidence of the models.

This study have shown that is feasible the arduous and costly process of sample collection over different times and from different locations was effectively replaced in a simple manner to use different types of biomass wastes to build single multivariate predictive models to analyze multiple constituents. Three quite distinct feedstocks (coffee, banana and coconut) and also different botanic fractions of each plant were considered. So, from the 10 different parameters (total lignin, klason lignin, acid insoluble lignin, acid soluble lignin, extractives, moisture, ash, glucose, xylose and total sugars) of physical-chemical composition analyzed, one model was built for each constituent, but useful for three singular feedstocks.

It proves that the NIR associated to multivariate analysis can be used for screening calibration and quality control to estimate physical-chemical content in biomass residues.

2. Material and methods

2.1. Sample collection

A total of 104, 101 and 28 samples of banana, coffee and coconut residues of different botanical parts were collected as illustrated in figure 1.

Also, among the different fractions, samples from different locations, soils, cultivars, species and harvest time were sampled to ensure the variability.

Of the 233 samples collected, not all were subjected to the wet analysis steps. All the 233 samples were analyzed for moisture, extractive and ash. The analyses of

soluble and insoluble lignin contents were carried out for 137 samples, and for sugars only 94 samples were analyzed.

2.2. Physical-chemical analysis

All the samples were dried, mill and then sieved to a homogeneus particle size of 180–850 µm. The biomass analyses (all in duplicate) of extractives, lignins and sugars were carried out using standard National Renewable Energy Laboratory (NREL) methods [12,13]. For extractives (NREL/TP-510-42619, 2008), the accelerated solvent extraction with 95% ethanol in a Dionex ASE 200 system (Thermo Fisher Scientific, Waltham, MA, USA), was used. Acid hydrolysis (NREL/TP-510-42618, 2011) on the extracted samples was carried out with sulfuric acid 72% in a water bath in the first step, followed by hydrolysis for 1 h at 120 ºC (in autoclave) and an acid concentration of 4%. In the hydrolysis step the lignin (soluble and insoluble) and sugar contents were determined. The acid soluble lignin (ASL) content was determined by UV-spectroscopy in a Shimadzu UV-1700 spectrometer (Shimadzu, Kyoto, Japan), at wavelength of 205 nm. Insoluble lignins (klason lignin (KL) and acid insoluble residue (AIR) were determined by gravimetry, and sugars were determined by high pH anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) for the monossacharides.

The moisture (105 ºC) and ash (600 ºC) analyses were carried out using ASTM 3173-87 [14] and ASTM D 3174-04 methods [15], respectively.

2.3. Multivariate calibration models

The Vis-NIR spectra (400-2500 nm) were collected using a FOSS XDS instrument (FOSS, Hillerød, Denmark). Each spectrum was generated by averaging 32 scans, with 0.5 nm of increment. Two spectra were collected for each sample and the average spectrum was used for data analysis.

Initially, all the 233 raw spectra were submitted to PCA with varimax rotation to reveal the data structure and identify similarity/dissimilarity among the three feedstocks.

Partial least squares regression (PLS1) was used to obtain the multivariate calibration models using the Unscrambler 10.2 (Camo Software, Oslo, Norway). The data set was randomly split into two subsets: the calibration set consisting of 75% of the samples and the external validation set with the remaining 25% of samples. The external validation set may be used to determine the number of latent variables (LV), and is often cited as the most realistic estimate, particularly of the prediction errors. However, it requires a large amount of samples [16,17], such as in the present study. These models were developed with the spectra transformed by taking the Savitzky-Golay second (2D) derivative using a second-order polynomial, with a window of 15 and 25 points [18]. For the extractive model, the best results were obtained by combining the standard normal variate (SNV) with first (1D) derivative transformations using a second-order polynomial, with a window of 2 points [19].

For each model, the coefficient of determination (R^2_{cal} and R^2_{val}), the root mean square error of calibration (RMSEC), the root mean square error of prediction (RMSEP), the standard error of calibration (SEC), the standard error of prediction (SEP) and the numbers of outliers and LV, were obtained. The error vector, **e**, which is the difference between the reference values and their estimates in calibration set ($\mathbf{e}_{cal}$) and validation set ($\mathbf{e}_{val}$), were calculated. Also the relative error (RE), the range error ratio (RER), the ratio performance deviation (RPD), the bias, the test t and the test F of Snedecor were calculated and used as the criterions of performance for the predictions on the calibration and validation sets according to the ASTM-1655-05 [20] rules and Fearn [21]. The regression coefficients were interpreted to show the physical meaning of the models.

All collected data were organized in plots and figures using Origin 8.0 (Northampton, MA, USA).

3. Results and discussion

The Vis-NIR raw spectra and the second derivative spectra of the banana, coffee and coconut samples are shown in Figures 2 (A) and 2 (B). The main bands for the three biomass sets are located in the same wavelength region for the raw spectra as also for the 2D spectra. The 2D spectrum is a measure of the change in the infinitesimal slope of the curve and can help solve nearby peaks and sharpen spectral characteristics. However, the maximum of the bands undergo a minimum (Figure 2B).

The bands at 460 and 670 nm are both attributed to lignin and chlorophyll structures, included conjugated π-bond system chromophores. Other bands appear at 1170 nm (2^{nd} overtone of C-H stretch of lignin structures), 1434-1470 nm (assigned to 1^{st} overtone of O-H stretch of structures of polysaccharides of OH groups with H-bonds) and 1724 nm (with two overlaps, assignment to C-H stretch of 1^{st} overtone of CH_2 of lignin or than of CH of furanose or pyranose due to hemicellulose). The band at 1920 nm is probably assigned to O-H stretch and OH bend of polysaccharides structures which overlaps with water. Above 2000 nm, there are the combinations bands, at 2090 (O-H combination band of carbohydrates) and 2329 nm attributed to C-H stretch or C-H combination band of polysaccharides [22-27].

Figure 3 presents descriptive statistics (mean and standard deviation), for the chemical constituents (%) of the sets comprising samples of all biomasses (Total) as well as of each feedstock separately.

It can be seen that total sugar (TS) is the major constituent (40.0% on average for the total set of samples and 51.8% for banana) and the minor constituents are ash and ASL for coconut samples (average of 1.35 and 1.44, respectively). The highest and lowest range were observed for TS from coconut samples, with a wide variation in the

standard deviation (12.4%) and for ash from coffee samples (standard deviation of 0.46%), respectively. The coefficient of variation (CV) ranged from 16.7 (KL) to 93.7 (ASL), for coffee and coconut, respectively. The average values found for CV and standard deviations for the constituents in all sets were 35.4% and 5.0%, respectively.

The range in composition for these lignocellulosic constituents is wide, as result of the sampling used, that included different lignocellulosic biomasses and botanical fractions.

Table 1 shows the Pearson correlation coefficient between the 10 constituents, using the average value of each parameter of the 230 samples. Only absolute values above 0.5 were considered significant.

The ash content is strongly negatively correlated with all constituents, except for sugars (Glu and TS, with 0.744 and 0.749, respectively). The opposite occurs to the moisture content, which shows positive correlation with lignins and negative correlation with sugars. Hayes et al., [5] observed the opposite, where the ash content was negatively correlated to most of the sugars.

Extractives show a positive correlation with insoluble lignins and xylose. A negative correlation might be expected for ASL. This positive relationship could indicate that extractives present some components that can be attributed to lignins and sugars. Ethanol extractives, for example, can include non-structural sugars, organic acids, chlorophyll and other components [28, 29]. On the other hand, the negative correlation with ASL indicates that the extractives were not condensed or precipitated, even under the strong acidic conditions used in the acid hydrolysis stage [30].

The KL, AIR and LT are negatively correlated with the sugars (what is expected) and are strongly negatively correlated with ash. The higher the content of lignin in the lignocellulosic biomass, the lower the level of sugar content.

Sugars present a positive correlation with ash content and a negative correlation with all lignin constituents. However strong positive correlations between glucose and TS were observed (> 0.999) and also a minor but positive correlation with xylose was found. This is expected once glucose and xylose are the major components of TS. Hayes et al., [5] observed the same significant correlation between TS and glucose.

3.1. Principal component analysis

The results from PCA applied to the raw spectra of banana, coconut and coffee, on the mean centered data, after performing varimax rotation, are shown in Figure 4.

The first two PC explained 39 and 26% of the total variance, respectively. The remaining PC explained 35% of the cumulative variance.

Except for PC1 (Figure 4A), the visible region of spectrum presented high loadings in all PC, with high weights in this region. PC1 is characterized by negative loadings around 1900 nm, typical of OH first stretch overtone probably due to cellulose [25]. This indicates that the coffee husks have lower percentages of cellulose than coconut and banana biomasses, which is supported by analytical data (Reference method) shown in Figure 3, where is possible to observe a higher average value of glucose (assigned to cellulose) for coconut and banana, than for coffee. On the other hand, PC2 was characterized by positive loadings at 540 nm, characteristic bands of chlorophyll [22] and, as expected, that differentiates leaf samples rich in these photosynthetic pigments, as well as the coffee samples with positive scores (Figure 4B). PC3 and PC4 have positive loadings at 470 and 677 nm, both associated to chlorophyll [22]. Most coffee samples and some banana samples show negative scores in PC3, probably assigned to C-H or CH_2 stretch of lignin structures, associated with the negative loadings of 760 nm in PC3.

By the analysis of the scores plot it was not possible to distinguish between banana and coconut samples based on their NIR spectra. In addition, the coffee samples were somewhat distant from the other groups in the scores plots. But, good calibration models have been already reported in the literature by combining coffee with banana biomasses [3]. Consequently, the above discussion justifies the combination of three different biomasses in a single calibration model, with the advantage of covering a wide range of variation and being as generic as possible.

3.2. Partial least squares regression

All the mathematical equations and statistics used are in accordance to the ASTM 1655-05. The results obtained for the multi-product (banana, coffee and coconut) calibration models from ten constituents of interest (total lignin (TL), Klason lignin (KL), acid insoluble lignin (AIR), acid soluble lignin (ASL), extractives (Extrac.), moisture (Moisture), ash (Ash), glucose (Gluc.), xylose (Xyl.) and total sugars (TS)) are summarized in Tables 2 and 3.

In table 2 it is possible to observe that all models were built with a maximum of 7 LV and no more than 6.2% of outliers were removed. The RE were high for extractives, ash and xylose (> 19.0%). Satisfactory results of RE were found for TS and TL (≤10), as well as good RER values, above > 10.0 indicating models acceptable for screening procedures. Prediction capacity of the models can be evaluated with the RPD, where values > 4.4 means that the models have good prediction accuracy [21]. According Williams [31] RPD values above 9.0 indicate excellent models, what occurs for TL and TS models.

The other parameters (soluble and insoluble lignins, moisture and glucose) presented reasonable results, with RE lower than 14.40%, RPD above 4.4 and RER above 8.50.

Liu et al., [8] evaluated the performance of broad-based models including three different biomass species; corn stover, switchgrass and wheat-straw samples. For the same constituents modeled in this work; glucose, xylose, lignin and ash, the authors found very good results, with RE values less than 14% and RER values higher than 11.23. For the constituent "TL" they have obtained RER and RE values of, 11.23 and 3.62%, respectively, what it in good agreement with the results obtained in this work for the same parameter (13.0 and 7.70, respectively). For the carbohydrates, glucose and xylose, Liu et al., [8] shows RE less than 2.37% and RER values of 12.58 and 12.87, respectively.

The Ash model in this work can be considered as moderately useful for prediction (semi-quantitative), because shows a $R^2_{val} < 0.80$ and high error (>20%). Liu et al., [8] also foundd high RE values (13.85%) when modeling ash.

The work proposed by Liu et al., [8] involved samples that presented a certain similarity in their chemical composition, which can facilitate the performance of the models. In the present study, one large variability was sampled, with quite different biomasses grouped in one single calibration model.

Hayes et al., [5] analyzed the lignocellulosic components of peat samples by near infrared spectroscopy and chemometric models for rapid quantitative predictions. All the results found were satisfactory, with $R^2_{val} > 0.87$ and RER> 8.5, except for extractives model, where these statistical parameters were found to be 0.769 and 7.04, respectively.

Godin et al., [32] predicted chemical characteristics of fibrous plant biomasses from NIR spectra and found $R^2_{val} = 0.92$ for KL, which is the same value obtained in the present study (Figure 5). The model was considered successful for prediction, because presented a RPD [21] value higher than 3.0. In this work, the RPD value obtained was

8.30 for KL, and therefore considered also acceptable for quality control. Godin [32] evaluated also the properties, TS and glucose, with values of R^2_{val} of 0.94 and 0.00. Comparing with the models obtained in this work the values are better for glucose (0.85), but worse for TS (0.84). Glucose and TS models propose in the present work are successful prediction models.

Is important to note that in this study three different biomass samples were included in the development of the ten calibration models, instead of one as Hayes [5] and Godin [32] used in their works. As a consequence, the performance of the calibration models can be affected.

The regression plots (calibrations and external validation) of the reference versus the predicted values from the multivariate models are show in Figure 5.

The bias is an indication of the systematic error that occurs when a plant species is predicted without being in the calibration set [32]. So, the *t* Test (95% probability) was used to determine if the validation estimates show a statistically significant bias. Except for the TS model (Table 3), all other chemical properties presented values lower than the $t_{critical}$ value, indicating that the analyses based on multivariate models are expected to give essentially the same average result as the measurements conducted by the reference method. For TS model, there is a 95% probability that the values estimated by the model will not give the same average results as the reference methods, indicating that the validation estimates show a statistically significant bias.

The calibration and validation error vectors ($e_{c,v}$) were lower than 5.00% for all parameters, except for extractive and sugar models (Table 3). According to the F test (95%), moisture, extractives and xylose presented significant differences between SEP and SEC values.

4. Conclusion

The multivariate models were reliable for the prediction of chemical composition of different biomasses species in a single multi-product model, and can be used for screening calibration, quality control and quantitative analyses of the main chemical component contents in biomasses. It was observed by the statistical parameters, that all the models show RER and RPD values higher than 4.0, RE less than 20.0% and R^2 cal, val > 0.80, except for ash and xylose models.

The results showed the potential of a robust and reliable predictive model using multiple biomass species, with great variability in the chemical composition. Furthermore, this alternative sampling approach avoids some problems, such as expensive costs and time-consuming collection of diverse sample throughout years and different locations, favoring the fast biomass compositional analysis. In this work, three biomasses were investigated but this number can be even higher for a biomass belonging to the same applicability domain.

Acknowledgments

The authors thank the Sugarcane Research Center-CTC, Brazilian Agricultural Research Corporation-EMBRAPA and Maria Helena Monteiro from Fazenda Monte Alto for kindly providing the samples and facilities to perform this study. The DIBANET research leading to these results has received funding from the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement n° 227248.

References

[1] B. Hames, S. Thomas, A. Sluiter, C. Roth, D. Templeton, Rapid biomass analysis:new tools for compositional analysis of corn stover feedstocks and process intermediates from ethanol production., Appl. Biochem. Biotech. 105 (2003) 5-16.

[2] C. Pasquini, Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. J. Braz. Chem. Soc. 14 (2003) 198-219.

[3] M.K.D. Rambo, E.P. Amorim, M.M.C. Ferreira, Potential of visible-near infrared spectroscopy combined with chemometrics for analysis of some constituents of coffee and banana residues. Anal. Chim. Acta, 775 (2013) 41-49.

[4] F. Xu, J. Yu, T. Tesso, F. Dowell, D. Wang, Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review. Appl. Energy 104 (2013) 801-809.

[5] D.J.M. Hayes, M.H.B. Hayes, J.J. Leahy. Analysis of the lignocellulosic components of peat samples with development of near infrared spectroscopy models for rapid quantitative predictions. Fuel 150 (2015) 261–268.

[6] M.K.D. Rambo, A.R. Alves, W.T. Garcia, M.M.C. Ferreira. Multivariate analysis of Coconut residues by Near infrared spectroscopy. Talanta 138 (2015) 263-272.

[7] M.K.D. Rambo, M.M.C. Ferreira, Determination of cellulose crystallinity of banana residues using near infrared spectroscopy and multivariate analysis. J. Braz. Chem. Soc. 26 (2015) 1491-14999.

[8] L. Liu, X. P. Ye, A. R. Womac, S. Sokhansanj, Variability of biomass chemical composition and rapid analysis using FT-NIR techniques. Carbohydr. Polym. 81 (2010) 820–829.

[9] T.A. Lestander, C. Rhen, Multivariate NIR spectroscopy models for moisture, ash and calorific content in biofuels using bi-orthogonal partial least squares regression. Analyst 130 (2005) 1182–1189.

[10] Y.X. Philip, L. Liu, D. Hayes, A.Womac, K. Hong, S. Sokhansanj, Fast classification and compositional analysis of cornstover fractions using Fourier transform near-infrared techniques. Bioresour. Technol. 99 (2008) 7323–7332.

[11] S. Wold, K.Esbensen, P. Geladi, Chemometrics and Intelligent Laboratory Systems, 2 (1987) 37-52.

[12] NREL/TP-510-42619. Determination of Extractives in Biomass. National Renewable Energy Laboratory, 2008.

[13] NREL/TP-510-42618. Determination of Structural Carbohydrates and Lignin in Biomass. National Renewable Energy Laboratory, 2008.

[14] ASTM D 3173-87. Standard method for determination of moisture content in biomass. Society for Testing Materials, International, 2003.

[15] ASTM D 3174-04. Standard method for ash in the analysis sample of coal and coke. Society for Testing Materials, International, 2004.

[16] N.K.M. Faber, R. Rajkó, How to avoid over-fitting in multivariate calibration — The conventional validation approach and an alternative. Anal. Chim. Acta. 595 (2007) 98–106.

[17] K. H. Esbensen, P. Geladi, Principles of Proper Validation: use and abuse of re-sampling for validation. J. Chemometr. 24 (2010) 168–187.

[18] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chemom. 36 (1964) 1627-1638.

[19] T. Fearn, Are two pretreatments better than one? NIR news 14 (2003) 9-11.

[20] ASTM 1655-05. Standard Practices for Infrared Multivariate Quantitative Analysis, 2005.

[21] T. Fearn, Assessing calibrations: SEP, RPD, RER and R^2. NIR news 13 (2002) 12-14.

[22] A. A. Gitelson, Y. Gritz, M.N. Merzlyak, Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J. Plant Physiol. 160 (2003) 271–282.

[23] S. Tsuchikawa, H. Siesler, Near-Infrared Spectroscopic Monitoring of the Diffusion Process of Deuterium-Labeled Molecules in Wood. Part II: Hardwood. Appl. Spectrosc. 57 (2003) 675-681.

[24] A. Watanabe, S. Morita, Y. Ozaki, Temperature-dependent structural changes in hydrogen bonds in microcrystalline cellulose studied by infrared and near infrared spectroscopy with perturbation-correlation moving-window two-dimensional correlation analysis. Appl. Spectrosc. 60 (2006) 611-618.

[25] J. S. Shenk, J. J. Workman, M. O. Westerhaus, Application of NIR Spectroscopy to Agricultural Products. In: (3 ed.). Burns, D. A. & Biurczak, E. W. Handbook of Near-Infrared Analysis, Florida, 2008.

[26] T. C. M. Pastore, C.C.K.Oliveira, J.C. Rubim, K.O. Santos, Efeito do intemperismo artificial em quatro madeiras tropicais monitorado por espectroscopia de infravermelho (DRIFT). Quim. Nova 31 (2008) 2071-2075.

[27] K. Mitsui, T. Inagaki, S. Tsuchikawa, Monitoring of hydroxyl groups in wood during heat treatment using NIR spectroscopy. Biomacromolecules 9 (2008) 286-288.

[28] Sluiter, J. B. *et al*. Compositional Analysis of Lignocellulosic Feedstocks. Review and Description of Methods. J. Agr. Food Chem. 58 (2010) 9043–9053.

[29] S.F. Chen, R.A. Mowery, C.J. Scarlata, C.K. Chambliss, Compositional analysis of water-soluble materials in corn stover. J. Agric. Food Chem. 55 (2007) 5912–5918.

[30] K. Thammasouk, D.Tandjo, M.H. Penner, Influence of extractives on the analysis of herbaceous biomass. J. Agric. Food Chem. 45 (1997) 437–443.

[31] P.C. Willians in Near Infrared Technology in the Agricultural and Food Indsutries, 2nd Edition , Ed by P.C. Wilians and K.H.Norris. American Association of cereal Chemists, St Paul, USA, p.164, 2001.

[32] B. Godin, R. Agneessens, J. Délcarte, P. Dardenne, Prediction of chemical characteristics of fibrous plant biomasses from their near infrared spectrum: comparing local versus partial least square models and cross- validation versus independent validations. J. Near Infrared Spectrosc. 23 (2015) 1–14.

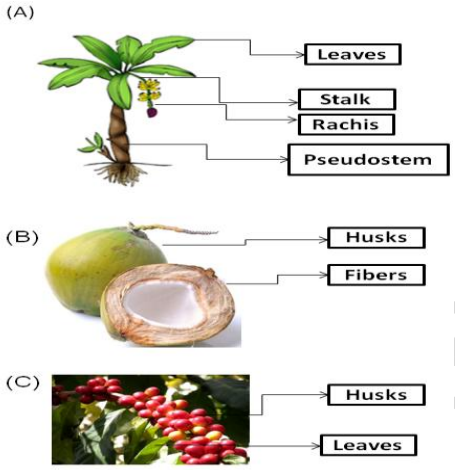


Fig. 1

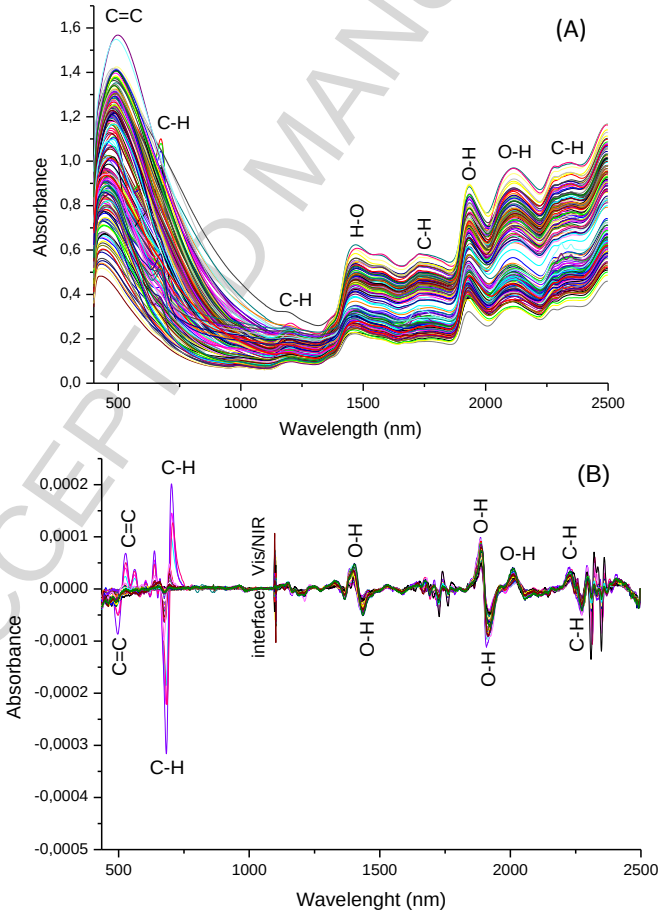


Fig. 2

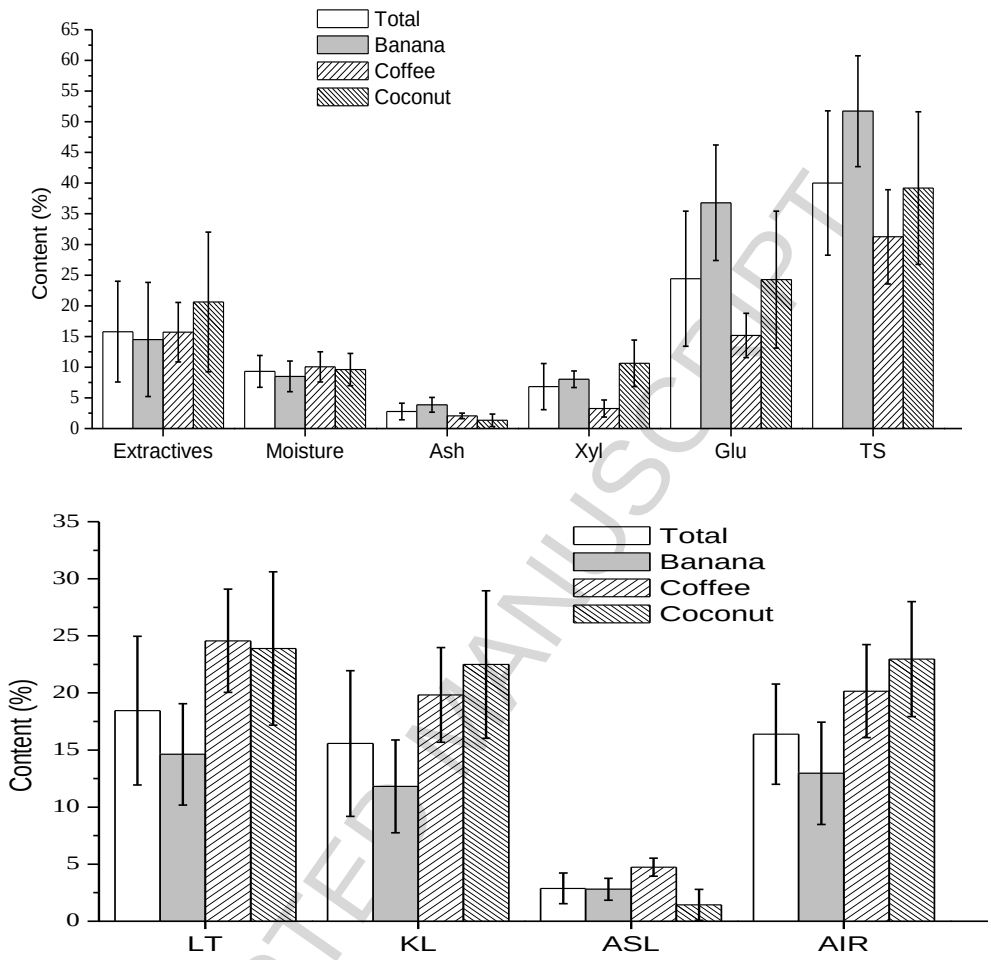


Fig.3

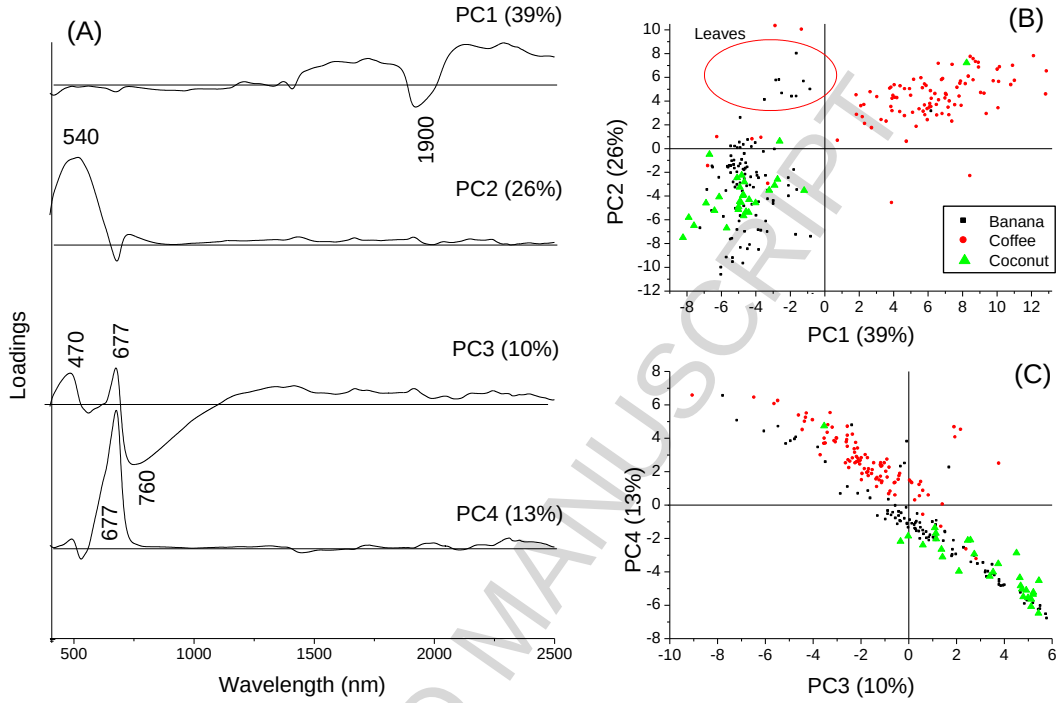
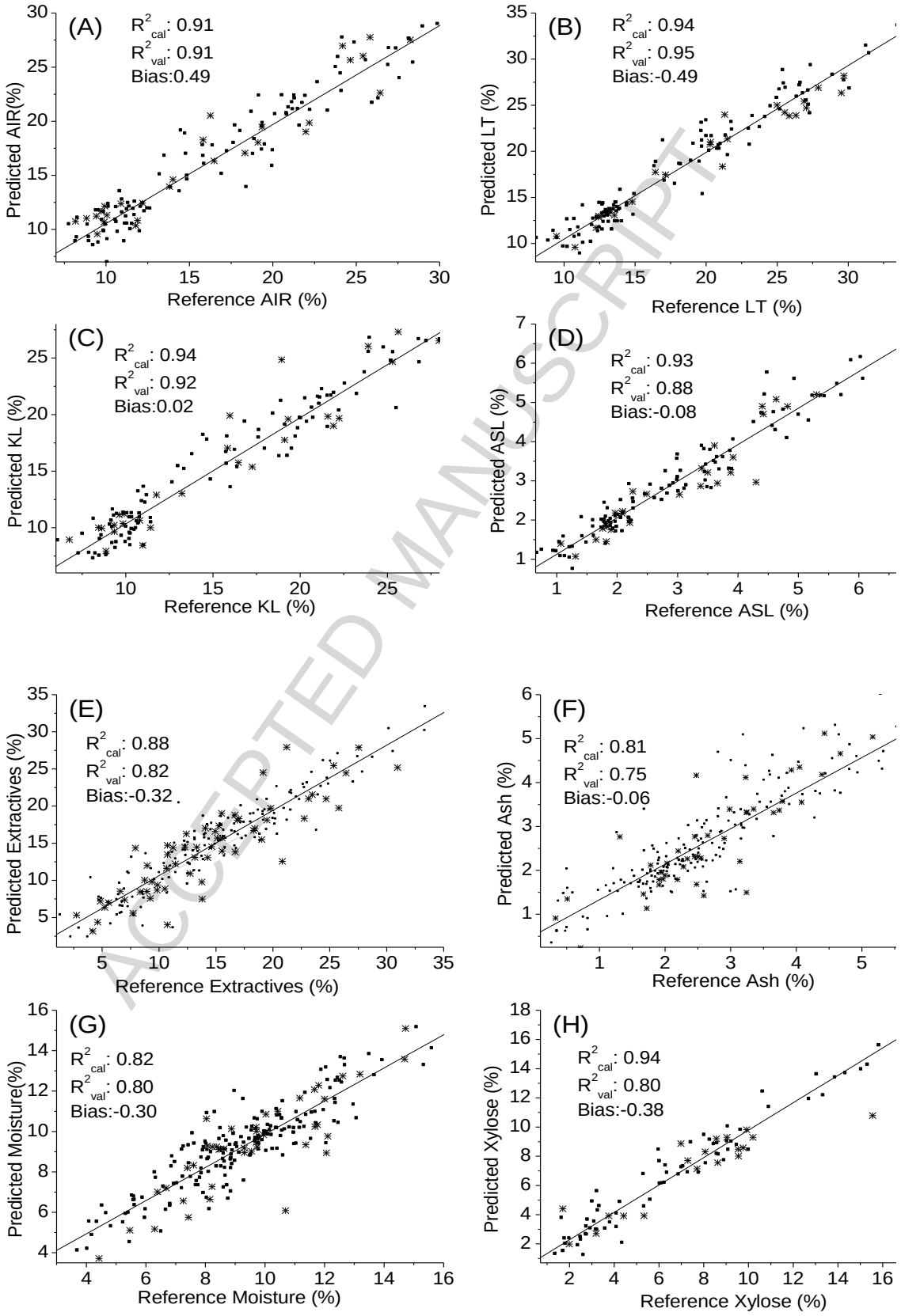


Fig.4



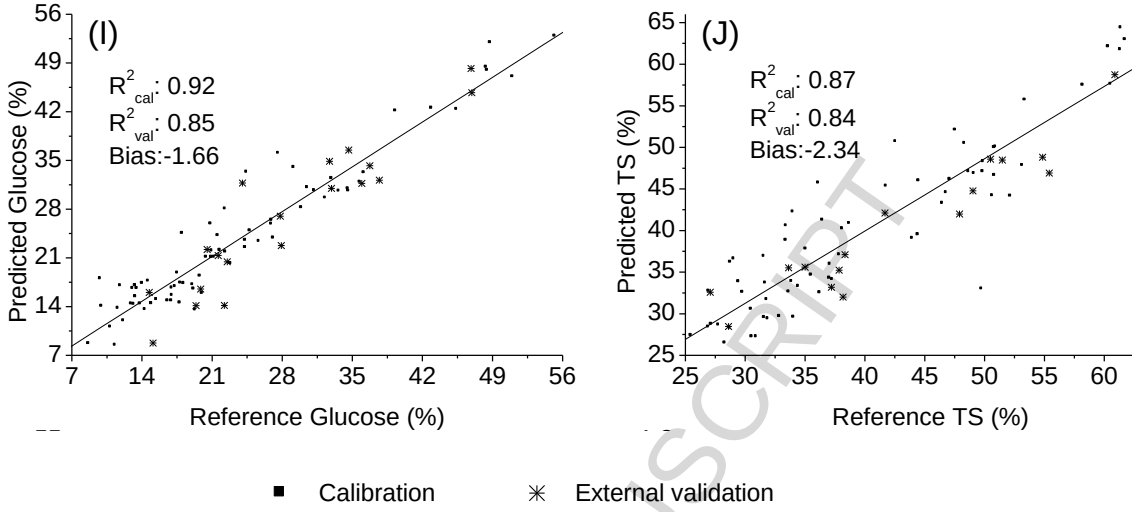


Fig.5

Table 1. Linear correlation between the constituents of samples from the three biomasses.

	Ash	Moist	Extrac.	ASL	KL	AIR	LT	Glu	Xyl	TS
Ash		**-0,850**	**-0,835**	0,187	**-0,999**	**-0,997**	**-0,966**	**0,744**	-0,155	**0,749**
Moistu			0,422	0,356	**0,837**	**0,815**	**0,956**	**-0,984**	-0,386	**-0,985**
Extract				**-0,696**	**0,849**	**0,868**	**0,667**	-0,255	**0,672**	-0,262
ASL					-0,212	-0,250	0,069	**-0,515**	**-0,999**	**-0,509**
KL						**0,999**	**0,960**	**-0,727**	0,180	**-0,732**
AIR							**0,948**	**-0,700**	0,218	**-0,705**
TL								**-0,890**	-0,102	**-0,893**
Glu									**0,543**	**0,999**
Xyl										**0,537**
TS										

Table 2. Statistical results of single calibration models for the ten chemical properties from the three biomasses.

y	Pre-treatment	Matrix size	LV	Outliers	RMSEC	RMSEP	RE	RER	RPD
TL	2D(25)	129X2800	7	8	1.550	1.478	7.70	14.24	13.9
KL	2D(25)	130X2800	5	7	1.444	1.803	11.80	11.55	8.30
ASL	2D(25)	133X2800	7	4	0.352	0.410	14.00	10.40	7.36
AIR	2D(25)	136X2800	5	1	1.766	1.899	11.35	9.44	7.63
Moisture	2D(25)	228X2800	7	5	1.038	1.335	14.15	12.43	7.40
Extrac.	SNV+1D(3)	222X2800	7	11	2.515	3.203	19.0	12.06	4.70
Ash	2D(25)	228x2800	6	8	0.587	0.619	21.80	9.03	4.56
Xyl.	2D(15)	89x2800	4	5	0.939	1.520	21.70	9.16	4.91
Gluc.	2D(15)	92x2800	7	2	3.106	4.058	14.40	19.13	7.50
TS	2D(15)	90x2800	7	4	4.290	4.151	10.00	10.02	12.13

Table 3. Statistics used in evaluating to data in calibration and validation set.

y	e_c	SEC	e_v	SEP	*t Test*	*F Test*
LT	5.00	1.55	3.19	1.42	1.72	1.19
KL	3.68	1.45	3.69	1.83	0.05	1.59
ASL	1.24	0.35	1.33	0.40	1.12	1.30
AIR	4.62	1.77	4.27	2.12	1.24	1.43
Moisture	2.96	1.04	4.59	1.31	1.55	1.58*
Extrac.	8.79	2.31	8.92	3.24	0.74	1.96*
Ash	1.66	0.58	1.74	0.62	0.70	1.14
Xyl.	2.38	0.94	4.75	1.51	1.09	2.54*
Gluc.	8.67	3.12	8.05	3.80	1.91	1.48
TS	16.94	4.32	8.48	3.54	2.64*	1.06

*: *t* value is greater than the tabulated *t* value; and F test presented significant differences.

Captions to figures

Fig. 1. Botanical fractions sampling of (A) banana (B) coconut and (C) coffee.

Fig. 2. Vis-NIR raw spectra (A) and Vis-NIR second derivative spectra.

Fig. 3. Mean and standard deviation of reference analysis for all samples and each biomass separately.

Fig. 4. (A) Loadings plot from PCA analysis. (B) Scores plot of the first two principle components. (C) Scores plot of the third and four principle components for the Banana, Coffee and Coconut biomasses.

Fig. 5. Plot of reference *versus* predicted values from the calibration and external validation models from (A) AIR content; (B) LT content; (C) KL content; (D) ASL content; (E) extractives content; (F) ash content; (G) moisture content; (H) xylose content; (I) glucose content; (J) TS content.

Highlights

>Potential of near-infrared spectroscopy and chemometrics for screening calibration, quality control and quantitative analyses of the biomass components > Principal component analysis (PCA) to demonstrate the possibility for combining three biomasses into one calibration model> robust and reliable predictive PLS models using multiple biomass species.