D. Sabatier, P. Dardenne and L. Thuriès, *J. Near Infrared Spectrosc.* **19**, 199–209 (2011)
Received: 14 February 2011 , Revised: 2 May 2011 Accepted: 12 May 2011 Publication: 4 July 2011

JNIRS JOURNAL OF NEAR INFRARED SPECTROSCOPY

Near infrared reflectance calibration optimisation to predict lignocellulosic compounds in sugarcane samples with coarse particle size

Damien Sabatier,[a,*] Pierre Dardenne[b] and Laurent Thuriès[c]

[a]CIRAD, UPR SCA, F-97494 Sainte-Clotilde, La Réunion, France. E-mail: damien.sabatier@cirad.fr

[b]CRA-W, Quality of Agricultural Products Department, Chaussée de Namur n° 24, 5030 Gembloux, Belgium

[c]CIRAD, UPR Recyclage et Risque, F-97408 Saint-Denis, La Réunion, France

Frequent variations in spectral intensity due to particle size and/or of particle size distribution are observed in plant products processed in powder form and scanned with near infrared reflectance (NIR). In this study, two grinders, with differences in time consumption, practicality and providing homogenates with different particle size range and distribution, were tested to evaluate their effects on NIR spectra. Optimisation of NIR calibration was necessary before predicting lignocellulosic compounds in sugarcane (*Saccharum spp.*) samples with coarse particle size to supply a pre-existing ecophysiological growth model. Sixty samples from three varieties, grown in four contrasting pedoclimatic areas and from five anatomical parts were scanned and then analysed by biochemical fractionation. Different calibration methods, resulting in a combination of multiple linear regressions (MLR) applied to three calibration sets (fine, coarse and mixed particle sizes) treated with six data pretreatments—first derivative (D), second derivative (D^2), multiplicative scatter correction (MSC), standard normal variate and detrend (SNVD), standard normal variate and detrend successively followed by first derivative (SNVD-D) or second derivative (SNVD-D^2)—were investigated. The best NIR model statistical values were obtained by calibration developed on a mixed calibration set treated by SNVD-D^2. Results confirmed that NIR spectroscopy could be an accurate and efficient method to predict lignocellulosic compounds in different botanical parts of sugarcane samples when used as input to an ecophysiological growth model.

Keywords: near infrared spectroscopy, calibration, spectral pretreatment, grinding methods, particle size, chemical composition, sugarcane

Introduction

Near infrared (NIR) reflectance spectroscopy is currently used for the rapid determination of chemical compounds in granular or powdered samples.[1] However, for products scanned in powder form, the optical thickness cannot be controlled by the operator; it depends on the granular distribution of the particles of the ground matrix. Moreover, samples with coarser particle size appear to have a more intense spectrum, primarily due to lower specular reflection.[2] Blanco *et al.*[3] reported that there is a decrease in spectral reproducibility with an increase in the particle size of the sample. Williams and Thompson[4] have demonstrated that the most important factors affecting the accuracy of NIR spectroscopy in the analysis of hard red spring wheat were mean particle size and particle size distribution. Osborne *et al.*[5] reported that the effects of particle size and particle size distribution were quantitatively the most important factors in NIR spectra. Dhanoa *et al.*[6] showed that

ISSN: 0967-0335
doi: 10.1255/jnirs.929

the diversity of material particle size affected scattering and was a major source of variation in NIR spectra.

A significant number of studies conducted on the effect of particle size over recent years underlined that which Barnes *et al.*[7] pointed out as a long-standing problem, continues to be a problem during NIR calibration. These studies can be divided into two broad categories: those linked to the development of NIR models for the determination of physical characteristics such as particle size;[8-11] and others focused on the effect of particle size on spectral data acquisition.[12-20] NIR calibration involves maximising the desired signal and minimising the unwanted signal.[1] In our study, the effect of particle size was considered to be an unwanted signal because it may mask the chemical information contained in raw NIR spectra. Norris and Williams[12] reported that normalised second derivative treatment of spectral data best predicted the protein content of samples with widely varying particle size. These authors also demonstrated that including particle size variation in the calibration samples improved predictions. In fact, different methods of normalising or "correcting" (i.e. pretreatment of) spectra are used in NIR calibration to remove physical information such as the particle size of products in powder form.[5-7,12,21-24] To date, no study has been conducted on sugarcane and no model of prediction of lignocellulosic compounds in sugarcane samples from different botanical parts has been developed. The primary aim of this study was to investigate two different grinding methods both for their advantages and to evaluate their effects on NIR spectra. The second aim was to develop an optimum calibration with a combination of multivariate analysis, data pretreatments and calibration sets, to predict lignocellulosic compounds of sugarcane samples with a wide range of particle sizes. Finally, we wanted to know if we could predict, without loss of accuracy, lignocellulosic compounds in different botanical parts of sugarcane samples prepared using a faster and more convenient grinding method which resulted in coarser particles. Data on these compounds are required as input for a sugarcane growth model.[25]

Materials and methods

Experimental sugarcane samples

The study was based on 60 samples of three sugarcane varieties (*Saccharum spp.* R570, R579 and R585) from Reunion Island. These three varieties are highly variable in their agronomic characteristics. They were grown in four different areas of La Reunion Island and harvested at different physiological ages (200 to 298 days after planting). After harvesting, the whole plants were separated into five anatomical fractions (millable stalk, top of the stalk, green leaf blade, green leaf sheath and trash) which increased the variability of the samples (Table 1). It results from these sampling campaigns 60 samples = three varieties * four sites * five anatomical fractions.

After separation, approximately 15 kg of millable stalk were roughly crushed with a cutter grinder (Jeffco Cutter Grinders, model L118C, Jeffress Engineering Pty Ltd, Dry Creek, Australia) to obtain sugarcane pulp. A sub-sample of 1000 g of sugarcane pulp was weighed and pressed for 1.5 min at 200 bars with a hydraulic press (Pinette Emidecau press model OB-103, Pinette Emidecau Industries, Chalon sur Saone, France). After crushing and pressing, a filter press cake of bagasse was obtained. The other anatomical parts were simply chopped to facilitate drying. The whole filter press cake of bagasse and a 500 g sub-sample of each of the other anatomical parts were dried for three days at 70°C in a ventilated incubator to obtain dry matter at a constant weight (DM).

Grinding methods

Once the samples were dried, they were first ground with a Thomas Wiley (TW) knife crusher (Swedesboro, NJ, USA) to pass through a 2 mm screen and then a subsample of each samples ground with TW were ground with a Foss-cyclone crusher Cyclotec (CT; Nanterre, France) through a 1 mm screen, as recommended for subsequent fibre analyses by Van Soest and Wine.[26] Grinding time was 1 min for both methods. Mean particle size of the TW and CT samples were respectively 428 µm and 286 µm. The elementary particles of TW samples were mainly coarse and of varying size, ranging from up to 1 mm to 160 µm; and CT samples varied in a more narrow range of size, from 630 µm to 160 µm (Figure 1).

Chemical analysis

Only CT samples were analysed according to the method reported by Van Soest *et al.*[27] and reviewed by Mertens *et al.*[28] CT samples were successively extracted for neutral detergent fibre (NDF) with alpha amylase, acid detergent fibre (ADF), and acid detergent lignin (ADL). To remove residual humidity before and at each extraction step, the raw matrix and the products obtained were dried at 70°C overnight and weighed. Then the raw matrix and each product was ignited gradually by raising the temperature to 525°C and maintaining it for five hours to measure ash content (Ash, Ash_{NDF}, Ash_{ADF} and Ash_{ADL}, respectively), in order to calculate ash free lignocellulosic fractions (NDF_{om}, ADF_{om} and Lig).

Spectra collection

Two replicates of TW and CT samples were packed into circular cups (50 mm in diameter) closed with a cardboard lid, then scanned in reflectance mode on a monochromator spectrometer (NIRS XDS, Silver Spring, MD, USA). Spectra were obtained for each replicate at 2 nm intervals over the 400–2498 nm wavelength range. The spectra were recorded as log 1/reflectance (log 1/*R*). Each sample was scanned twice (two different cup fillings) and spectra were averaged for each sample and saved in a different file for each grinding method (TW and CT sets). TW and CT sets were used in the calibration sets CS_{TW} and CS_{CT}, respectively (Figure 2). A third calibration set (CS_{MIX}) was made by mixing the CS_{CT} with 30 samples chosen at random in CS_{TW} to include wide variations in particle size in CS_{CT} (ratio 2CT : 1TW).

Table 1. Description and chemical composition of the experimental sugarcane (*Saccharum sp.*) samples from four contrasting pedoclimatic areas of Reunion Island collected in 2009 including both calibration and validation sets.

Sugarcane samples				Chemical composition % DM							
Botanical parts	Varieties			NDF_{om}		ADF_{om}		Lig		Ash	
	R570	R579	R585	Mean+SD	Range	Mean±*SD*	Range	Mean±*SD*	Range	Mean±*SD*	Range
Calibration											
Trash	4	4	4	73.9±4.0	64.3–78.8	41.4±2.4	35.8–45.1	4.5±0.6	3.2–5.5	9.4±4.1	4.0–17.4
Green leaf blade	4	4	4	73.5±1.8	70.6–76.3	38.9±1.3	36.2–41.0	3.6±0.4	2.9–4.1	6.3±1.4	4.4–8.5
Green leaf sheath	4	4	4	75.1±3.4	69.7–80.3	41.9±2.3	38.1–46.7	5.8±1.2	4.1–7.8	8.0±2.5	4.7–13.7
Bagasse	4	4	4	84.6±3.1	78.1–88.8	50.7±3.3	46.5–55.1	5.5±0.8	4.6–7.0	1.8±0.5	1.1–2.8
Top of the stalk	4	4	4	64.3±4.6	55.9–69.1	36.2±2.5	31.7–39.3	3.0±1.0	1.2–4.1	9.2±1.4	5.9–10.7
Validation set											
Trash	2	2	2	74.0±1.8	70.7–76.0	41.9±1.9	39.7–45.1	4.4±0.4	4.0–4.9	9.9±3.0	4.0–12.0
Green leaf blade	2	1	1	72.9±2.3	70.6–75.8	38.3±1.5	36.2–39.7	3.4±0.6	2.9–4.1	6.0±1.1	5.0–7.4
Green leaf sheath	2	1	—	73.6±3.4	69.7–75.7	41.7±2.8	39.3–44.7	5.7±1.0	4.8–6.7	8.6±1.2	7.2–9.4
Bagasse	2	—	—	87.6±1.8	86.3–88.8	53.9±1.0	53.2–54.7	6.8±0.3	6.6–7.0	1.8±0.6	1.4–2.2
Tope of the stalk	1	—	4	65.1±4.6	58.6–69.1	37.0±2.2	34.1–39.3	3.1±1.0	1.5–4.0	9.3±1.5	6.8–10.5

NDF_{om} = neutral detergent fiber, ADF_{om} = acid detergent fibre, Lig = lignin-like, Ash = total ash, *SD* = standard deviation

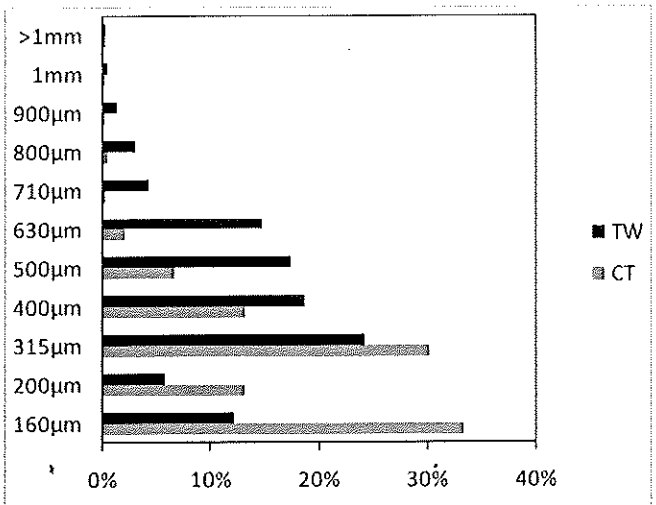

Figure 1. Particle size distribution of sugarcane samples ground by using different grinding methods. TW = Thomas Wiley knife crusher through a 2 mm screen and CT = Cyclone crusher Cyclotec through a 1 mm screen.

Principal components (PC) were computed on TW and CT sets to define the optical properties of the samples, and the Mahalanobis distance of each spectrum was calculated with respect to the average spectra. The aim of this analysis was to evaluate population structuring and, in the present study, it was used as a preliminary way to see whether the grinding methods had a segregating effect on the spectra of the sugarcane samples.

Optimisation of calibration method

Data pretreatment

At the first step of optimisation, six pretreatments were tested: first derivative (D), second derivative (D^2), multiplicative scatter correction (MSC), standard normal variate and detrend (SNVD)[7], standard normal variate and detrend successively followed by first derivative (SNVD-D) or second derivative (SNVD-D^2). Spectral pretreatment and multiple linear regression (MLR) were performed with the WIN-ISI 3 software (Infrasoft International, Port Matilda, PA, USA) package. Seven MLR calibrations with the same number of terms were developed on CS_{CT} without pretreatment and treated with the above six pretreatments. The standard error of calibration (*SEC*), the coefficient of determination of calibration (R^2_C), the predicted error sum of squares (*PRESS*), the coefficient of determination of cross-validation (R^2_{CV}) and the ratio performance to deviation for cross-validation ($RPD_{CV} = [1/\sqrt{(1-R^2_{CV})}]$, as described by Dardenne,[29] were used to evaluate the performance of each MLR calibration. The method described by Fearn[30] was used to evaluate which pretreatment method(s) led to the significantly lowest overall *SEC*.

Calibration sets

The second step of optimisation is in the comparison of statistical values from calibration models ($Model_{CT}$, $Model_{TW}$ and $Model_{MIX}$) developed on each calibration set (CS_{CT}, CS_{TW} and CS_{MIX}, respectively), pretreated with the most effective pretreatment determined at the first step of optimisation. Relationships were developed by regressing the observed chemical values against the three calibration sets using the MLR regression technique. The accuracy of each calibration with the same number of terms was assessed using the statistics *SEC*, R^2_C, *PRESS*, R^2_{CV} and RPD_{CV}.

Validation of calibration method

The last step utilised an independent validation of the three calibration models ($Model_{CT}$, $Model_{TW}$ and $Model_{MIX}$) developed at the second step on three validation sets (VS_{CT}, VS_{TW} and VS_{MIX}). Each validation set contained 20 sugarcane samples not used for calibration, chosen at random and collected on the same experimental sites at a different harvest time

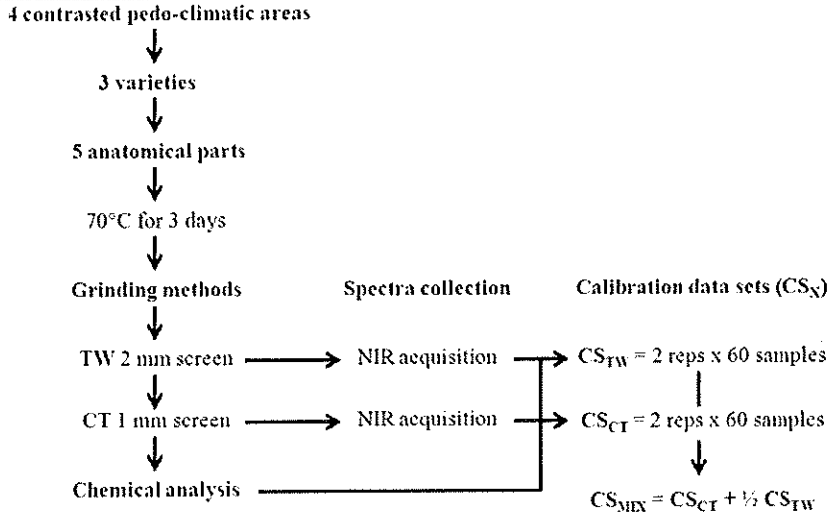

Figure 2. Flow chart of experimental procedures used for each sample, spectra collection and calibration sets (CSx).

(Table 1). Validation set VS_{MIX} was obtained by mixing ten samples randomly chosen in VS_{CT} and VS_{TW}, respectively. These three calibration models were evaluated, as recommended by Dardenne,[29] by using prediction parameters such as the standard error of prediction (*SEP*), coefficient of determination of prediction (r^2_P), bias (Bias) and average global *H* (Av. *GH*).

Results and discussion

Preparing samples, in particular using fine grinding, is costly and time-consuming. We propose an optimum calibration method by using a faster treatment of samples used for NIR analysis. We first present the results of a small economic study to compare the two grinding methods. At the experimental station, sugarcane samples are routinely ground with the coarse-grinding TW because of its convenience and rapidity to process a large number of samples (the treatment of each sample takes about five minutes). Even though the CT is very efficient in providing homogeneous samples, it is still necessary to pre-grind the samples with the TW, which is time-consuming (using the fine-grinding CT, the treatment of each sample takes about 10 min). A high number of samples underlined the need for NIR calibration optimisation to predict lignocellulosic compounds in sugarcane samples prepared using a faster method resulting in a wider and coarser particle size.

Chemical composition of sugarcane samples

Table 1 shows the characteristics of the 60 samples from different anatomical fractions of the sugarcane plant. Samples originated from (i) three contrasting varieties, (ii) planted in four pedoclimatic environments and (iii) harvested at different times. These factors contributed to a wide range of chemical compositions in the resultant database.

Chemical composition analysis of sugarcane has particularly focused on sugarcane bagasse[31–37] and few studies[38] have dealt with other anatomical parts. Results presented by others[31–37] are closer to those presented in Table 1. An average of 34% and 45% were measured in sugarcane bagasse against ranges from 23% to 58% and from 32% to 58% for hemicellulose (hemicellulose = NDF_{om} – ADF) and cellulose (cellulose = ADF_{om} – ADL), respectively. The lignin content (Lig) of our sugarcane bagasse samples was lower (4.6% to 7.0%, average 5.5%) than values reported in the literature[31–37] (18% to 28%). Table 1 shows wide variations in chemical composition between different anatomical parts. Sharma *et al.*[38] also reported differences in the composition of stalk and leaf sheath. However, other authors[39] have found lower NDF, ADF and Lig content than ours in the leaf sheath. Differences in the chemical composition of the same anatomical parts may be caused by (i) natural variability between varieties, ages or pedoclimatic growth conditions, (ii) variations in the characteristics of the sample (for example, between industrially processed bagasse and sugarcane stalk prepared in the laboratory) and (iii) problems due to the method used (different protocols or difficulties in separating lignocellulosic compounds without spoiling them).

Spectral variability

Figures 3(a) and (b) show the raw NIR spectra of TW and CT sets, respectively. Variations in spectral intensity are clearly apparent in the two figures. To make these variations more readable, the raw NIR spectra of TW and CT sets were averaged (Figure 3(c)). TW set spectra presented more variations in spectral intensity than CT set spectra, as reported by Bertrand *et al.*[2] However Figure 3(c) shows that variation due to grinding method is low with regard to variation due to different anatomical parts and their agro-climatic characteristics. Figure 4 shows the first two PCs of TW and CT sets as a function of the first PC score. The samples are clearly separated according to the grinding method used. For example, samples of green leaf blade (Gb) from TW and CT sets are circled by a dotted line (Gb TW) and a solid line (Gb CT), respectively, in so doing demonstrating the effect of the grinding method on the spectral features. However, samples tended to cluster according to the anatomical part concerned. This indicates that the chemical information contained in the raw NIR spectra can be masked or modified by physical information, however anatomical and growth characteristics of samples are the main sources of variation.

Optimisation of calibration method

Data pretreatment

Table 2 shows statistical values from the calibration and the full (leave-one-out) cross-validation parameters of a MLR model developed on the raw and pretreated CS_{CT}. To compare these statistical values, we performed a paired *t*-test[30] which evaluates the significant differences among the six pretreatments we investigated. Significant differences were found ($P < 0.05$) between models developed on the raw and pretreated spectra for each constituent, ADF_{om} excepted. These results suggest that performance ranking of each pretreatment depends on the constituent being considered. However, in the comparison between pretreatment methods, the SNVD-D^2 pretreatment produced the best calibration and cross-validation parameters for at least three out of the four variables to be predicted. These results are in agreement with results in both older and recent publications[7,12,20,22,24,39] showing that the pretreatments (SNV, Detrend and D^2, or their combination) performed well in calibration. So, a combination of these three pretreatments, resulting in a well known pretreatment called SNVD-D^2, was shown to be the most powerful pretreatment among those tested in the present study. In parallel, we obtained higher R^2_{CV} (0.95 and 0.98 for Lig and Ash, respectively) than others[24] for the same variables measured on comparable plant matrices (corn and switchgrass): values ranging from 0.44 to 0.90 and from 0.62 to 0.92 for Lig and Ash, respectively. Even though Liu *et al.*[24] reported data with a wide range of Lig and Ash content, they had access to fewer samples with a smaller range than us, which was probably why they found lower R^2_{CV}. Besides,

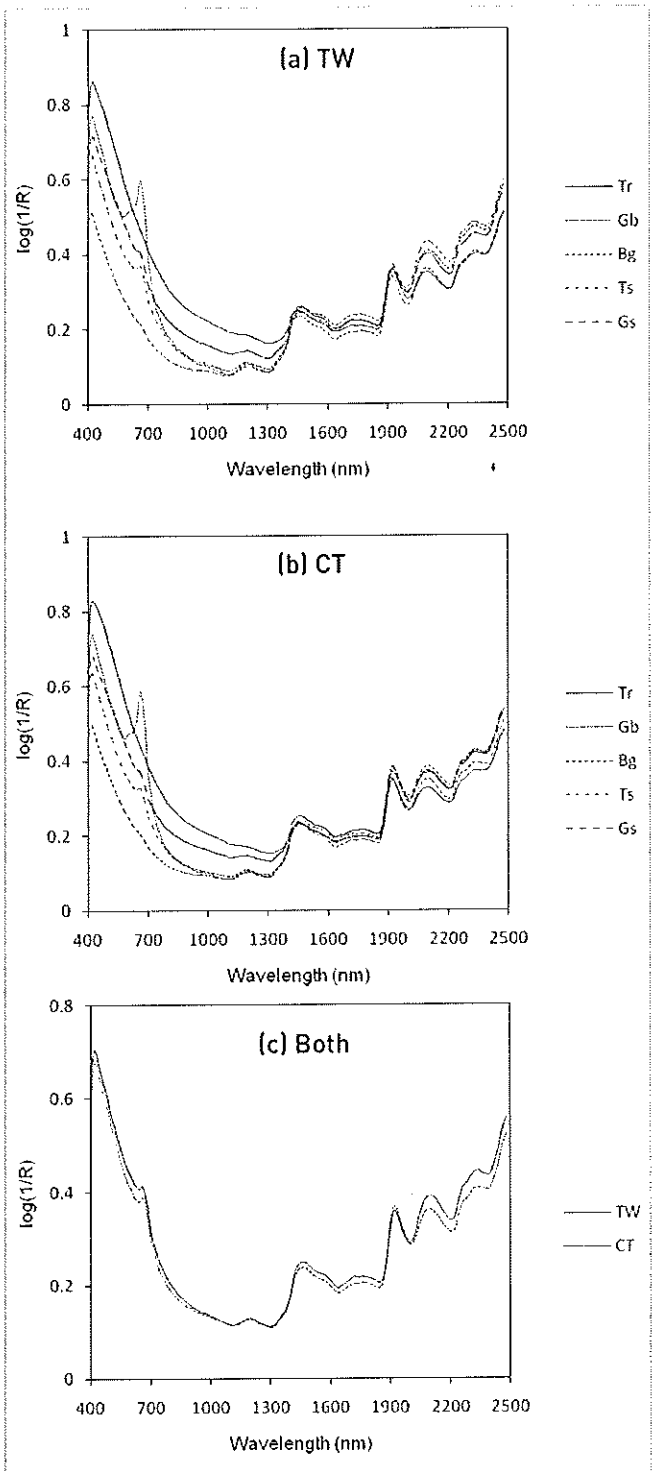

Figure 3. Raw log 1/reflectance spectra of sugarcane samples ground using two different methods: (a) TW = samples ground with a Thomas Wiley through a 2 mm screen, (b) CT = samples ground with a Cyclotec through a 1 mm screen and (c) 1/reflectance mean spectra for both methods.

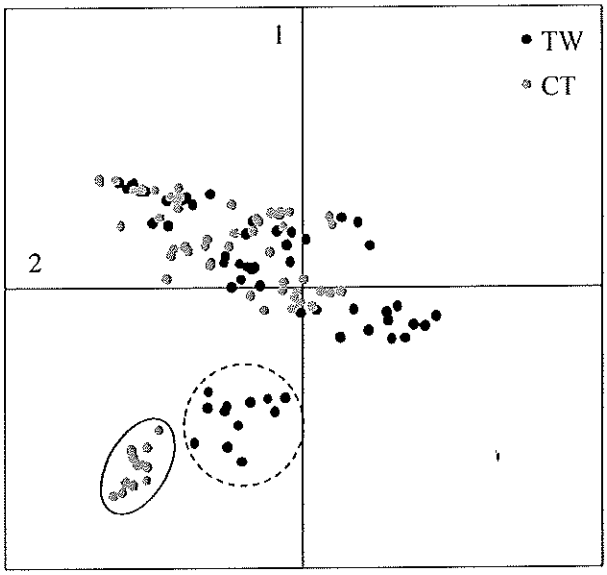

Figure 4. PCA of raw spectra of sugarcane samples subjected to different grinding methods: ● TW (Thomas Wiley through a 2 mm screen) and ○ CT (Cyclotec through a 1 mm screen). Samples are plotted according to the scores of PC1 (vertical) and PC 2 (horizontal). Samples of sugarcane green leaf blade (Gb) ground with TW and CT were circled by a dashed line (Gb, TW) and a solid line (Gb, CT), respectively, to illustrate spectral dissimilarity between treatments

contrary to Liu *et al.*,[24] calibrations developed on raw spectra were not always significantly different from those calibrations developed on treated spectra.

Consequently, as with other authors working on different matrices,[20,22] we believe that the SNVD-D² pretreatment has the greatest potential for removing physical interferences in the spectra acquired on samples in powder form.

Calibration sets

Restrained by the reference method,[26] the NIR model had to be developed using chemical data acquired on finely ground samples (CT) in order to make predictions using coarsely ground samples (TW). Table 3 shows good calibration statistics and the values obtained were better and comparable, in spite of fewer numbers of samples and narrow range, with those reported by others.[39] As expected, $Model_{CT}$ always showed higher R^2_C, R^2_{CV} and RPD_{CV} and lower *SEC* and *PRESS* than $Model_{TW}$ (excepted for Lig). This seems logical, given that the correlation level between spectral and chemical data was optimised when chemical and spectral data were acquired on samples with the same particle size. The converse is also true excepted for Lig. We also shows that $Model_{MIX}$ improved the calibration and cross-validation parameters for two out of four variables, for example RPD_{CV} rose from 4.1 to 4.7 and from 3.3 to 4.0 for NDF_{om} and ADF_{om}, respectively. However, this was not the case for Lig and Ash. These results confirmed that inclusion of samples with a wide variation in particle size in a calibration set improved the performance of data pretreatment, as demonstrated by Norris and Williams.[12] Nevertheless, our results show that the improvement in the calibation also depends on the constituent considered.

Table 2. Statistical values of NIR calibrations developed on calibration set CS_{CT} without pretreatment and treated with six pretreatments depending on the chemical composition of the sugarcane sample analysed.

Chemical composition % DM	NDF_{om}							ADF_{om}						
Mean	74.3							41.8						
Range	52.3–96.2							25.4–58.2						
SD	7.3							5.5						
Pretreatment	**Terms**	***SEC***	R^2_c	***PRESS***	R^2_{cv}	RPD_{cv}	**Rank**	**Terms**	***SEC***	R^2_c	***PRESS***	R^2_{cv}	RPD_{cv}	**Rank**
No pretreatment	6	1.7	0.95	1.8	0.94	4.0	bc	6	1.4	0.94	1.5	0.93	3.7	a
D	6	1.7	0.94	1.9	0.93	3.9	ab	6	1.4	0.93	1.5	0.92	3.6	a
D^2	6	1.6	0.95	1.7	0.95	4.3	bc	6	1.2	0.95	1.3	0.94	4.1	a
MSC	6	1.6	0.95	1.7	0.94	4.2	bc	6	1.2	0.95	1.4	0.94	4.0	a
SNVD	6	1.5	0.96	1.6	0.95	4.4	bc	6	1.3	0.94	1.4	0.94	4.0	a
SNVD+D	6	1.4	0.96	1.5	0.96	4.8	c	6	1.3	0.94	1.5	0.93	3.7	a
SNVD+D^2	6	1.5	0.96	1.7	0.95	4.4	bc	6	1.2	0.95	1.3	0.94	4.1	a
Chemical composition % DM	**Lig**							**Ash**						
Mean	4.5							6.9						
Range	0.4–8.6							0.0–17.8						
SD	1.4							3.6						
Pretreatment	**Terms**	***SEC***	R^2_c	***PRESS***	R^2_{cv}	RPD_{cv}	**Rank**	**Terms**	***SEC***	R^2_c	***PRESS***	R^2_{cv}	RPD_{cv}	**Rank**
D	6	0.6	0.83	0.6	0.80	2.2	a	6	0.7	0.96	0.7	0.96	4.9	b
D^2	6	0.4	0.93	0.4	0.91	3.4	c	6	0.6	0.97	0.6	0.97	5.5	bc
MSC	6	0.3	0.94	0.3	0.94	4.0	cd	6	0.6	0.98	0.6	0.97	5.9	bc
SNVD	6	0.3	0.95	0.3	0.94	4.0	cd	6	0.9	0.94	1.0	0.93	3.8	a
SNVD+D	6	0.4	0.92	0.4	0.91	3.3	b	6	0.8	0.95	0.9	0.94	4.1	a
SNVD+D^2	6	0.3	0.95	0.3	0.94	4.1	cd	6	0.5	0.98	0.6	0.98	6.4	c
D	6	0.3	0.96	0.3	0.95	4.5	d	6	0.5	0.98	0.5	0.98	7.0	c

NDF_{om} = neutral detergent fibre, ADF_{om} = acid detergent fibre, Lig = lignin-like, Ash = total ashes, *SD* = standard deviation, D = drivative, D^2 = second derivative, MSC = multitplicative scatter correction and SNVD = standard normal variate and detrend, Terms = nuumber of terms, *SEC* = standard errorof calibration, R^2_c = correlation coefficient for calibration, *PRESS* = predicted error sum of squares, R^2_{cv} = correlation coefficient of determination for cross-validation, RPD_{cv} = ratio performance to deviation for cross-validation, Rank = performance ranking of pretreatments

Table 3. Statistical values of three NIR calibration models ($Model_X$) developed according to a combination of three calibration sets (CS_{TW}, CS_{CT} and CS_{MIX}) treated with standard normal variate and detrend followed by second derivative (SNVD + D^2) and multiple linear regression technique

Model	Consistuent %DM	*N*	Outliers	Mean	Range	*SD*	Terms	*SEC*	R^2_c	*PRESS*	R^2_{CV}	RPD_{CV}
$Model_{TW}$	NDF_{OM}	60	0	74.3	52.3–96.2	7.3	6	1.7	0.95	1.8	0.94	4.1
	ADF_{OM}	60	0	41.8	25.4–58.2	5.5	6	1.5	0.93	1.6	0.91	3.3
	Lig	60	0	4.5	0.4–8.6	1.4	6	0.3	0.96	0.3	0.96	4.9
	Ash	60	0	6.9	0–17.8	3.6	6	0.5	0.98	0.5	0.98	6.6
$Model_{CT}$	NDF_{OM}	60	0	74.3	52.3–96.2	7.3	6	1.5	0.96	1.7	0.95	4.4
	ADF_{OM}	60	0	41.8	25.4–58.2	5.5	6	1.2	0.95	1.3	0.94	4.1
	Lig	60	0	4.5	0.4–8.6	1.4	6	0.3	0.96	0.3	0.95	4.5
	Ash	60	0	6.9	0–17.8	3.6	6	0.5	0.98	0.5	0.98	7.0
$Model_{MIX}$	NDF_{OM}	90	0	73.7	51.0–96.4	7.6	6	1.6	0.96	1.6	0.95	4.7
	ADF_{OM}	90	0	41.5	24.8–58.2	5.6	6	1.3	0.94	1.4	0.94	4.0
	Lig	90	0	4.4	0.1–8.7	1.4	6	0.3	0.96	0.3	0.96	4.7
	Ash	90	0	7.1	0–17.7	3.5	6	0.7	0.96	0.7	0.96	4.9

NDF_{om} = neutral detergent fibre, ADF_{om} = acid detergen fibre, Lig = lignin-like, Ash = total ash, *N* = number of samples in calibration set, *SD* = standard deviation, Terms = number of terms, *SEC* = standard error of calibration, R^2_c = correlation coefficient of determination for calibration, *PRESS* = predicted error sum of squares, R^2_{CV} = correlation coefficient for cross-validation, RPD_{CV} = ration performance to deviation for cross-validation

Validation of calibration method

Globally, Table 4 shows better statistical values for predictions when we considered each model independently vs its associated validation set (see grey shaded). As expected, $Model_{MIX}$ performed well across both VS_{TW} and VS_{CT} samples relative to the others models. Moreover, inclusion of maximum variations in particle size was associated with decreases in biases and better representation of the variation in particle size. In fact, we show lower bias for $Model_{MIX}$ for three out of four variables than $Model_{TW}$ and much closer Av. *GH* between both. On the other hand, $Model_{CT}$ failed to perform well on VS_{TW} samples, especially in respect of ADF_{om}, Lig and Ash fractions which show elevated biases and Av. *GH* statistics. This suggests that samples were not represented in the calibration set[29] and were approaching spectral outlier status brought about by differences due to the grinder used. $Model_{TW}$ presents equal and higher *SEP*s than $Model_{MIX}$, for Lig, NDF_{om} and ADF_{om}, respectively. The r^2_P of $Model_{TW}$ values were smaller than those of $Model_{MIX}$ for two out of four variables and equal for Lig. The r^2_P of both models were better for NDF_{om} and Ash but smaller for ADF_{om} than those observed by Cozzolino *et al.*[39] on maize, although our *SEP*s were lower than those except for Ash. The results show that including samples with variations in particle size in a calibration set can improve the prediction of lignocellulosic compounds of sugarcane samples with coarser and wider particle size. It confirms the result of Norris and Williams.[12] Even though we tested only two grinders, our sample sets had a wider range of particle size (from 160 µm to up to 1 mm) than they used. (from 150 µm to 335 µm) as well as different variables to predict in different matrices. Indeed, we confirmed that improvement in prediction accuracy depends on the inclusion of samples with variations in particle size in calibration sets. We also demonstrated that improvement depends on the constituent we considered which is especially true for Lig. We affirmed that $Model_{MIX}$ proved to be the best calibration model to predict lignocellulosic compounds of sugarcane samples with the coarse particle size (VS_{TW}) we investigated.

Conclusions

The SNVD-D^2 pretreatment enabled a significant reduction in variations in spectral intensity observed on NIR spectra of sugarcane samples caused by the grinding method. We showed that the calibration method developed on a mixed calibration set improved the prediction of lignocellulosic compounds in coarse sugarcane samples with a wide range in particle size. We also showed that improvement in prediction depends of the variable we considered. In conclusion, we confirmed that it is possible to predict—without loss of accuracy—lignocellulosic compounds in samples of all parts of the sugarcane prepared using a faster method resulting in coarse particle size. Results from this study should give analysts using NIR spectroscopy confidence to avoid time consuming fine grinding of sugarcane samples. Moreover, analsis of samples with a coarse particle

Table 4. Statistical values of NIR validations according to three calibration models ($Model_{TW}$, $Model_{CT}$ and $Model_{MIX}$), developed on three calibration sets (CS_{TW}, CSC_T and CS_{MIX}, respectively) treated with SNVD + D^2, tested on three validation sets (VS_{TW}, VS_{CT} and VS_{MIX}).

	VS_X	VS_{TW}				VS_{CT}				VS_{MIX}			
$Model_X$	Chemical	NDF_{OM}	ADF_{om}	Lig	Ash	NDF_{om}	ADF_{OM}	Lig	Ash	NDF_{OM}	ADF_{OM}	Lig	Ash
$Model_{TW}$	*N*	20	20	20	20	20	20	20	20	20	20	20	20
	Outliers	0	0	0	0	0	0	0	0	0	0	0	0
	Mean	72.9	4.11	4.3	8.0	72.9	41.1	4.3	8.0	72.8	40.9	4.4	8.6
	Range	58.6–88.8	34.1–54.7	1.5–7.0	1.41–12.0	58.6–88.8	34.1–54.7	1.5–7.0	1.4–12.0	58.6–86.4	34.1–53.2	1.5–7.0	1.4–12.0
	SD	6.8	5.2	1.4	3.1	6.8	5.2	1.4	3.1	7.0	5.3	1.7	3.2
	SEL	0.4	0.3	0.1	0.2	0.4	0.3	0.1	0.2	0.4	0.3	0.2	0.3
	SEP	1.4	1.6	0.2	0.4	2.9	2.5	0.5	0.9	1.7	1.6	0.4	0.7
	r^2_P	0.96	0.91	0.98	0.99	0.82	0.79	0.87	0.92	0.94	0.93	0.95	0.96
	Bias	0.4	0.7	0.1	−0.1	0.9	−0.2	0.1	0.4	0.7	−0.1	0.1	0.1
	Av. *GH*	1.0	1.0	1.0	1.1	1.6	1.4	1.8	1.9	1.2	1.0	1.2	1.2
$Model_{CT}$	*N*	20	20	20	20	20	20	20	20	20	20	20	20
	Outliers	0	0	0	0	0	0	0	0	0	0	0	0
	Mean	72.9	41.1	4.3	8.0	72.9	41.1	4.3	8.0	72.8	40.9	4.4	8.6
	Range	58.6–88.8	34.1–54.7	1.5–7.0	1.4–12.0	58.6–88.8	34.1–54.7	1.5–7.0	1.4–12.0	58.6–86.4	34.1–53.2	1.5–7.0	1.4–12.0
	SD	6.8	5.2	1.4	3.1	6.8	5.2	1.4	3.1	7.0	5.3	1.7	3.2
	SEL	0.4	0.3	0.1	0.2	0.4	0.3	0.1	0.2	0.4	0.3	0.2	0.3
	SEP	2.1	1.8	0.7	2.7	1.7	1.1	0.3	0.4	1.8	1.8	0.7	2.8
	r^2_P	0.92	0.88	0.76	0.62	0.95	0.96	0.97	0.99	0.94	0.89	0.82	0.67
	Bias	0.1	−2.1	−0.9	3.9	0.0	0.1	0.0	0.0	−0.1	−1.0	−0.5	1.9
	Av. *GH*	1.5	1.9	3.2	2.0	1.0	0.9	0.9	1.0	1.1	1.4	1.9	1.4
$Model_{MIX}$	*N*	20	20	20	20	20	20	20	20	20	20	20	20
	Outliers	0	0	0	0	0	0	0	0	0	0	0	0
	Mean	72.9	41.1	4.3	8.0	72.9	41.1	4.3	8.0	72.8	40.9	4.4	8.6
	Range	58.6–88.8	34.1–54.7	1.5–7.0	1.4–12.0	58.6–88.8	34.1–54.7	1.5–7.0	1.4–12.0–	58.6–86.4	34.1–53.2	1.5–7.0	1.4–12.
	SD	6.8	5.2	1.4	3.1	6.8	5.2	1.4	3.1	7.0	5.3	1.7	3.2
	SEL	0.4	0.2	0.1	0.2	0.4	0.3	0.1	0.2	0.4	0.3	0.2	0.3
	SEP	**1.3**	**1.1**	**0.2**	**0.6**	1.9	0.9	0.2	0.7	1.6	0.8	0.3	0.7
	r^2_P	**0.97**	**0.96**	**0.98**	**097**	0.94	0.97	0.97	0.95	0.98	0.98	0.98	0.96
	Bias	−0.2	0.1	0.0	0.2	0.0	0.3	0.0	−0.1	−0.2	0.2	0.0	0.0
	Av. *GH*	1.2	1.2	1.2	1.2	0.8	1.0	1.0	0.9	0.9	0.9	1.0	1.1

NDF_{om} = neutral detergent fibre, ADF_{om} = acid detergent fibre, Lig = lignin-like, Ash = total ash, *N* = number of samples in validation set, *SD* = standard deviation, Terms = number of terms, *SEL* = standard error of laboratory, *SEP* = standard error of prediction, r^2_P = correlation coefficient of determination for prediction, Av *GH* = average global H (Mahalanobis distance of the PLS score space from the average spectrum).

size can be an alternative technique in the characterisation of all parts of the sugarcane plant in ecophysiology research and growth model development.[25] With more representative calibration sets, including more samples with wider characteristics and new variables, especially hemicellulose and cellulose, robust and accurate NIR models can be developed to predict lignocellulosic compounds of sugarcane for biomass quantification.

Acknowledgements

Special thanks to the technical staff of "eRcane" Experimental Station for the treatment of sugarcane stalk samples. The authors also thank Dr F. Chiroleu and Dr P. Letourmy for statistical advice.

References

1. M.C. Pasikatan, J.L. Steele, C.K. Spillman and E. Haque, "Near infrared reflectance spectroscopy for online particle size analysis of powders and ground materials", *J. Near Infrared Spectrosc.* **9(3),** 153–164. (2001). doi: 10.1255/jnirs.303
2. D. Bertrand et E. Dufour. *La Spectroscopie Proche Infrarouge Et Ses Applications Analytiques. Collection Sciences Et Techniques Agroalimentaires*, Ed by J.-L. Multon. Lavoisier, Paris, France. (2000).
3. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, F. Gonzales, and R. Pous, *Near Infrared Spectroscopy: Bridging The Gap Between Data Analysis And NIR Applications*, Ed by. K.I. Hildrum, T. Isaksson, T. Naes and A. Tanberg. Ellis Horwood, Chichester, UK (1992).
4. P.C. Williams and B.N. Thompson, "Influence of whole meal granularity on analysis of HRS wheat for protein and moisture by near infrared reflectance spectroscopy (NIRS)", *Cereal Chem.* **55(6),** 1014–1037. (1978).
5. B.G. Osborne, "Near-infrared spectroscopy in food analysis", in *Encyclopedia of Analytical Chemistry*, Ed by Robert A. Meyers. John Wiley & Sons Ltd, Chichester, UK (1993).
6. M.S. Dhanoa, S.J. Lister, R. Sanderson and R.J. Barnes, "The link between multiplicative scatter correction (MSC) and standard normal variate (SNV) transformations of NIR spectra", *J. Near Infrared Spectrosc.* **2,** 43–47 (1994). doi: 10.1255/jnirs.30
7. R.J. Barnes, M.S. Dhanoa and S.J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra", *Appl. Spectrosc.* **43(5),** 772–777 (1989). doi: 10.1366/0003702894202201
8. J.L. Ilari, H. Martens and T. Isaksson, "Determination of particle size in powders by scatter correction in diffuse near-infrared reflectance", *Appl. Spectrosc.* **42(5),** 722–728 (1988). doi: 10.1366/0003702884429058
9. A.J. O'Neil, R.D. Jee and A.C. Moffat, "The application of multiple linear regression to the measurement of the median particle size of drugs and pharmaceutical excipients by near-infrared spectroscopy", *Analyst* **123,** 2297–2302 (1998). doi: 10.1039/a806001k
10. A.J. O'Neil, R.D. Jee and A.C. Moffat, "Measurement of the cumulative particle size distribution of microcrystalline cellulose using near infrared spectroscopy", *Analyst* **124,** 33–36 (1998). doi: 10.1039/a807134i
11. M. Otsuka, "Comparative particle size determination of phenacetin bulk powder by using Kebulka-Munk theory and principal component regression analysis based on near-infrared spectroscopy", *Powder Technol.* **141,** 244–250 (2004). doi: 10.1016/j.powtec.2004.01.025
12. K.H. Norris and P.C. Williams, Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat. I. Influence of particle size", *Cereal Chem.* **61(2),** 158–165 (1984).
13. C.R. Bull, "Compensation for particle size effects in near infrared reflectance", *Analyst* **116,** 781–786 (1991). doi: 10.1039/an9911600781
14. A.A. Christy, Y.-Z. Liang, C. Hui and O.M. Kvalheim, "Effect of particle size on diffuse reflectance infrared spectra of polystyrene spheres", *Vib. Spectrosc.* **5,** 233–244 (1993). doi: 10.1016/0924-2031(93)87073-3
15. M.-F. Devaux, N. Nathier-Dufour, P. Robert and D. Bertrand, "Effects of particle size on the near-infrared reflectance spectra of wheat and rape seed meal mixtures," *Appl. Spectrosc.* **49(1),** 84–91 (1995). doi: 10.1366/0003702953963256
16. M. Papini, "Study of the radiative properties of powdered and fibrous polymers", *Vib. Spectrosc.* **11,** 61–68. (1995). doi: 10.1016/0924-2031(95)00058-5
17. D.J. Dahm, "Effect of particle size on absorbance", *NIR news* **16(3),** 5–6 (2005).
18. D.J. Dahm, E.W. Ciurczak and K.H. Norris, "Effect of particle size on absorbance (for real)", *NIR news.* **16(4),** 4–5 (2005).
19. D.K. Lovett, E.R. Deaville, D.I. Givens, M. Finlay and E. Owen, "Near infrared reflectance spectroscopy (NIRS) to predict biological parameters of maize silage: effects of particle comminution, oven drying temperature and the presence of residual moisture", *Anim. Feed Sci. Technol.* **120,** 323–332 (2005). doi: 10.1016/j.anifeedsci.2005.02.001
20. D.R. Ely, M. Thommes and M.T. Carvajal, "Analysis of the effects of particle size and densification on NIR spectra", *Colloids and Surfaces A: Physicochem. Eng. Aspects* **331,** 63–67 (2008). doi: 10.1016/j.colsurfa.2008.07.017
21. R.J. Barnes, M.S. Dhanoa and S.J. Lister, "Correction of the description of standard normal variate (SNV) and de-trend (DT) transformation in *Practical Spectroscopy With Applications In Food And Beverage Analysis*-2nd edition. *J. Near Infrared Spectrosc.* **1,** 185–186 (1993). doi: 10.1255/jnirs.21
22. T. Azzouz, A. Puigdoménech, M. Aragay and R. Tauler, "Comparison between different data pre-treatment

methods in the analysis of forage samples using near-infrared diffuse reflectance spectroscopy and partial least-squares multivariate calibration method", *Anal. Chim. Acta* **484,** 121–134 (2003). doi: 10.1016/S0003-2670(03)00308-8

23. Stephen R. Delwiche and James B. Reeves, III, "The effect of spectral pre-treatments on the partial least squares modelling of agricultural products", *J. Near Infrared Spectrosc.* **12(3),** 177–182. (2004). doi: 10.1255/jnirs.424

24. L. Liu, X.P. Ye, A.M. Saxton and A. Womac, "Pretreatment of near infrared spectral data in fast biomass analysis", *J. Near Infrared Spectrosc.* **18(3),** 317-331 (2010). doi: 10.1255/jnirs.898

25. J.F Martiné and P. Todoroff, "Le modèle de croissance Mosicas et sa plateforme de simulation Simulex: Etat des lieux et perspectives", 10e Congrès S.T.A.S.M. pp. 133–147 (2002).

26. P.J. van Soest and R.H. Wine, "Use of detergents in the analysis of feeds. IV. Determination of plant cell-wall constituents", *J. Assoc. Offic. Anal. Chem.* **50,** 50 (1967).

27. P.J. Van Soest, J.B. Robertson and B.A. Lewis, "Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition", *J. Dairy Sci.* **74,** 3583–3597 (1991). doi: 10.3168/jds.S0022-0302(91)78551-2

28. D.R. Mertens, "Gravimetric determination of amylase-treated neutral detergent fiber in feeds with refluxing in breakers or crucibles: collaborative study", *J. Assoc. Off. Anal. Chem. Int.* **85(6),** 1217–1240 (2002).

29. P. Dardenne, "Some considerations about NIR spectroscopy: Closing speech at NIR-2009", *NIR news.* **21(1),** 8–14 (2010).

30. T. Fearn. "Comparing standard deviations (continued)", *NIR News.* **20(7),** 24–25 (2009). doi: 10.1255/nirn.1153

31. J.E. Irvine and G.T.A. Benda, "Genetic potential and restraints in *Saccharum* as an energy source", Symposium proceedings: *Alternative uses of sugarcane for development*, Caribe Hilton Hotel, San Juan, Puerto Rico (1979).

32. I.M. Wood, "The utilization of field crops and crop residues for paper pulp production", Commonwealth Bureau of Pasture and Field Crops, *Field Crop Abstracts* **34(7),** 557–568. (1981).

33. L. Sene, A. Converti, M.G.A. Felipi and M. Zilli, "Sugarcane bagasse as alternative packing material for biofiltration of benzene polluted gazeous streams: A preliminary study", *Bioresource Technol.* **83,** 153–157 (2002). doi: 10.1016/S0960-8524(01)00192-4

34. A.R. Conçalves, P. Benra, S.M. Costa, D.S. Ruzene, R.Y. Moriva, S.M. Luz and L.P. Ferretti, "Integrated processes for use of pulps and lignins obtained from sugarcane bagasse and straw", *Appl. Biochem. Biotechnol.* **121–124,** 821-826 (2005).

35. M. Rinaudo and J.P. Joseleau, "La biomasse végétale peut-elle concurrencer le pétrole? Centre de Recherche Sur les Macromolécules Végétales, CNRS. 32 Diapo (2008).

36. S.K. Batra, "Other long vegetable fibers: abaca, banana, sisal, henequen, flax, hemp, sunn and coir, in *Handbook of Fiber Science and Technology*, Vol. IV, Ed by M. Lewin and E.M. Pearce. Marcel Dekker, New York, USA, pp. 727–807 (1983).

37. A. Pandey, C.R. Soccol, P. Nigam and V.T. Soccol, "Biotechnological potential of agro-industrial residues I: Sugarcane bagasse", *Bioresource Technol.* **74(1),** 69–80 (2000). doi: 10.1016/S0960-8524(99)00142-X

38. B.L. Sharma, S. Jaipal and O.P. Chaudhary, "Screening of sugarcane genotypes against stalk borer, *Chilo auricilius* Dudgeon", *Cooperative Sugar.* **39(4),** 27 (2000).

39. D. Cozzolino, A. Fassio and A. Giminez, "The use of near-infrared reflectance spectroscopy (NIRS) to predict the composition of whole maize plants", *J. Sci. Food Agric.* **81,** 142–146 (2000). doi: 10.1002/1097-0010(20010101)81:1<142::AID-JSFA790>3.0.CO;2-I