Near infrared spectroscopy determination of sucrose, glucose and fructose in sweet sorghum juice

Maria Lúcia F. Simeone a,⁎, Rafael A.C. Parrella a, Robert E. Schaffert a, Cynthia M.B. Damasceno a, Michelle C.B. Leal a, Celio Pasquinib

a Embrapa Milho e Sorgo, MG 424, Km 45, 35701-970 Sete Lagoas, MG, Brazil
b University of Campinas, Chemistry Institute, C. P. 6154, 13083-970 Campinas, SP, Brazil

A R T I C L E   I N F O

Article history:
Received 20 October 2016
Received in revised form 3 April 2017
Accepted 28 May 2017
Available online 29 May 2017

Keywords:
Biofuel
Multivariate calibration
Bioenergy crop
Partial least squares regression

A B S T R A C T

Sweet sorghum is a very robust crop which has the potential to be used in ethanol production due to its high fermentable sugar content present in its stem juice, very similar to sugarcane. Therefore, for breeding purposes it is relevant to analyze sugar composition in the juice to characterize sweet sorghum genotypes and their period of industrial utilization within different environments for maximum ethanol yield. In this work we developed a rapid, low cost and efficient method to determine the profile of sugars (sucrose, glucose and fructose) in sorghum juice by near infrared spectroscopy and partial least square regression, and validation of the method was performed according to the high-performance liquid chromatography method. Developed models provided root mean square error of prediction of 4, 1 and 0.6 mg·mL⁻¹ and ratio performance deviations of 8, 5 and 5% for sucrose, glucose and fructose, respectively. Relative standard deviations of three sweet sorghum juice samples were reported with content variation (low, medium and high) 0.2, 0.3, 0.8% for sucrose; 1, 2, 2% for glucose; and 1, 2, 3% for fructose. Sugar profile is an asset for crop breeders to take decisions for the development of more productive cultivars and higher sugar content.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Sweet sorghum is one of the most promising alternative crops to sugarcane for ethanol production due to the presence of sweet juice in its stem [1]. Sugar content in sweet sorghum juice varies between 14 and 23% Brix and may be extracted by protocols similar to those used for sugarcane [2]. The juice from the fresh stem contains sucrose, glucose and fructose, with sucrose being the main sugar [3,4].

One of the measures undertaken by the sugar industry to assess sweet sorghum quality is the determination of the contents of soluble solids (Brix) of the juice extracted. However, the Brix is an indirect measure that relates the soluble solids dissolved in water based on refractive index changes. It is a measure widely used in the technological qualification of sugarcane juice [5], fruit juice [6] without specifying the sugar present. Brix in sweet sorghum samples has been strongly correlated with sucrose content, albeit not correlated with glucose and fructose [7].

Since the sugar extracted from sweet sorghum is a function of biomass yield, fiber content and juice quality, it is important to know the composition of the sugars in sorghum juice to better qualify the sweet sorghum genotypes and their period of industrial utilization (PIU) in different environments to provide maximum yield of ethanol during the fermentation process [2]. PIU should be the longest possible, with a minimum threshold of 30 days. In fact, PIU comprises the period in which the cultivar may remain in the field maintaining productivity and quality at optimal levels, according to the minimum standards established to ensure the viability of the crop until it is harvested and processed by the ethanol industry.

Chromatographic techniques, such as high performance liquid chromatography (HPLC) [8], ion chromatography (IC) [9], gas chromatography (GC) [10] or enzymatic methods [11], are commonly used to determine the chemical composition of sugars in sorghum juice.

However, all these techniques, coupled to several chemicals and inputs needed for sample preparation allow only a few analyses per day. The Embrapa Sorghum Breeding Program requires a great number of sugar content analyses of sweet sorghum juice during the harvest period. The method we established in this work allowed a faster and low-cost alternative to the HPLC method to detect hybrids with high sugar yield potential during their PIU. The method employs near infrared spectroscopy (NIR) associated to the development of multivariate chemometric regression models. PLS regression is a multivariate method and uses information of the NIR spectrum to establish the calibration equation. NIR
region contains information on the relative proportions of C—H, N—H and O—H bands which are the primary structural components or organic molecules [12].

This approach has been widely used in numerous agricultural and food products [13] and offers decisive advantages over traditional methods, such as little sample handling, no chemicals, high precision and accuracy, inexpensiveness and faster results [12]. The evaluation of sugar quality by near infrared spectroscopy has been reported in the literature for fruit juice [14], sugar beet [15], sugar-cane [16] and sweet sorghum in dry samples [17,18]. Chen et al. [17] extracted sucrose and glucose from dry sorghum stalks using distilled water and autoclave at 121 °C for 15 min. Mid-infrared spectroscopy was used to predict sucrose, glucose and fructose contents in juice samples of sweet sorghum [4].

This work aimed at developing a multivariate calibration-based method using near infrared transfectance spectroscopy as a source of analytical information to determine sucrose, glucose and fructose contents in sweet sorghum juice with the minimal pretreatment of samples for high-throughput screening phenotyping.

2. Materials and methods

2.1. Preparation of samples

The experiment was conducted in the field experimental area of Embrapa Maize and Sorghum, in Sete Lagoas (19°28′S, 44°15′08″W), MG, Brazil, using cultivars of Embrapa’s sweet sorghum breeding program.

One hundred sixty juice samples, from eight genotypes of sweet sorghum (BRS 508, BRS 509, BRS 511, CMSXS643, CMSXS646, CMSXS647, CV 198, CV 568 with similar flowering patterns were harvested, at different stages of maturation, 72 days after sowing with an interval of seven days approximately. The samples were collected during 2015 and 2016.

Normal cultural practices were maintained to conduct the experiment, following May et al. [19].

2.2. Sugar analysis

Stalk panicles were removed and eight stalks were crushed in a forage chopper machine (Irbi, Araçatuba SP Brazil). Further, 500 g of the material were taken to the hydraulic press (Hidraseme, Ribeirão Preto SP Brazil) for 1 min with minimal constant pressure of 250 kgf·cm$^{-2}$. An 80 mL aliquot of juice extracted from each sample was stored in a polyethylene vial and frozen at −4 °C for later analysis, totaling 160 samples. Sucrose, glucose and fructose contents were analyzed by HPLC as follows: sorghum juice samples were thawed at room temperature and 3 mL of each sample were diluted 15 times with deionized water. The samples were then shaken at 45 rpm for 15 min and centrifuged at 3000 rpm for 15 min. Samples were filtered through a C18 cartridge, previously conditioned with 2 mL acetonitrile and 2 mL deionized water. After this process, 2 mL of the solution were filtered with 0.45 μm membrane filters (PTFE) and analyzed by HPLC (2695 Alliance Waters, Milford, MA, USA) using a Phenomenex column (RCM-Ca). The mobile phase used was ultrapure water flux 0.6 mL min$^{-1}$, column temperature 65 °C. The detector was the Refractive Index (Milford MA, USA) working at 40 °C. Analytical curves were produced by using sucrose, d-glucose and d-fructose as standards (Sigma-Aldrich) with 99.5% purity, respectively. Sucrose, glucose and fructose in the samples were detected by comparison to standard retention time. Three calibration curves ($R^2$ ≥ 0.999) were established for sucrose, glucose, fructose.

### Table 1
Sucrose, glucose and fructose contents as determined by HPLC from 160 samples of sweet sorghum juice.

<table>
<thead>
<tr>
<th>Component</th>
<th>Sucrose</th>
<th>Glucose</th>
<th>Fructose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>26.50</td>
<td>6.60</td>
<td>4.21</td>
</tr>
<tr>
<td>Maximum</td>
<td>169.52</td>
<td>36.16</td>
<td>17.5</td>
</tr>
<tr>
<td>Mean</td>
<td>89.40</td>
<td>17.58</td>
<td>9.97</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Units: mg mL$^{-1}$.
and fructose, respectively, from determinations at six different sugar concentrations.

2.3. Near infrared spectra data calibration and validation

Juice sweet sorghum samples (50 mL) were filtered in cotton and placed on a petri dish (100 mm in diameter) with a transfection accessory (total nominal optical path of 1.5 mm) to collect NIR spectra with NIRFlex N-500 FT-NIR spectrometer (Flawil, Switzerland). The spectrometer was controlled and data were retrieved by NIRWare Operator software and handled with Unscrambler X® (version 10.3, CAMO Software Inc., Woodbridge NJ USA) software. The spectra were recorded in triplicate from 10,000 to 4000 cm$^{-1}$ with 4 cm$^{-1}$ steps, averaging 32 scans, at 25 ± 2 °C. HPLC analyses were performed after NIR measurement.

Prior to calibration, several preprocessing techniques, standard normal variate (SNV) and first-derivative Savitzky-Golay (SG-1), with 9 points on the right and on the left, were applied to the spectra to obtain the best calibration equation. Two sample sets were prepared for calibration and external validation applying the Kennard-Stone algorithm [21] to the values of the PLS scores of the samples.

The partial least square (PLS) method was used to provide a prediction eq. [20]. Model performance was assessed by the coefficient determination ($R^2$) of calibration and validation, root mean square error of calibration (RMSEC), (RMSECV, a full cross-validation) and prediction (RMSEP, for the external validation set). A full cross-validation following the random method was performed to determine the optimum number of factors for the model and to detect any outliers. Accuracy of the generated PLS models was attested by trueness and precision studies. Two other parameters, namely, ratio performance deviation (RPD) and range error ratio (RER), were used to evaluate the model’s prediction capacity [22].

3. Results and discussion

Soluble sugars are major components of sweet sorghum juice, with a wide range of sucrose, glucose and fructose concentrations [23].

Current study characterized 160 samples of sweet sorghum juice by HPLC analysis during maturation curve period of sweet sorghum development. We observed that sugar profiles changed according to sorghum’s developmental stage and the genotype analyzed.

The overall average sugar content in sweet sorghum juice (Table 1) was 89.40 mg mL$^{-1}$ sucrose, 17.58 mg mL$^{-1}$ glucose and 9.97 mg mL$^{-1}$ fructose. Juice from sweet sorghum genotypes exhibited total fermentable sugars ranging between 105.43 and 204.99 mg mL$^{-1}$ and averaging 171.92 mg mL$^{-1}$.

The raw spectra set in Fig. 1 show baseline offsets due to light scattering or refractive index variation due to concentration variation. All NIR spectra showed that vibration bands from O—H and C—H groups were correlated with sugar components. While the structures of sugars are similar and they exhibit similar NIR absorption peaks, they may be probably differentiated by their absorption magnitude due to the different numbers of O—H groups, and slight changes in

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of statistical indicators for calibration and validation of sucrose, glucose and fructose content (mg mL$^{-1}$) in sweet sorghum juice determinate by the optimized NIR based PLS models.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component</th>
<th>Sucrose</th>
<th>Glucose</th>
<th>Fructose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of samples</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>LV$^a$</td>
<td>5</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>RMSEC$^b$</td>
<td>0.8</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>R$^2c$</td>
<td>0.99</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>RPD$^d$</td>
<td>9</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Validation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of samples</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>RMSECV$^e$</td>
<td>4</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>RMSEP$^f$</td>
<td>0.89</td>
<td>−0.21</td>
<td>−0.01</td>
</tr>
<tr>
<td>R$^2c$</td>
<td>0.98</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>RPD$^d$</td>
<td>8</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>RER$^g$</td>
<td>35</td>
<td>25</td>
<td>30</td>
</tr>
</tbody>
</table>

$^a$ LV = latent variable.

$^b$ RMSEC = root mean square error of the calibration.

$^c$ $R^2$ = determination coefficient.

$^d$ RPD = ratio performance deviation.

$^e$ RMSECV = root mean square error of cross-validation.

$^f$ RMSEP = root mean square error of prediction.

$^g$ RER = range error ratio.

Fig. 2. Plots of regression coefficients for the sucrose, glucose and fructose PLS models. Sucrose (A), glucose (B) and fructose (C).
the O—H and C—H absorption band positions caused by inter-molecular hydrogen bonds. The absorption bands due to O—H and C—H groups in sugars in which sucrose contains eight O—H functional groups and glucose and fructose contain five groups each largely influence the spectral variation although they may still be differentiated [24]. Spectral ranges between 7200 and 6600, 6000 and 5500, 5400 and 4600, and 4600 and 4000 cm⁻¹ may be attributed to O—H stretch first overtone, C—H stretch first overtone, O—H combination bands and C—H combination band regions, respectively [25]. Strong peaks between 7400 and 6400 cm⁻¹ and between 5400 and 4600 cm⁻¹ are mainly related to the first overtone of O—H stretching and O—H combination bands of water, respectively and were not use to develop the PLS models. Spectral regions between 5800 and 5400 cm⁻¹ and between 4600 and 4000 cm⁻¹ are related to the first overtone of C—H stretching and C—H + C—H and C—H + C—C combination bands, respectively, both attributed to vibrations of the molecules of sugars [24,26]. Considering the absence of significant signals in this region between 10,000 and 7800 cm⁻¹, it was deleted previously to the development of the models.

All NIR spectra collected were preprocessed with mean centering, whilst the presence of scattering and baseline deviations were corrected by SNV (standard normal variate) and first derivative with 9-point Savitzky-Golay (9 on the right, 9 on the left). Samples are divided into calibration (n = 100) and validation (n = 60) sets utilizing the Kennard-Stone algorithm [21]. Calibrations sets cover the widest range of sugar concentration (Table 2).

The preprocessed spectra (5800–5400 cm⁻¹ and 4600–4000 cm⁻¹) were submitted to PLS calibration to give the most accurate models for sucrose, glucose, and fructose content. RMSEC for calibration set, RMSECV a full cross-validation, RMSEP for prediction set and R² were considered to evaluate results. RMSEC provides information about the adjustment of the model to calibration data.

Latent variables (LVs) can be used to reduce the dimensionality of data, and the optimal number of latent variables (LVs) was determined by the lowest value of predicted residual error sum of squares (PRESS) [27]. Consequently, the calibration optimal models were selected to high R², and low RMSEC, RMSECV, RMSEP and bias [28].

Fig. 2 shows the regression coefficients for the models. The coefficients for sucrose, glucose and fructose present a great similarity among them. The highest variation was associated with frequencies in the 7800–4000 cm⁻¹ region. In general, coefficients associated with water vibrations are negative, while the coefficients associated with sugars are positive [15,17]. The information-rich region from 4600 to 4000 cm⁻¹ can be ascribed to combinations of O—H bend/hydrogen-bonded O—H stretch (4428 cm⁻¹), O—H stretch/C—C stretch (4393 cm⁻¹) and combinations of C—H/C—C (4385–4063 cm⁻¹) vibrations of the sugar molecules [28].

Accuracy of the generated PLS models was attested by trueness and precision studies. Trueness of multivariate methods is evaluated by RMSECV, RMSEC and RMSEP. All the models presented good correlation between reference values and NIR predicted ones. Fig. 3 shows the correlation between values determined by the reference analysis method and values predicted by the NIR for external validation.

---

**Fig. 3.** Plots of sugar content predicted by the proposed NIR method versus reference values obtained by liquid chromatography using an independent test set of sweet sorghum juice samples. Sucrose (A), glucose (B) and fructose (C).
A model with five latent variables (LVs) minimized the root mean squared error of cross validation (RMSECV) and maximized $R^2$ for sucrose, or rather RMSECV $= 4$ mg mL$^{-1}$ and $R^2 = 0.99$. A model with eight LVs was selected for glucose, with RMSECV $= 1$ mg mL$^{-1}$ and $R^2 = 0.97$. In the case of fructose, a model with seven LVs showing RMSECV $= 0.6$ mg mL$^{-1}$ and $R^2 = 0.95$ was selected. RMSECV expresses the degree of agreement between estimated values by a model previously constructed and a real or reference value [29]. When the predicted values were plotted against the reference values for sucrose, glucose and fructose, the validation samples achieved a root mean squared error of prediction (RMSEP) of 4.1 and 0.6 mg mL$^{-1}$, respectively.

The precision was only estimated at the level of repeatability because the sugar content of sweet sorghum juice changes over time and the intermediate precision cannot be evaluated. Consequently, repeatability was evaluated by estimating relative standard deviations (RSD) for triplicates of three sweet sorghum juice samples with low, medium and high sugar contents. RSD varied 0.2, 0.3, 0.8% for sucrose; 1, 2, 2% for glucose; 1, 2, 3% for fructose, respectively. These values can be compared with the expected values issued from the Horwitz eq. [30] and acceptable RSD (~4%) were obtained.

The accuracy of the method was evaluated by the elliptical joint confidence region (EJCR) test, which is frequently used to evaluate accuracy of new analytical methods. This ellipse must contain values of intercept $= 0$ and slope $= 1$, which indicate the absence of systematic error of new analytical methods. This ellipse must contain values of intercept $= 0$ and slope $= 1$, which indicate the absence of systematic error of new analytical methods. This ellipse must contain values of intercept $= 0$ and slope $= 1$, which indicate the absence of systematic error of new analytical methods.

The authors are grateful to Fapemig for providing student fellowships and research grants (process number 3031/2014) to Embrapa Maize and Sorghum.

References
