

Evaluation of Near Infrared Spectroscopy for the Direct Analysis of Cane Quality Characters

S. S. Koonjah^{1,*}, A. Beekharry², M. G. H. Badaloo¹, C. Henderson³, A. Dookun Saumtally⁴

¹Department of Plant Breeding, Mauritius Sugarcane Industry Research Institute, Mauritius

²Ex-Research, Mauritius Sugarcane Industry Research Institute, Mauritius

³Rex Consulting Pty Ltd, Australia

⁴Department of Plant Biotechnology, Mauritius Sugarcane Industry Research Institute, Mauritius

Received July 10, 2019; Revised September 11, 2019; Accepted September 22, 2019

Copyright©2019 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract The use of near infrared spectroscopy, JEFFCO Infracana II, in the breeding and selection programme of the Mauritius Sugarcane Industry Research Institute (MSIRI) has led to a more reliable and rapid approach in estimating essential laboratory cane quality characters. The assessment of sucrose and fibre content now takes less than one minute per sample of cane with increased analytical precision and requiring considerably less labour. The latest calibration model in 2017 is based on approximately 3000 samples for Brix % cane, Pol % cane and fibre % cane. With this calibration model, the number of outliers generated that necessitate a separate laboratory analysis has been reduced to 5% or below. In fact, with the latest calibration model developed in 2018, the number of outliers has been considerably reduced and is now below 2%. The inclusion of more cane samples from high biomass and high sucrose sugar cane varieties will improve the robustness of the calibration models and further limit the number of outliers. Results from wet laboratory analyses were linearly regressed against those obtained from NIR giving a Pearson R² values of 0.89 for Pol % cane, 0.90 for Brix % cane and 0.74 for fibre % cane.

Keywords Cane Analysis, Sucrose Content, Juice Purity, Sugarcane, NIR

1. Introduction

Sugar cane samples at the Mauritius Sugarcane Industry Research Institute (MSIRI) originating from the preliminary and final selection stages were analysed for cane quality characters including sucrose and fibre content since the 1950s. The traditional analytical method consists

of crushing whole cane samples, extracting the sucrose using a wet disintegrator, separating the fibre from the extracted juice and filtrating the juice using lead acetate as a clarifying agent (Saint Antoine, 1968). The process is time consuming, labour intensive, expensive and is subject to large sampling errors.

Choosing NIRS over chemical analysis brings many advantages. In routine operation, NIRS provides rapid analysis, requires less labour, is environment friendly, has a low cost of analysis, reduces sample handling due to its non-destructive behaviour and increases safety by avoiding the use and disposal of dangerous chemicals. Additionally, electronic production and handling of data replaces manual handling, improving data accuracy and reliability. However, initial development of calibration models can be time consuming and expensive, requiring sometimes large volumes of primary wet laboratory analysis. Entry costs for NIR instrumentation can be a barrier and a high level of chemometric knowledge is required for effective modelling outcomes (Broad et al., 2002; Li-Chen et al., 2002). Moreover, NIRS is a secondary method of analysis, and as such relies on a library of primary laboratory analysis to provide it with its inherent "experience".

The infrared region of the electromagnetic spectrum is usually divided into three spectral regions: near-, mid- and far-infrared. Near-infrared spectroscopy (NIRS) is used to measure chemical and physical characteristics of biological materials. The application of electromagnetic radiation to a sample results in either reflection, transmittance or absorbance. Radiated energy excites hydrogen molecular bonds (-CH, -OH, -SH and -NH) resulting in a harmonic vibration (stretching and stretching/bending combinations) that absorb energy at a frequency (wavelength) dependent on the mass of the atoms in the molecular bond. Most absorbance occurs as overtones of localised vibrational modes such as hydrogen bond stretching (Metrohm, 2013).

NIRS is a type of vibrational spectroscopy, which corresponds to the wavelength range of 700 to 2500 nm (Pasquini, 2003), i.e., 0.7 to 2.5 microns (Figure 1). Sensitive detectors measure the total radiated energy and the total less absorbed energy at the sample and describe a difference spectrum across the range of wavelengths of the radiation (figure 2). Variations in absorbance in the spectra can be correlated to patterns in the matched primary analysis, using multivariate analysis techniques, providing predicted measurement of qualitative and quantitative characteristics of interest.

NIR multicomponent analyses of forage, fibre, grain and cereal are well documented (Williams and Norris, 1987). The possibility for application of NIRS to cane quality measurement (for payment systems and for variety development) has prompted several feasibility studies on reflectance spectroscopy of finely chopped cane (Berding, et al., 1991; Schaffler and Meyer, 1996). These early studies used a scanning NIR instrument with a long light

path that was affected by humidity variation.

Clarke et al., (1994) placed emphasis on cane quality assessment for grower payment using cane samples and conventional analyses directly from a Louisiana core laboratory. Recently, the Sugar Milling Research Institute in South Africa has also taken an interest in NIR analysis (reflectance) of finely chopped cane for the prediction of DAC (direct analysis of cane for Pol, Brix and moisture) analyses. Their initial study based on less than 200 samples without validation was encouraging. Today NIR systems are in widespread use across many sugar industries (Australia, South Africa, Brazil, Guatemala, Philippines, Mauritius, Columbia, Nicaragua and El Salvador) for cane quality assessment and cane payment determination (Edey and Clarke, 1996; Ochola et al., 2015; Ferraz and Molin, 2016). The use of NIR in sugar industry research facilities is becoming widespread and the variety of analytical applications continues to grow both in research and in commercial factory operations.

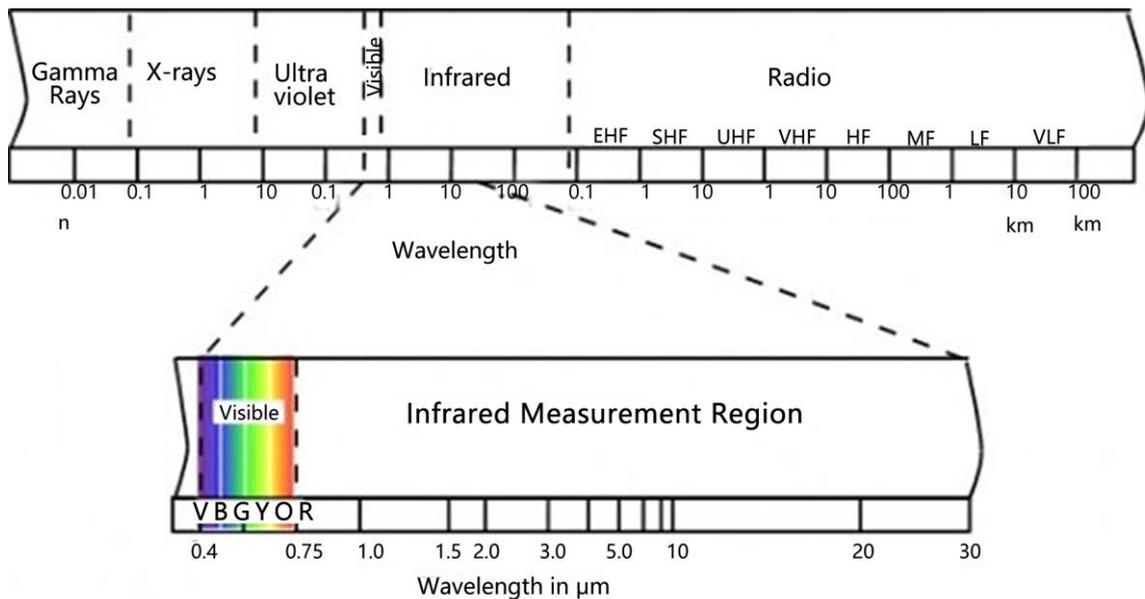


Image source: (<http://www.ipi-infrared.com.au/how-do-infrared-cameras-work/>).

Figure 1. The near-infrared region corresponds to wavelengths ranging between 0.7 and 2.5 μm

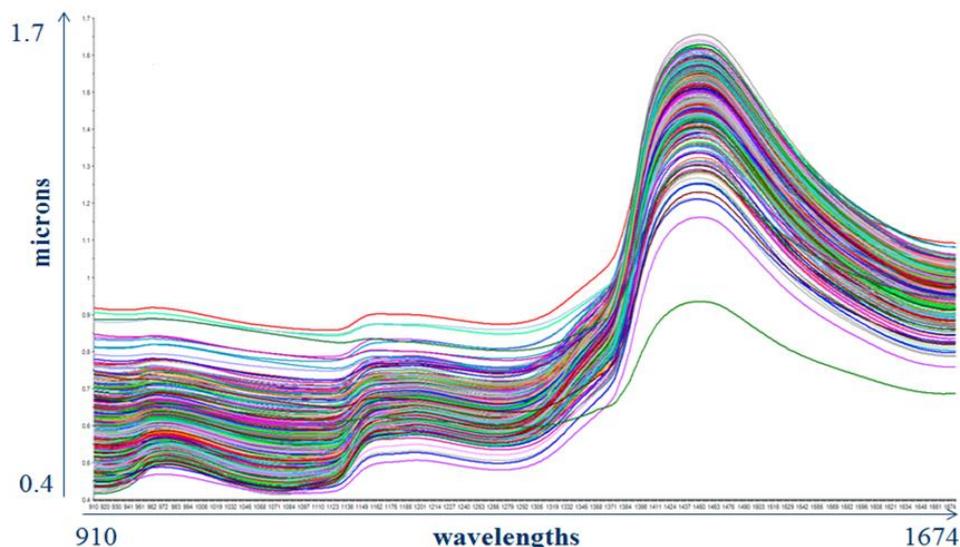


Figure 2. Typical sugar cane spectra

Each year, the Cane Analysis Laboratory of the MSIRI processes some 13,000 samples issuing from selection trials, both preliminary and final stages. In addition, crop monitoring is an ongoing process. Each year, assessment of cane samples is carried out during the period May to November across 113 sites in five sugar cane sectors of the island. These sites cover the various agro-climatic zones, different varieties under cultivation and stage of development and ripening of the crop. This paper describes the process of evaluation and development of cane analysis using the NIR InfraCana equipment together with the calibration model developed and its reliability.

2. Materials and Methods

2.1. Data Acquisition

Cane samples received at the laboratory for analysis were mainly obtained from trials on preliminary and final stages of selection, specific research projects, and for monthly national crop monitoring purposes. A wide range of cane characteristics were thus processed ranging from high fibre to high sucrose. Samples received were usually clean cane, i.e., devoid of trash, and were received at the Cane Analysis Laboratory during the period of April to December covering sugar cane varieties adapted for early-to late-harvest season.

Data acquisition for calibration covered period August 2015 to May 2017 and three consecutively improved calibration models were developed for the NIRS. Each cane sample received at the laboratory consisted of 6 millable stalks (approximately 5 to 8 kg), cut in pieces of 50 to 60 cm in length and was bar-coded. Each sample was processed using a JEFFCO InfraCana II Automated Cane Analysis system incorporating a Jeffco Cutter Grinder, a conveyor, and a Jeffco NIR spectrophotometer integrated

into an automated system for shredding, conveying and analysing finely shredded sugarcane for cane quality characters.

2.2. Data Collection by Conventional Laboratory Analysis (Primary Analysis) and NIR

To develop a first calibration model for the NIRS, all samples processed through the JEFFCO InfraCana II were sent for wet laboratory analyses to collect reference values. Spectra and predicted values were recorded using a core sampled sugar cane calibration model developed by the Cane & Arbitration Department in Mauritius. Once sufficient data was collected, an initial MSIRI calibration model was developed and used for further processing of MSIRI samples with improved predictive performance. Development of further models proceeded following an iterative process. All scanned sugar cane samples were processed in the wet laboratory as per the following methodology:

Exactly 329 g of finely crushed cane was collected from the InfraCana II, added to a JEFFCO wet disintegrator followed by 1L of water and 10 mL of 5% sodium bicarbonate solution. The mixture was disintegrated for 10 min to extract sucrose based on the methodology of Saint Antoine (1968). The diluted extract obtained was sieved to separate the fibre from the diluted juice. The latter was divided into two 250 ml parts, one for Brix and the other one for Pol reading, respectively. Brix reading was obtained using an automatic microprocessor controlled critical angle refractometer (Schmidt + Haensch GmbH & Co., Berlin, Germany). Prior to polarimetric analysis, the second part of the juice was thoroughly mixed with 2 g of Octapol Plus (Baddley Chemicals, Louisiana, USA), a lead-free clarifying reagent, and the mixture poured through filter paper (Whatman No. 91). The filtrate was collected and Pol reading was obtained using the

Saccharomat® (Schmidt + Haensch GmbH & Co., Berlin, Germany), an automatic sugar polarimeter. The fibre was thoroughly washed with tap water to remove maximum juice and collected in a tared fibre bag. The fibre was left to dry in a drying oven at 105 °C for 48 h (Saint Antoine, 1968). Fibre % cane was then derived as follows:

Where, $W1$ = weight of empty bag

$W2$ = weight of bag + fibre after drying (STASM, 1991)

The polarimeter reading refers to the Pol % cane of the sample directly whereas Brix % cane, which is equivalent to the proportion of total soluble solids, is derived from the diluted Brix measured in the laboratory according to the following formula (STASM, 1991):

2.3. Pre-treatments, Data Preparation and Multivariate Data Analysis

To increase signal from the characteristics being analysed and to reduce background information (noise), the collected spectra were subjected to pre-treatments using: (1) standard normal variate (SNV) to correct for light scattering and path length by centering and scaling each spectrum individually, so each has a mean equal to 0 and standard deviation equal to 1, and (2) Savitzky-Golay second order derivatives to remove the influence of any baseline variations. A Savitzky-Golay filter is a digital filter that can be applied to a set of digital data points for the purpose of smoothing the data in order to increase the signal-to-noise ratio without greatly distorting the signal.

The NIR spectra of outlier cane samples and their respective matched reference values, i.e., values for Pol % cane, Brix % cane and fibre % cane obtained from wet laboratory analyses, were used to develop the chemometric models for prediction using the statistical software Unscrambler® X version 10.3 (CAMO Software, Oslo, Norway). A Kennard-Stone sample selection procedure was used to select about half of the data for calibration. The calibration model was created by applying a mathematical algorithm to the data which correlates spectral data of the calibration samples with the respective reference data to derive the model. Validation of the calibration model is a crucial step to determine its suitability to predict new cane

samples and was done with the remainder independent and unused samples not included in the calibration development process.

Partial least-squares (PLS) regression, a multivariate analysis technique, was used to calibrate the spectral data with the reference data (i.e. the wet laboratory data). The maximum number of latent variables (LVs) used in the development of PLS models was eight. PLS regression is a useful tool to construct predictive models when there are many predictor variables which are highly collinear. The PLS regression algorithm selects successive orthogonal factors that maximise the covariance between predictor (spectra) and response variables (reference data). The final PLS calibration models were evaluated by a number of statistical parameters generated by both calibration and validation processes.

3. Results and Discussions

From 2015 to 2017, three calibration models for Pol % cane, Brix % cane and fibre % cane were developed and installed into the JEFFCO InfraCana II with consecutively improved robustness and reliability. The first model 100M2016R installed in April 2016 used approximately 200 sample reference data with corresponding spectral data. As a result, 83% of all cane samples processed during the following months could be directly analysed with the NIRS, i.e., 17% of samples (outliers) were sent for wet laboratory analyses. With more data collection, a second calibration model was developed (100M2016JUL) using 1076 samples for both Brix % cane and Pol % cane, and 2285 samples for fibre % cane. The model was installed in August 2016 with slight improvement in robustness (i.e., 12% outliers). With a high number of varied cane characteristics ranging from high biomass to high sucrose processed from 2015 to mid-2017, a third model (100M2017JUN) was developed using about 3000 samples for the three cane quality characters, and installed in June 2017. This new model showed further improvement in robustness and reliability with only 5% of cane samples requiring analysis using conventional methods (Figures 3a, 3b and 3c).

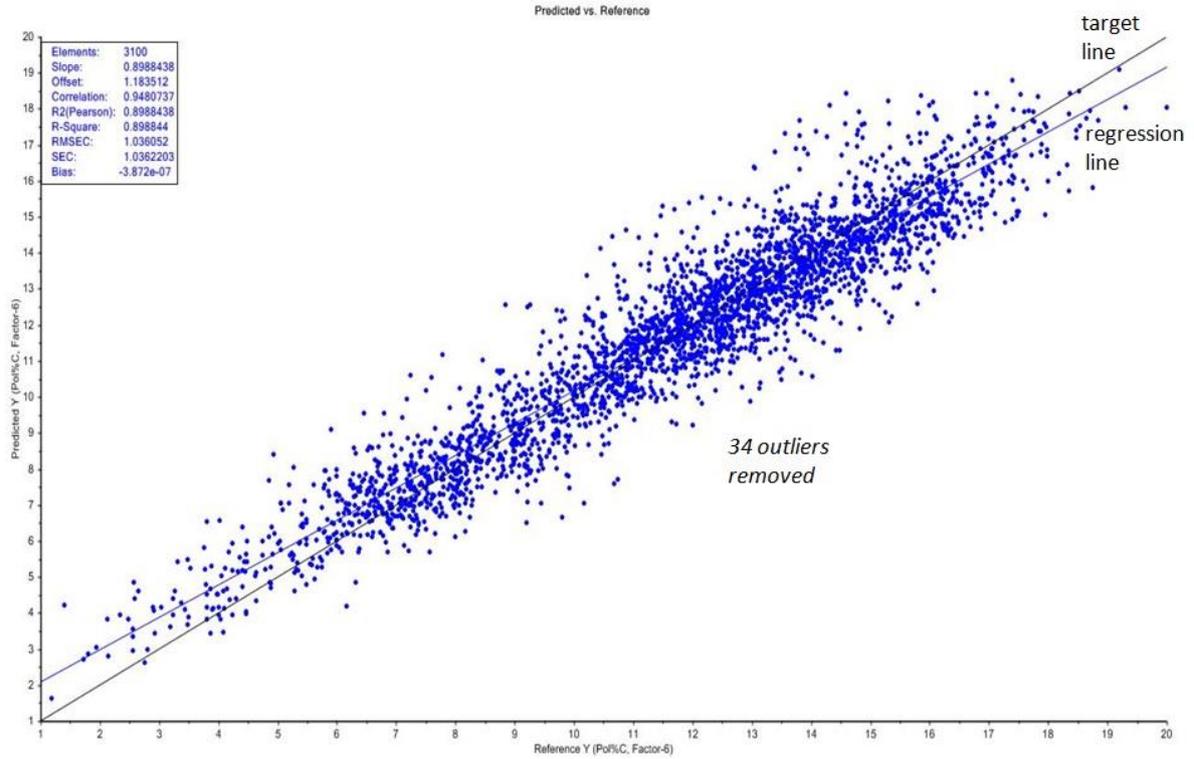


Figure 3a. Calibration model for Pol % cane

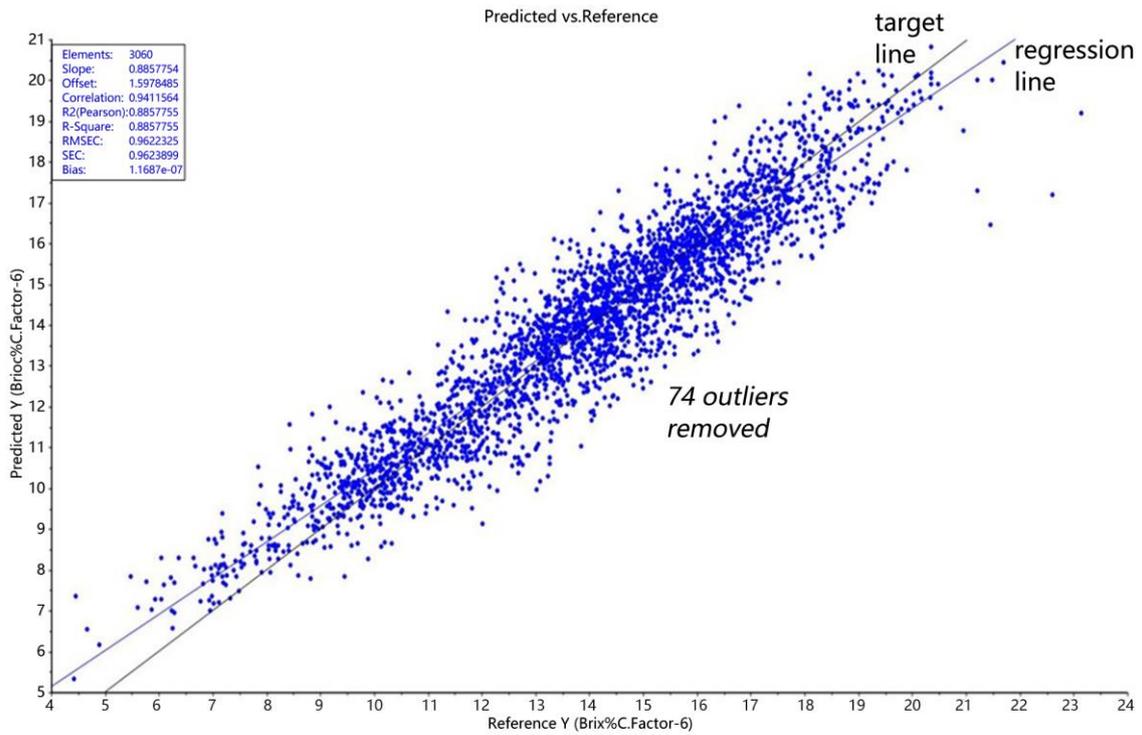


Figure 3b. Calibration model for Brix % cane

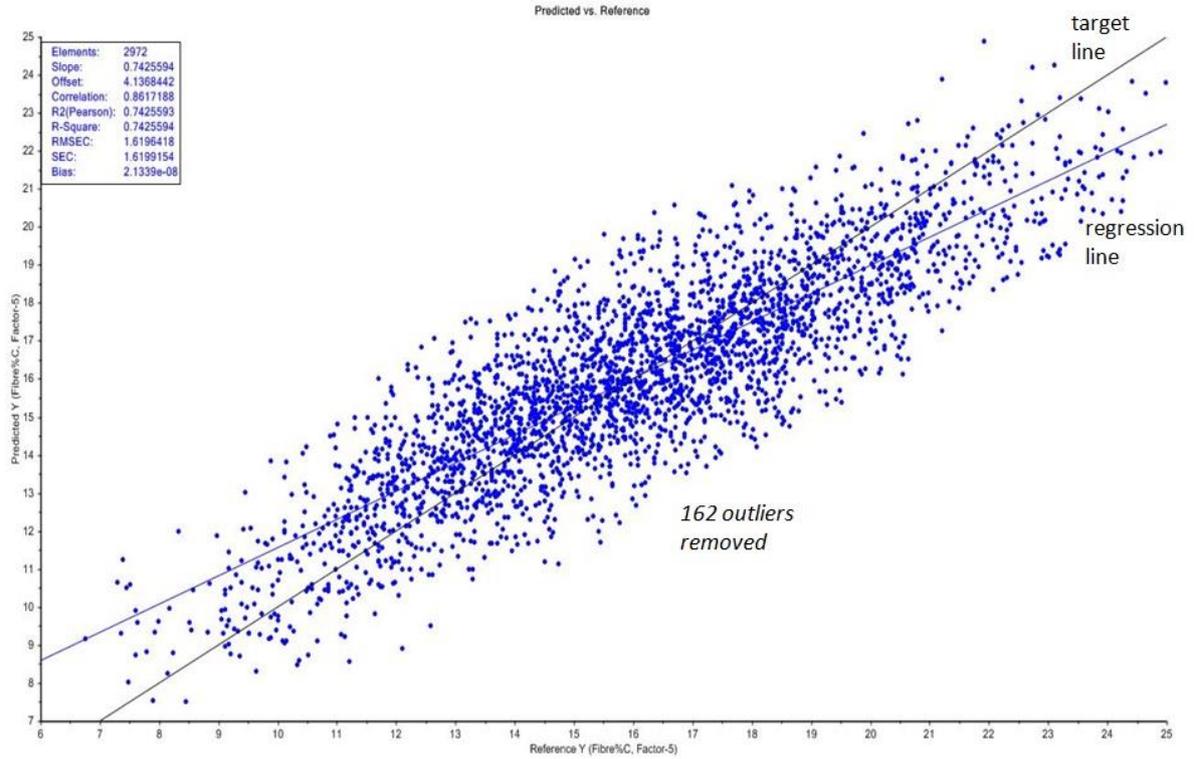


Figure 3c. Calibration model for fibre % cane

Table 1. Calibration and validation results for calibration models 100M2016JUL and 100M2017JUN

Calibration Model	Parameters	Calibration							Validation			
		N	Mean	Median	Range		RMSEC	R ²	N	R ²	RMSEP	RPD
					Min.	Max.						
100M2016JUL	Brix % cane	1076	12.16	11.85	4.23	20.69	0.99	0.85	1076	0.83	1.21	2.11
	Pol % cane	1076	9.36	9.00	1.18	18.66	0.88	0.90	1076	0.89	1.20	2.28
	Fibre % cane	2285	16.00	14.59	6.44	32.21	1.55	0.73	2285	0.71	2.44	1.34
100M2017JUN	Brix % cane	3060	14.00	14.31	4.42	23.14	0.96	0.89	3034	0.87	1.01	2.61
	Pol % cane	3100	11.7	12.19	1.18	20.00	1.04	0.90	3050	0.90	1.00	3.04
	Fibre % cane	2972	16.07	16.02	6.75	25.00	1.62	0.74	2904	0.75	1.56	1.74

N = number of sugar cane sample processed

R² = Coefficient of Determination (Pearson)

RMSEC = Root Mean Square Error of Calibration

RMSEP = Root Mean Square Standard Error of Prediction

RPD = Ratio of Standard Error of Prediction to Standard Deviation

CAMO's Unscrambler X software visualisation and statistical tools were used to help identify residual outliers (differences between laboratory and NIR results) and spectral outliers (spectra that are significantly different from those used in the training set, i.e., the set of spectral data initially used to develop the calibration models). These spectral outliers affect the stability and robustness of the NIR calibrations and can be either deleted and the sample re-run or added to the training set prior to recalibration, depending on the conditions of the sourcing of the sample. Outlier data points have a wide variety of sources ranging from uneven surfaces or too much field mud at the sample surface, year to year variations in cane varieties and changes in general weather conditions (e.g. cane from year of severe drought followed by cane from a year of good rainfall) (Edye and Clarke, 1996).

The coefficient of determination (Pearson's R^2) provides an estimation of how much variance between reference and predicted values is explained versus the total variance. As shown in Table 1, slight improvement in R^2 was possible when the last model 100M2017JUN was developed using software Unscrambler X. The following R^2 values were obtained in the third model: 0.89 for Brix % cane, 0.90 for Pol % cane and 0.74 for fibre % cane based on sample size of 3034, 3050 and 2904 for Brix % cane, Pol % cane and fibre % cane respectively. Ochola et al., (2015) have reported R^2 values reaching 0.9787, 0.9503 and 0.8725 for Brix % cane, Pol % cane and fibre % cane respectively, based on 688 samples.

During the process of model development, apparent outlier values for each parameter were not systematically removed. These values were cross-verified to ensure that they were not produced out of bad laboratory practices or any other erroneous technical or mechanical issues. Regular standard error of laboratory measures was performed in order to ensure correct laboratory practices. The validation values of RMSEP, which account for bias and relate to the reliability and predictive ability of the model, showed higher accuracy for model 100M2017JUN. The RMSEP for Brix % cane, Pol % cane and fibre % cane was reduced from 1.21 to 1.01, 1.20 to 1.00 and 2.44 to 1.56, respectively as shown in Table 1.

The ratio of performance of deviation or relative predictive determinant (RPD) is indication of the model's robustness, i.e., its ability to predict future data in relation to the initial variability of the calibration data. Ranges of RPD values related to the calibration suitability have been provided by Williams (2001) where values above 3 represent an excellent model, while values below 2.3 indicate a poor calibration performance, with use for predicting new samples not advisable. As shown in Table 1, there is a significant improvement in the RPD for Brix % cane (2.61) and Pol % cane (3.04) for model 100M2017JUN. Because of the relatively higher bias (RMSEP) for fibre % cane, the latter's RPD tends to remain small. This can be remedied as more fibre reference

data are collected and incorporated in newer calibration models. However, a drastic change of bias is not expected because of the inherent characteristics pertaining to sugar cane fibre. In fact, natural fibres from sugar cane have water which is not associated with the juice water. It is known as hygroscopic water or Brix-free water. Brix-free water % fibre has been estimated to be around 25% (Anon, 1984), but many studies reveal that brix-free water content varies considerably with time (day to day, week to week and month to month), sugar cane varieties, time of harvest and climatic conditions (Fourmond, 1965). It may also vary with the sugar content of the plant (Wong Sak Hoi and Martincigh, 2013). The higher bias in the fibre % cane calibration model can therefore be attributed to the fluctuations in Brix-free water in fibre samples over the course of analyses made over time.

It is important to note that the method used to calculate fibre % cane (Saint Antoine, 1968) dates back to a time where relatively lower cane samples were analysed and the sugar cane varieties cultivated were not comparable to current commercial varieties. Since brix-free water cannot be removed from collected fibre using the current method of drying, as described previously, another potential alternative is to measure moisture % cane through wet laboratory analysis and then convert to fibre % cane.

4. Conclusions

Saint Antoine (1968) simplified the methods and calculations, and cut down on the time taken for the analysis of sucrose and fibre content in samples at the MSIRI. Traditional wet laboratory analyses take about 15 min to prepare one sample, analyse for Brix % cane and Pol % cane, and 48 h (oven drying) for the determination of fibre % cane. Since the adoption of direct analysis of cane samples using NIRS, at MSIRI significant changes have taken place in a short period of time, and one of the benefits obtained is the drastic reduction in analysis time. It now takes less than 1 minute to analyse for cane quality characters for one sample.

From 17% outliers since the installation of the first model to only 5% using the last model the NIR technology has proved to be rapid and reliable for the evaluation of sucrose and fibre content resulting in a drastic reduction in resources such as labour and chemical reagents. The technology will be of great benefit in the processing of thousands of samples from ongoing experimental trials as well as crop monitoring. Because the NIR technique can be applied with little or no sample preparation, analysis times are reduced from hours to minutes and furthermore several analytical results can be obtained from the same NIR data while the conventional analysis would often require another technique and more hours of work. Using more than 300 reference values from 2017 mid- and late-season harvest periods, the development of another more reliable

calibration model will further enhance the use of NIRS for direct analysis of cane samples.

Further to the application of NIR technology to analyse sucrose and fibre content directly, which have been the quality characteristics used for cane breeding purposes for a long time at the MSIRI, there is a need to introduce more quality characteristics such as moisture content and reducing sugar content. Knowledge of the percentage of fructose and glucose, in addition to sucrose content, would give a more precise value to the sugar content of a variety. Considering the further challenges in developing a robust model for fibre % cane, creating a calibration model for moisture % cane would bring more reliability to future NIRS models developed. If these improvements are brought about in the near future, the NIRS would prove to be an indispensable tool for use in the sugarcane breeding and selection programme of the Institute.

Acknowledgements

The authors thank the ACP secretariat, the European Commission and the Coordinating Unit of the African, Caribbean and Pacific Sugar Research Program (ACP-SRP) for financial support provided for the purchase of the NIR InfraCana cane analysis equipment used in this study and also express their gratitude to Dr. A Salem Saumtally, Director, Mauritius Sugarcane Industry Research Institute for his encouragement and interest in this study

REFERENCES

- [1] Anon (1984). The standard laboratory manual for Australian sugar mills. Bureau of Sugar Experiment Station 1, 2-3.
- [2] Berding, N., Brotherton, G.A. & J.C. Skinner (1991). Near Infrared Reflectance Spectroscopy for Analysis of Sugarcane from Clonal evaluation trials: I. Fibrated Cane. *Crop Science* 31, 1017-1023.
- [3] Broad, N., Graham, P., Hailey, P., Hardy, A., Holland, S., Hughes, S., Lee, D., Prebbe, K., Salton, N. & P. Warren (2002). Guidelines for the Development and Validation of NearInfrared Spectroscopic Methods in the Pharmaceutical Industry, in *Handbook of Vibrational Spectroscopy*, Vol. 5, Chalmers, J. M. and Griffiths, P. R. (Eds), Wiley, Chichester, UK, pp. 3590–3610.
- [4] Clarke, M.A., Edey, L.A. & W.S. Patout (1994). Sugarcane crop analysis by NIR. *Proc. Conf. Sugar Processing Research*, 244-264.
- [5] Edey, L., & M. Clarke (1996). Sugarcane quality analyses by near infrared spectroscopy. *Proc. S. Afr. Sug. Technol. Ass.* 70, 127-130.
- [6] Ferraz, M.N., & Molin, J.P. (2016). NIR spectroscopy to map quality parameters of sugarcane. *Proceedings of the 13th International Conference on Precision Agriculture*. St Louis, Missouri, USA. August, 1-13.
- [7] Fourmond, T.H. (1965). The influence of Brix Free Water when assessing milling performance. *Proceedings of the South African Sugar Technologists' Association*. March, 105-107.
- [8] Li-Chen, E.C.Y., Ismail, A. A. Sedman, J. & F.R. van de Voort, (2002). *Vibrational Spectroscopy of Food and Food Products*, in *Handbook of Vibrational Spectroscopy*, Vol. 5, Chalmers, J. M. and Griffiths, P. R. (Eds), Wiley, Chichester, UK, pp. 3629–3662.
- [9] Metrohm, (2013). *Monograph. NIR Spectroscopy. A guide to near-infrared spectroscopic analysis of industrial manufacturing processes*. Metrohm AG, CH-9101 Herisau, Switzerland, 46 pp.
- [10] (http://www.mep.net.au/wpmp/wp-content/uploads/2013/05/MEP_Monograph_NIRS_81085026EN.pdf)
- [11] Ochola, P., Apunda, E.O. & E.O. Yewa (2015). Evaluation of near infrared spectroscopy for sugarcane quality determination in Western Kenya. *Chemistry and Material Research* 10 (9), 30-37.
- [12] Pasquini, C. (2003). Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society* 14, 198–219.
- [13] Saint Antoine, J.D. de R. de (1968). A simplified method of calculating pol of cane harvested from experimental plots. *MSIRI Tech. Circular No. 32*: 13 pp.
- [14] Schaffler, K.J., & J.H. Meyer (1996). Near infrared analysis of shredded cane: a potential replacement for direct analysis of cane. *Proc. S. Afr. Sug. Technol. Ass.* 70, 134.
- [15] STASM (1991). *Laboratory Manual for Mauritius Sugar Factories (Official Methods)*. 207 pp.
- [16] Williams, P.C. (2001). Near-infrared technology: In the agricultural and food industries. In P. Williams, & K. Norris (Eds.), (2nd ed.). *American Association of Cereal Chemists*. (pp. 296) St. Paul, Minn., USA: American Association of Cereal Chemists.
- [17] Williams, P.C., & K. Norris (1987). Near-Infrared technology in the agriculture and food industries. *American Association Cereal Chemists*, St. Paul, MN 330 pp.
- [18] Wong Sak Hoi, L.W., & B.S. Martincigh (2013). An improved method to determine the 'Brix-free' water capacity of sugarcane fibres. *Sugar Industry* 138 (9), 583-594.