

Procedure for obtaining NIR spectra in sugarcane stalks using a portable instrument under field conditions

Andrea Carla Bastos Andrade¹, Luiz Alexandre Peternelli^{2*} , Márcio Henrique Pereira Barbosa² and Reinaldo Francisco Teófilo²

Abstract: Obtaining near-infrared spectra (NIR) directly from the sugarcane stalks in the field requires additional care due to the presence of wax on their surface. This study investigated the best cleaning procedure to be applied in sugarcane studies to obtain NIR spectra with reduced noise. NIR spectra were obtained from two sugarcane varieties using three cleaning methods (wax not removed, polyurethane abrasive sponge, and stainless steel sponge), and the sample sizes ranged from 2 to 8 stalks. Different pre-treatments were compared by principal components analysis (PCA). The p-value associated with the Mahalanobis distance measure was used to choose the best number of stalks to be sampled. PCA results demonstrated that the stainless steel sponge method is the most efficient procedure for distinguishing between the different sugarcane varieties once all wax is removed from the surface. Four stalks are necessary to obtain the average spectrum (p-value < 0.05).

Keywords: Near-infrared, sugarcane breeding, Mahalanobis distance

INTRODUCTION

Near-infrared (NIR) spectroscopy can produce a universal analytical response considering samples of organic composition (Pasquini 2018). This technique is widely used to measure various qualities of agricultural products, being accurate, reliable, non-destructive, rapid, and inexpensive, and it enables determining several properties simultaneously (Maraphum et al. 2018, Phuphaphud et al. 2020).

Sugarcane cultivation has great economic relevance due to the production of sugar, ethanol and energy from its biomass (Crystian et al. 2018). The NIR spectroscopy technique has been successfully applied in predictive models (Porto et al. 2019, Peternelli et al. 2020, Gonçalves et al. 2021) and genotype discrimination in sugarcane breeding programs (Peternelli and Andrade 2023). Also, studies with portable instruments have presented the possibility of obtaining NIR spectra on field conditions (Taira et al. 2013, Maraphum et al. 2018, Phuphaphud et al. 2020).

Naturally, spectral variability sources (including noise) are present in the NIR data and are not associated with the objective of the qualitative or quantitative models (Pasquini 2018). This noise might increase with the field conditions. The practicality of using NIR in the field is known to be constrained by the influence of unknown factors on the spectra in the NIR region, such

Crop Breeding and Applied Biotechnology
25(3): e520925311, 2025
Brazilian Society of Plant Breeding.
Printed in Brazil
<http://dx.doi.org/10.1590/1984-70332025v25n3n41>



***Corresponding author:**
E-mail: peternelli@ufv.br

Received: 22 February 2025
Accepted: 22 July 2025
Published: 26 August 2025

¹ Universidade Estadual de Maringá, Avenida Colombo, 5790, Zona 7, 87020-900, Maringá, PR, Brazil

² Universidade Federal de Viçosa, Avenida PH Rolfs, s/n, Campus Universitário, 36570-900, Viçosa, MG, Brazil

as temperature (Phuphaphud et al. 2020) and the presence of dirt on the sample's surface. Temperature and dirt are critical in sugarcane since most varieties contain wax on their stalk surface. Therefore, spectra collection practices in the field must be more careful.

Despite the efforts to evaluate sugarcane associated with NIR spectroscopy, an easy and practical procedure for obtaining spectrum must be established under field conditions as sugarcane breeding experiments can be very extensive, especially in the initial phases of genotype evaluations. In view of the above, this study aimed to compare different sugarcane stalk cleaning methods associated with varying sample sizes to obtain NIR spectra from a portable instrument in field conditions and acquire a more straightforward and faster procedure to be applied in sugarcane studies.

MATERIAL AND METHODS

Samples and NIR spectra collection procedure description

We evaluated two standard commercial varieties of sugarcane (Var1 and Var2) to assess cleaning efficiency in terms of reducing spectral variability and, therefore, increasing the signal-to-noise ratio. These commercial varieties are not only morphologically distinct, but also regarding their inner characteristics (UPOV 1961, Jördens 2011, Ćemalović and Petrović 2019). Var1 presents purple-colored internodes when exposed to the sun, with few cracks and weak waxiness, and Var2 has green-purplish internodes when exposed to the sun, with no cracks but strong waxiness. These varieties belong to the germplasm bank of the Sugarcane Genetic Breeding Program of Universidade Federal de Viçosa (UFV), located at the *Unidade de Ensino, Pesquisa e Extensão em Produção de Grandes Culturas e Bioenergia* (UEPE GCBE) in Viçosa, Minas Gerais.

The NIR spectra obtainment procedures are presented in Figure 1. They are described as a) 'wax not removed' method (WNR) – there is no cleaning of the sugarcane stalks, and the reading by the portable NIR instrument is done directly on the stalk (Figure 1A); b) 'polyurethane abrasive sponge' method (PAS) – cleaning the sugarcane stalks with a typical PAS, followed by reading by the portable NIR instrument (Figure 1B); and c) 'stainless steel sponge' method (SSS) – cleaning the sugarcane stalks with a more robust sponge, followed by reading by the portable NIR instrument (Figure 1C). Eight stalks were randomly selected from each variety to obtain the NIR spectra. The NIR spectra were ensured to be obtained from the same stalks chosen for all procedures evaluated in this experiment. Spectral samples were collected in the middle third of each stalk of the evaluated varieties. The collection spot was selected based on previous studies (data not shown), which also involved the NIR spectra acquisition.

The spectra were obtained using a portable MicroNIR instrument (VIAVI, USA) with a spectral range from 908 to 1676 nm, scans average equal to 100, and a reflectance mode configuration. The white reference was initially measured and again after every 8 sample measurements. One measurement was performed for each sample.

Chemometrics

The spectra initially underwent visual inspection and were subjected to different pre-treatments (Savitzky-Golay smoothing, first derivative, second derivative, multiplicative

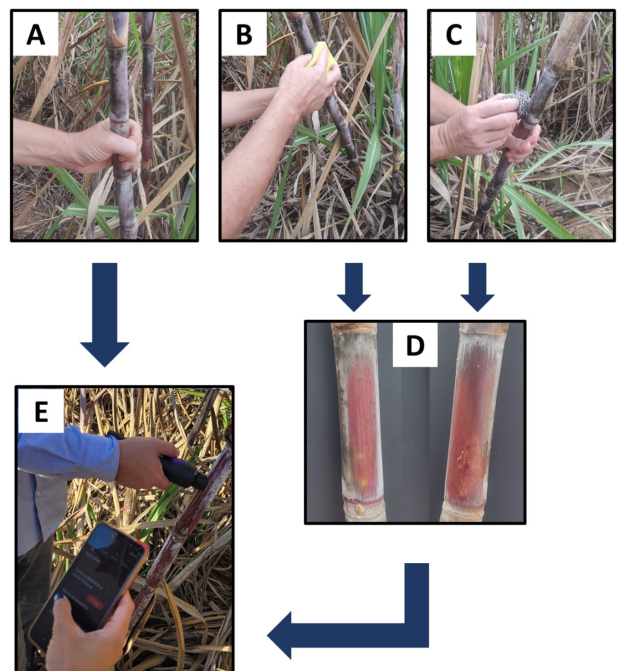


Figure 1. Details of the procedures for collecting NIR spectra in sugarcane. A) 'Wax not removed' method (WNR). B) 'Polyurethane abrasive sponge' method (PAS). C) 'Stainless steel sponge' method (SSS). D) Stalks cleaned by PAS and SSS methods. E) Obtaining NIR spectra using a portable MicroNIR instrument.

scattering correction, and mean centering) and some combinations of them. The process was the same as reported in Peternelli and Andrade (2023). A principal component multivariate exploratory analysis was performed to evaluate the best pre-treatment or combination of pre-treatments applied to the original data matrix and compared to no use of pre-treatments.

The Mahalanobis distance (D^2) (Manly 2004) was applied as a measure of dissimilarity to discriminate between both varieties. We adapted the procedure for decision-making on classifying pairs of individuals as previously described in Peternelli and Andrade (2023). In the present study, we aimed to obtain a more robust distribution of D^2 values and corresponding p-values, with 200 simulations being performed for each method (WNR, PAS, and SSS) associated with the number of stalks measured. In turn, the following steps were considered in each simulation:

Step 1: a cleaning method was chosen from the list: WNR, PAS, and SSS;

Step 2: the number of sampled stalks ($n = 2$ to 8) was defined to obtain the average spectra for each variety;

Step 3: a principal component analysis (PCA) was performed for each method and number of sampled stalks from steps 1 and 2 over the $2n$ spectra from both varieties. The aim was to reduce the dimensionality of the data due to the large number of wavelengths in the NIR spectra. This step involved considering the variety of scores associated with the first two principal components since two PCs captured most of the data variability;

Step 4: for each variety i ($i = 1, 2$), obtain a vector V_i containing the averages of the PC1 and PC2 scores and calculate the differences between these two vectors $V_{diff} = V_1 - V_2$;

Step 5: repeat step 4 $B = C_n^{2n}$ times under a permutation test approach (Davison and Hinkley 1997, Good 2005) to obtain a $B \times 2$ matrix containing the B resampled vector differences (V_{diff}^B);

Step 6: the combined covariance matrix (of dimension 2×2) was built considering the original vector V_{diff} (from step 4) and the B resampled (V_{diff}^B) vectors obtained in step 5;

Step 7: calculate Mahalanobis distance (D^2) for the vector V_{diff} obtained in step 4 and the mean vector $\bar{V}_{diff}^B$ obtained from averaging the B resampled vectors (step 5), weighted by the combined covariance matrix obtained in step 6. Consider the p-value associated with the D^2 value as a measure of decision-making (Peternelli and Andrade 2023).

All analyses were performed in the R environment (R Core Team 2024). The 'FactoMineR' (Lê et al. 2008) and 'factoextra' (Kassambara and Mundt 2020) packages were used for principal component analyses. The 'ggplot2' (Wickham 2009) and 'R-base' packages were also used to obtain other graphical visualizations and D^2 statistics. The authors developed all the scripts for this analysis, including simulations, in the Laboratory of Analysis and Research in Applied Statistics (LAPEA, www.lapea.ufv.br).

RESULTS AND DISCUSSION

Figure 2 shows the NIR spectra for the two sugarcane varieties considering the methods tested in the present study. Initial inspection of the NIR spectra revealed clear separation of the varieties for the SSS method (Figure 2C). It is also possible to note that the spectra obtained by the WNR method did not present a shift and inclination pattern in the baseline in all its samples (Figure 2A), characteristic of diffuse reflectance spectra (Ferreira 2015). Although the spectra obtained by the PAS method are more standardized concerning the displacement and inclination in the baseline, the spectral samples of the different sugarcane varieties are confused along the wavelengths (Figure 2B).

The spectral matrices, pre-treated or not, were subjected to principal component analysis. The results showed that the original data matrix, free of transformation and preprocessing techniques, was informative for the study objectives. It was then confirmed in an individual analysis with raw spectra for each tested method that the SSS method can better differentiate between samples of both sugarcane varieties (Figure 2F).

In the present study, the Mahalanobis distance dissimilarity measure (D^2) allowed us to identify the ideal number of stalks to obtain the average spectral value using the evaluated methods. Figure 3 shows the distributions of p-values associated with D^2 for each method evaluated and the number of stalks used. The p-value associated with D^2 generally decreases as the stalk counts increase to obtain the mean NIR spectra values. In turn, the p-values obtained for the SSS

method in both scenarios with a stipulated number of stalks stand out as the best for detecting divergence between both sugarcane varieties through NIR spectra. Interestingly, the PAS showed more significant variation among the methods studied and presented worse performance than the WNR method. One explanation for this behavior is the wax scattering from the stalks instead of its removal by the PAS cleaning method. Finally, considering the p-value = 0.05 threshold, the ideal number of stalks to obtain spectral averages in this study is equal to 4 using the SSS.

Corroborating these findings, an additional study was conducted to evaluate the accuracy of the methods using samples within the same variety (data not shown). In other words, the sample set of a single variety was randomly subdivided into two sample subsets, and the same procedure was performed to obtain the D^2 distance and its associated p-value between these two sample subsets. The results demonstrated that the SSS cleaning method also presented greater accuracy in inferring that the samples were of the same variety (median p-value = 0.63) compared to the PAS (median p-value = 0.46) and WNR (median p-value = 0.40) methods.

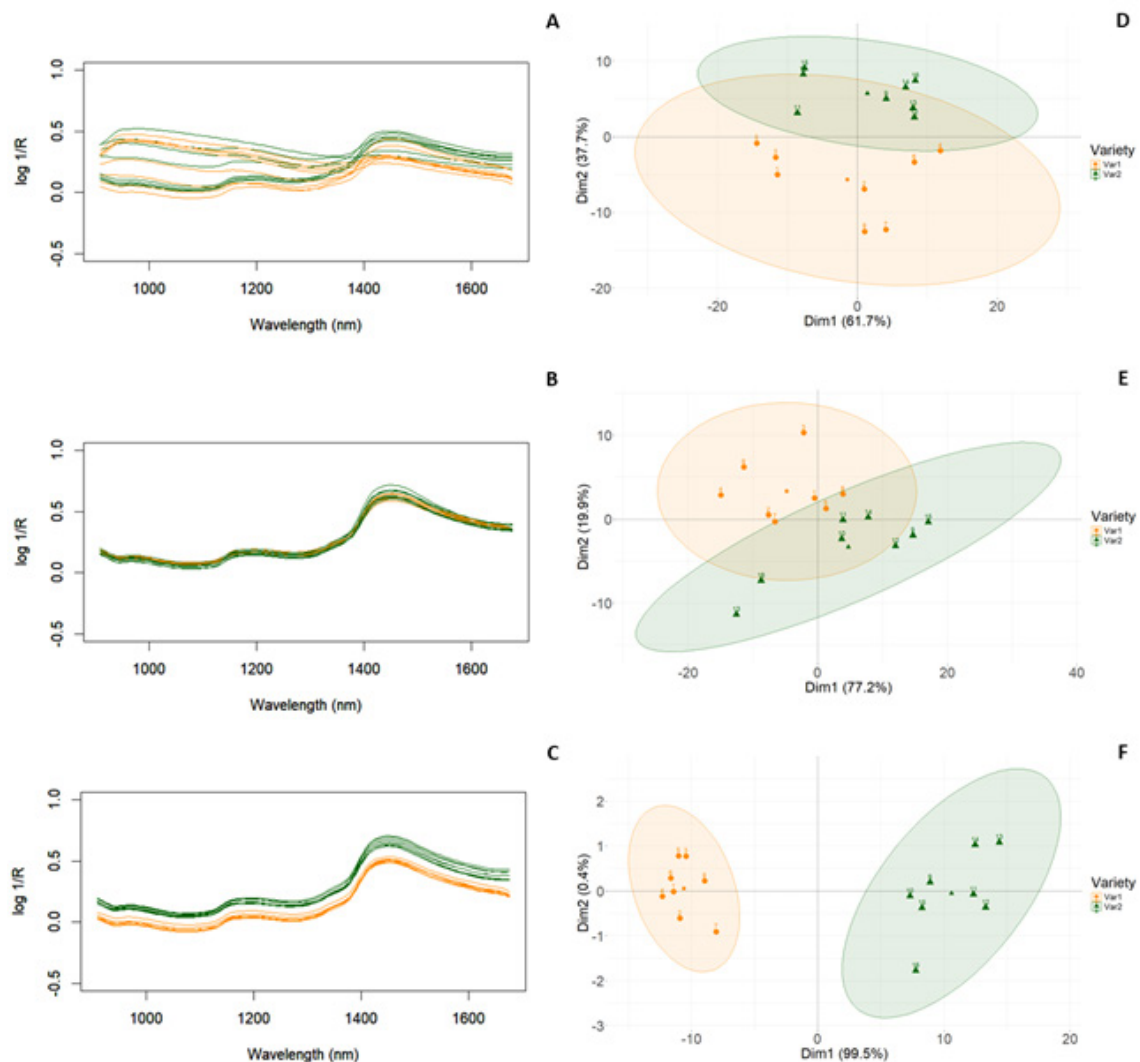


Figure 2. NIR spectra for each spectra collection procedure tested and principal components to differentiate the two sugarcane varieties used in this study, based on NIR spectra without pre-treatments. A) and D) 'Wax not removed' method (WNR). B) and E) 'Polyurethane abrasive sponge' method (PAS). C) and F) 'Stainless steel sponge' method (SSS). The orange and green colors represent the two sugarcane varieties (Var1 and Var2, respectively) used in this study.

Increasing the stalk sample size (or the number of NIR spectra measurements) promotes better discrimination between the varieties using any evaluated method to obtain spectral averages. In a study using a portable NIR instrument to measure sugar content in field conditions, the spectra collection was performed directly on the cane surface (WNR) with an average spectrum obtained from 8 readings (Phuphaphud et al. 2020). The predictive models provided good accuracy, with coefficients of determination (r^2) from the validation set ranging from 0.69 to 0.78. In addition, using an average spectrum obtained from 8 readings, Taira et al. (2013) demonstrated that the Pol (estimate of soluble sucrose) value could be predicted from the cane stalk with similar accuracy to the cane juice (r^2 from the calibration model developed using cross-validation equal to 0.87). Similar results were observed in Maraphum et al. (2018), in which r^2 increased with the increase in sample size (r^2 ranging from 0.65 to 0.82) in models of Pol value prediction. Although these findings address the evaluation of predictive models, they present the relevance of the increase in sample size to obtain the average spectrum. These findings corroborate the results presented in the present study, which showed that even when wax on the stalks was not removed, it was possible to distinguish both sugarcane varieties using portable NIR when the sample size increased.

However, considering the large number of genotypes to be evaluated in sugarcane experiments, the ideal method is to reduce the number of measurements required to obtain good analytical results. Furthermore, spectra from the WNR may show unpredictable behavior depending on the sample, as shown in Figure 2A. Maraphum et al. (2018) showed that the wax-removed surface using a 'hard steel plate' provided better performance in Pol prediction models with one scan (r^2 ranging from 0.73 to 0.83) compared with the WNR. Our results showed that the SSS method is another alternative to cleaning sugarcane stalks, which is essential for reducing the number of measurements and spectra variance, and that it is a faster and easier method to obtain the spectra.

CONCLUSION

The comparison of the three cleaning methods of cane stalks with the different sample sizes revealed that the SSS method associated with 4 NIR measures through a portable instrument allowed statistical discrimination between the two sugarcane varieties. This procedure is easy and practical to apply in extensive experiments of sugarcane breeding programs. It also enables reducing spectral noise.

ACKNOWLEDGMENTS

The authors are thankful to the *Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)* – grant number 312316/2023-2, the Sugarcane Genetic Breeding Program for the sample material, and *Financiadora de Estudos e Projetos (FINEP)* for the financial support for research projects.

DATA AVAILABILITY

The datasets generated and/or analyzed in this study are available from the corresponding author upon reasonable request.

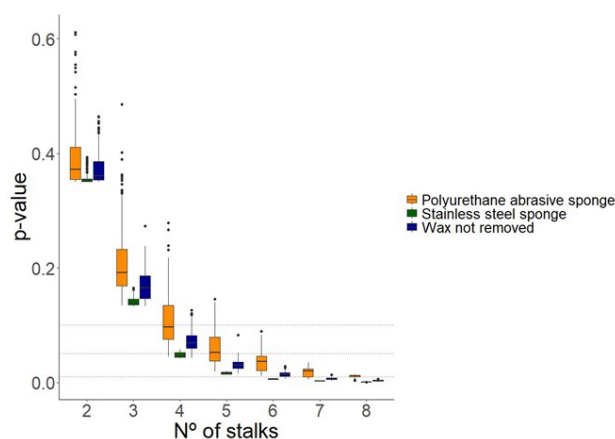


Figure 3. P-values associated with the Mahalanobis distance (D^2) to distinguish the two sugarcane varieties used in this study considering the average of different numbers of sugarcane stalks and the three methods evaluated to obtain NIR spectra. The red dashed lines represent the 0.01, 0.05, and 0.10 limit p-values.

REFERENCES

- Ćemalović U and Petrović M (2019) New varieties of plants and legal protection of breeder's right: The UPOV Convention and its major economic consequences. **Ekonomika poljoprivrede** **66**: 513-524.
- Crystian D, Santos JM, Barbosa GVS and Almeida C (2018) Genetic diversity trends in sugarcane germplasm: Analysis in the germplasm bank of the RB varieties. **Crop Breeding and Applied Biotechnology** **18**: 426-431.
- Davison AC and Hinkley DV (1997) **Bootstrap methods and their Application**. Cambridge University Press, Cambridge, 592p.
- Ferreira MMC (2015) **Quimiometria: conceitos, métodos e aplicações**. Editora Unicamp, Campinas, 496p.
- Gonçalves MTV, Morota G, Costa PMDA, Vidigal PMP, Barbosa MHP and Peternelli LA (2021) Near-infrared spectroscopy outperforms genomics for predicting sugarcane feedstock quality traits. **PLoS ONE** **16**: e0236853.
- Good P (2005) **Permutation, parametric and bootstrap tests of hypotheses**. Springer, New York, 336p.
- Jördens R (2011) Effective system of plant variety protection in responding to challenges of a changing world: UPOV perspective. **Journal of Intellectual Property Rights** **16**: 74-83.
- Kassambara A and Mundt F (2020) factoextra: Extract and visualize the results of multivariate data analyses. Package Version 1.0.7. R package version.
- Lê S, Josse J and Husson F (2008) FactoMineR: An R package for multivariate analysis. **Journal of Statistical Software** **25**: 1-18.
- Manly BFJ (2004) **Multivariate statistical methods: A primer**. Chapman and Hall/CRC, Boca Raton, 208p.
- Maraphum K, Chuan-Udom S, Saengprachatanarug K, Wongpichet S, Posom J, Phuphaphud A and Taira E (2018) Effect of waxy material and measurement position of a sugarcane stalk on the rapid determination of Pol value using a portable near infrared instrument. **Journal of Near Infrared Spectroscopy** **26**: 287-296.
- Pasquini C (2018) Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta** **1026**: 8-36.
- Peternelli LA and Andrade ACB (2023) Insights and protocols for discrimination of sugarcane clones by dissimilarity measures on RGB and NIR data. **PLOS ONE** **18**: e0288508.
- Peternelli LA, Gonçalves MTV, Fernandes JG, Brasileiro BP and Teófilo RF (2020) Selection of sugarcane clones via multivariate models using near-infrared (NIR) spectroscopy data. **Australian Journal of Crop Science** **14**: 889-896.
- Phuphaphud A, Saengprachatanarug K, Posom J, Maraphum K and Taira E (2020) Non-destructive and rapid measurement of sugar content in growing cane stalks for breeding programmes using visible-near infrared spectroscopy. **Biosystems Engineering** **197**: 76-90.
- Porto NA, Roque JV, Wartha CA, Cardoso W, Peternelli LA, Barbosa MHP and Teófilo RF (2019) Early prediction of sugarcane genotypes susceptible and resistant to *Diatraea saccharalis* using spectroscopies and classification techniques. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy** **218**: 69-75.
- R Core Team (2024) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
- Taira E, Ueno M, Saengprachatanarug K and Kawamitsu Y (2013) Direct sugar content analysis for whole stalk sugarcane using a portable near infrared instrument. **Journal of Near Infrared Spectroscopy** **21**: 281-287.
- UPOV - International Union for the Protection of New Varieties of Plants (1961) International Convention for the Protection of New Varieties of Plants adopted by the Diplomatic Conference on December 2, 1961 and Additional Act of November 10, 1972. Available at <<https://upovlex.upov.int/en/convention>>. Accessed on January 23, 2025.
- Wickham H (2009) **ggplot2: Elegant graphics for data analysis**. Springer, New York, 212p.