University of São Paulo
"Luiz de Queiroz" College of Agriculture

Proximal spectroscopy sensing for sugarcane quality prediction and spatial variability mapping

Lucas de Paula Corrêdo

Thesis presented to obtain the degree of Doctor in Science. Area: Agricultural Systems Engineering

Piracicaba
2021

Lucas de Paula Corrêdo
Agricultural Engineer

Proximal spectroscopy sensing for sugarcane quality prediction and spatial variability mapping

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **JOSÉ PAULO MOLIN**

Thesis presented to obtain the degree of Doctor in Science. Area: Agricultural Systems Engineering

Piracicaba
2021

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Corrêdo, Lucas de Paula

Proximal spectroscopy sensing for sugarcane quality prediction and spatial variability mapping / Lucas de Paula Corrêdo - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2021.

110 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Agricultura de precisão 2. Espectroscopia vis-NIR 3. Cana-de-açúcar 4. Qualidade tecnológica I. Título

to the greatest enthusiast of this stage, eternal friend and teacher,
to my father Jorge Luiz Corrêdo (*in memorian*), I dedicate.

ACKNOWLEDGEMENTS

To the One who, in his wisdom, incomprehensible to skeptical eyes, but clear to the meek, manifested himself in various ways, contributing to the construction of the journey, of learning, and of development: to God.

To the University of São Paulo, through the Luiz de Queiroz College of Agriculture, for the acceptance, support, and infrastructure provided.

To Capes, CNPq and mainly FAPESP, for providing the essential financial funds for the research accomplishment.

To the partnership of Iracema sugar mill (São Martinho group) and Spectral Solutions to provided infrastructure and operational support for the data acquisition.

To Professor José Paulo Molin, for providing the means, structures and teachings necessary for the learning process to happened in a concrete, fluid and natural way.

To the colleagues from the precision agriculture laboratory (LAP), Helizani, Leonardo, Chan, Tatiana, Martello, Tiago, Guia, Cássio, Orlando, and Ricardo, from Biosystems Engineering Department, Aureo, Chicão and Juarez, and all from gMAP, especially Firuleta.

To my college friends who, even at a distance, actively contributed to this journey: Vitor, Althoff, Abrahão and Matheus.

To my first friends in Piracicaba, Bruno, Ivanka, Antônio, Rogério, Dan and Vívian, for the good moments of friendship.

To my brother Pedro, for his companionship and fidelity.

To my family in Piracicaba, especially my in-laws Cláudio and Rosária, for the affection and the good moments during this stage.

To my beloved mother, Maria Gomes, my first teacher, in life and in mathematics, who, with few and strong words, formulated advice that served and serves as a basis for my constant evolution.

In special, to my lovely and admirable Ana Cláudia, which has contributed directly or indirectly In each stage of the construction of this study. Thank you for this period of construction together, for the kind advice, patience, examples and affection.

"Simplicity is the ultimate sophistication"

Leonardo Da Vinci

CONTENTS

RESUMO

Sensoriamento espectroscópico proximal para predição de qualidade de cana-de-açúcar e mapeamento da variabilidade espacial

Sensores para estimativa de atributos relacionados à qualidade de produtos agrícolas têm sido avaliados e implementados desde linhas de produção em setores industriais até iniciativas em âmbito de produção agrícola. Aplicações em campo buscam fonecer informações espacializadas ao longo do talhão relacionadas à qualidade do produto colhido. Com aplicações avançadas no setor industrial, a espectroscopia no infravermelho próximo (NIR) apresenta-se como a melhor alternativa devido à precisão das medições, realização de análises não destrutivas e rápidas, facilidade de operação, custo reduzido e sustentável. Além disso, a tendência de miniaturização dos equipamentos têm possibilitado maior flexibilidade de aplicações. Sensores NIR tem sido utlizados em iniciativas para identificar a variabilidade espacial da qualidade de uva, grãos e forragens. Entretanto, as aplicações para monitoramento da qualidade da cana-de-açúcar ainda são incipientes. O primeiro estudo deste documento (Capítulo 2) focou na avaliação da variabilidade espacial de atributos de qualidade de cana-de-açúcar em um campo comercial, a partir de amostras coletadas manualmente em campo e processadas para mensuração por espectroscopia NIR sob a forma desfibrada. Além disso, os mapas produzidos foram avaliados comparativamente a mapas produzidos a partir de resultados obtidos por métodos convencionais de análise em laboratório. O segundo estudo (Capítulo 3) buscou avaliar o potenial de predição de parâmetros de qualidade de cana-de-açúcar com espectroscopia NIR em diferentes níveis de preparo de amostra: sem preparo (toletes), com leituras em diferentes secções, cana desfibrada, e caldo cru. Além disso, buscou-se alcançar variabilidade nos modelos de calibração em função da variação climática com coletas realizadas em diferentes períodos ao longo de uma safra. Nessa etapa o experimento foi realizado em um laboratório de qualidade de uma usina, a fim de que análises NIR e convencionais fossem realizadas simultaneamente. Além disso, os modelos de calibração e predição de ambos os estudos foram desenvolvidos por análise multivariada, com regressões por mínimos quadrados parciais (PLSR), e avaliada a importância das bandas espectrais na predição de compostos orgânicos com base no reportado na literatura. O terceiro estudo (Capítulo 4) foi conduzido com um micro espectrômetro NIR embarcado no elevador de uma colhedora de cana-de-açúcar para coleta de informações em tempo real em três áreas de uma lavoura comercial. Para isso, foi construída uma plataforma de mensuração com fontes de iluminação externa. Durante a colheita, foram coletadas amostras diretamente da máquina, após leitura pelo sensor, para calibração e validação dos modelos. Além disso, foram retiradas subamostras de cana desfibrada, a partir das amostras processadas para análise por métodos convencionais, para mensuração em bancada, de forma similar ao realizado no primeiro estudo. Esses espectros foram utilizados para contrução de modelos de transferência de calibração para ajuste dos espectros coletados em tempo real na colhedora. Em seguida foram construídos modelos de calibração, validados e utilizados para estimativa de atributos de qualidade de cana-de-açúcar coletados em tempo real. Ao final da colheita, foram coletadas amostras de solo para avaliação de relações de causa e efeito com os dados de qualidade estimados. O método proposto permitiu a construção de modelos variográficos com dependência espacial e a espacialização dos dados de qualidade de cana-de-açúcar obtidos por mensurações com o sensor embarcado. Além disso, as relações de causa e efeito corroboraram com os resultados estimados, ao apresentarem relação entre os parâmetros de qualidade e atributos físicos do solo. Os resultados dessa pesquisa constituem uma nova etapa no direcionamento de pesquisas para viabilizar a obtenção de dados espacializados de qualidade de cana-de-açúcar por meio de sensores embarcados.

Palavras-chave: Espectroscopia de infravermelho próximo, Qualidade tecnológica, Sensor embarcado, Agricultura de precisão

ABSTRACT

Proximal spectroscopy sensing for sugarcane quality prediction and spatial variability mapping

Sensors for predicting attributes related to the quality of agricultural products have been evaluated and implemented from production lines in industrial sectors to some initiatives in the field of agricultural production. Field applications seek to provide spatial information along the field related to the quality of the harvested product. In constant evolution, and with advanced applications in the industrial sector, the near infrared spectroscopy (NIR) presents itself as the best alternative due to the precision of the equipment, fast analysis, non-destructive, easy operation, low cost and sustainable. In addition, the trend towards miniaturization of the equipment has enabled greater flexibility of applications. The production of thematic maps of product quality, associated to yield data, represents an advance for the practice of management with precision agriculture techniques for understanding spatial variability, cause and effect relationships during the crop cycle, and rational agronomic management of production inputs. NIR sensors have been used in initiatives to understand the spatial variability of grape, grain, and forage quality. However, applications for sugarcane quality monitoring are still incipient. The first study reported in this document (Chapter 2) focused on assessing the spatial variability of sugarcane quality attributes in a commercial field, from samples manually collected in the field and processed for measurement by NIR spectroscopy in defibrated form. In addition, the maps produced were evaluated against maps produced from results obtained by conventional laboratory analysis methods. The second study (Chapter 3) sought to evaluate the potential for predicting sugarcane quality parameters with NIR spectroscopy at different levels of sample preparation: no preparation (stalks), with measurements in different sections, defibrated cane, and raw juice. In addition, we sought to achieve variability in the calibration models as a function of climatic variation with sample collections performed in different periods throughout a harvest. In this step, the experiment was performed in a quality laboratory of a mill, so that NIR and conventional analyses could be performed simultaneously. The calibration and prediction models for both studies were developed by multivariate analysis, with partial least squares regressions (PLSR), and the importance of spectral bands in the prediction of organic compounds was evaluated based on what has been reported in the literature. The third and last study (Chapter 4) was conducted with a on-board micro spectrometer in the elevator of a sugarcane harvester to collect real-time information in three areas of a commercial crop. For this, a measurement platform was built with external light sources. During the harvest, samples were collected directly from the machine, after being read by the sensor, for model calibration and validation. In addition, sub-samples of defibrated cane were taken from the samples processed for analysis by conventional methods, for bench measurement, similar to the first study. These spectra were used to build calibration transfer models to fit the spectra collected in real time at the harvester. Calibration models were then built, validated, and used to estimate sugarcane quality attributes collected in real time. At the end of the harvest, soil samples were collected to evaluate cause and effect relationships with the estimated quality data. The proposed method allowed the construction of variogram models with spatial dependence and the spatialization of sugarcane quality data obtained by measurements with the on-board sensor. Moreover, the cause and effect relationships corroborated the estimated results with previously reported by other studies, by presenting a relationship between quality parameters and soil physical attributes. The results of this research constitute a new stage in the direction of research to make it feasible to obtain spatialized data of sugarcane quality by means of on-board sensors.

Keywords: Near-infrared spectroscopy, Technological quality, On-board sensor, Precision agriculture

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Knowledge about the complex biotic and abiotic interactions that govern the development of agricultural crops drives the development of technologies and management strategies seeking optimization of product results, in yield and quality, economy in the use of inputs and resources, and reduction of environmental impacts. In this way, precision agriculture (PA) is a management strategy that allows to achieve these objectives from the gathering, processing and analysis of data from the crops that characterize their inherent temporal and spatial variability. Thus, from the resulting information allows management according to the variability estimated to improve efficiency in resource use, productivity, quality, profitability and sustainability of agricultural production (ISPA, 2021).

On a first step, the application of PA management strategies requires an accurate assessment of fine-resolution spatial variability of some crop or soil attribute (CASTRIGNANÒ *et al.*, 2018). The use of proximal sensing techniques assists in gathering relevant data. Proximal sensing (PS) is the use of sensors combined with a positioning device (GNSS) to collect soil or plant data through signals with physical or chemical measurements that can be related to the properties of interest (PALLOTTINO *et al.*, 2019). Then, it helps in the collection of high-density data throughout the crop.

High-density data acquisition is the first step for characterizing crop spatial variability (MALDANER; MOLIN, 2020). There are many factors, such as weather, soil and plants attributes, which affect crop development and could provide informations to perform data layers to interactions analysis. Among the available information, yield constituting the noblest data layer for handling spatial variability (VEGA *et al.*, 2019), because it characterizes the result of all management strategies adopted during the crop cycle, combined with environmental factors, some inherent factors of the field and corresponds to the economic result of the activity (COLAÇO *et al.*, 2021). Yield mapping allows to perform variable rate inputs application based on the site-specific information of export nutrients by the crop (MOLIN *et al.*, 2020), for example.

While studies for many agricultural crops have focused on variable-rate input application strategies or yield measurement methods, on the other hand, for some crops the interest has advanced on selective harvesting or product quality-based management (BRAMLEY, 2012). Quality sensing is desired for most agricultural commodities at different stages of the production chain (RUIZ-ALTISENT *et al.*, 2010). Post-harvest quality control is widely applied at the industrial level (ZAREEF *et al.*, 2020), but resources applieds to in-field level is still lacking. Site-specific management based on final product quality tends to be beneficial especially for higher value-added products.

For crops predominant in temperate regions, studies related to the application of PA techniques for site-specific management have been employed with greater vigor for grapevine management (SANTESTEBAN, 2019). The monitoring of spatial and temporal variability quality-related attributes, such as total soluble solids, pH and phenolic content, allows the identification vineyard zones with potential for the production of higher quality wine, and consequently, higher value (BALUJA *et al.*, 2013, 2012b). In this way, studies have been performed evaluating the spatial variability monitoring of these attributes by means of different sensing types as orbital, aerial, and proximal sensing (BALUJA *et al.*, 2012a; BORGOGNO-MONDINO *et al.*, 2018; FERNÁNDEZ-NOVALES *et al.*, 2019; MESAS-CARRASCOSA *et al.*, 2020).

The main tropical crops whose added value is associated with product quality and have been the subject of PA studies are coffee and sugarcane. Some studies present criteria and the basis for designing quality-based management of the spatial variability of these products (CORRÊDO *et al.*, 2021; SILVA *et al.*, 2014). However, they are still incipient regarding the essential automation of this information in the field to obtain high-density data for

better characterization of this spatial variability (RUIZ-ALTISENT *et al.*, 2010). In this way, some studies have sought to development and adaptation equipments for PS of agricultural products for obtaining the high-density data in the field (WALSH *et al.*, 2020).

Proximal sensors based on different physical principles have been evaluated for the characterization of chemical compounds related to the quality of various agricultural products, such as devices based on image processing technology, acoustic, ultrasound, impact test methods, electronic, and spectroscopy (ABASI *et al.*, 2018). Based on spectroscopy, especially on the visible and near-infrared ranges, proximal sensing has been more relevant for this objective for to allow easy, fast, accurate, non-destructive, environmental friendly and inexpensive quality assessments (CORTÉS *et al.*, 2019). The technique is based on the interaction of the eletromagnetic radiation with the matter. Wavelengths in the visible (380-700 nm) and near infrared (NIR, 780-2500 nm) range have been widely used, including in commercial equipment for post-harvest analysis of agricultural products (EL-MESERY; MAO; ABOMOHRA, 2019; WALSH; MCGLONE; HAN, 2020).

Visible and near-infrared (VNIR) and only near-infrared (NIR) spectroscopy is a vibrational spectroscopy analytical technique from which the radiation interaction occurs mainly by exciting overtones, combinations, and resonances of fundamental vibrational modes of molecular functional groups containing heavy atoms as C, N, O and S attached to a hydrogen atom, the main atoms and molecules present in organic compounds (PASQUINI, 2018). Sensors based on this physic principle have been widely adapted for in-line application in industrial food production, also for diverse matrices such as intact solid samples, free-flowing solids, pasty, and fluid samples (POREP; KAMMERER; CARLE, 2015). In sugar mills it is used in routine analyses to measure the apparent sucrose (Pol) and soluble solids content (Brix) of the juice (SEXTON *et al.*, 2020). Moreover, it has also been widely used in research to monitor the quality of moving products (POREP; KAMMERER; CARLE, 2015), as for Brix of apples (XU; XU; *et al.*, 2019), oil content, moisture and free acidity of olive fruits (SALGUERO-CHAPARRO *et al.*, 2013, 2012), Brix of watermelons (JIE *et al.*, 2014), for example. The main overcome to achieve an desired accuracy for in-line measurements (*in situ*), i.e. when the sample is analised within the running production line (CORTÉS *et al.*, 2019), are the sample moviment, temperature fluctuations and the sensor-target distance (POREP; KAMMERER; CARLE, 2015; WALSH *et al.*, 2020).

In the field, the motivation for quality monitoring has been to identify the spatial variability of parameters of interest related to product quality, to provide an additional layer to yield for site-specific crop management (CORRÊDO *et al.*, 2020). To achieve this goal, the main proposal presented in several works has been to embed sensors, spetially VNIR or NIR sensors, in agricultural harvesters (DIGMAN; SHINNERS, 2008; LONG; ENGEL; SIEMENS, 2008; LONG; MCCALLUM, 2020; NAWI; CHEN; JENSEN, 2014). However, some basic technical aspects are required to implement an embedded NIR sensor system as the incompatibility with any intense sample processing, the huge variations of temperature at harvest, and presence of impurities and undesirable material (MONTES; PAUL, 2008). For measuring grain constituents, there are initiatives in advanced stages, with on-board sensor evaluations (LONG; ENGEL; SIEMENS, 2008; LONG; MCCALLUM, 2020). On the other hand, the measurement of constituents in forages already has on-board sensor solutions in harvesters based on NIR spectroscopy (MONTES; WILLIAMS; PAUL, 2008).

Some recent studies have evaluated NIR spectroscopy-based technology for embedded PS of sugarcane quality (NAWI *et al.*, 2013; NAWI; CHEN; JENSEN, 2014, 2013; PHETPAN; UDOMPETAIKUL; SIRISOMBOON, 2018; UDOMPETAIKUL; PHETPAN; SIRISOMBOON, 2021). These studies have proven the potential for predicting sugarcane quality attributes in off-line (in bench) measurements based on NIR spectroscopy

(NAWI *et al.*, 2013), and obtained promising results for measurements with the sample in motion (PHETPAN; UDOMPETAIKUL; SIRISOMBOON, 2018). However, unlike research using proximal sensors based on NIR spectroscopy to measure forage (DIGMAN; SHINNERS, 2008) and grain (LONG, Dan S.; MCCALLUM, 2020) quality in real time at the harvesters, this technique has not yet been evaluated for sugarcane quality measurements in the field. It is essential to evaluate this technique for on-the-go measurements under hostile field conditions (NAWI; CHEN; JENSEN, 2014).

The main motivation to proceded in-line measurements of on-the-go sugarcane quality, i.e., by on-board sensing, is based on the understanding of the spatial variability of quality attributes along the crop field. This information could help in industrial and logistical planning, based on the quality of the recently harvested product, and in agronomic management, to identify cause-effect relationships between product quality and edaphic variables, and varieties and input management according to the potential of different field zones. Thus, the hypothesis of this study is that it is possible, by means of proximal sensing based on vis-NIR spectrometry, to measure sugarcane quality data in the field, from unprepared samples, in order to map the spatial variability of quality in sugarcane crops. To address this goal, this thesis presents three chapters given as complete manuscrips that respond the hypothesis presented. Firstly, it was investigate the spatial dependence of sugarcane quality attributes along a commercial field, with model-based predictions from NIR spectroscopy (Chapter 2). In addition, different levels of minimum processing of sugarcane samples to predict quality attributes were evaluated by means of robust calibration models that encompass the variability existing throughout a harvest season (Chapter 3). Finally, it was evaluate a method for measuring sugarcane with an on-board NIR spectroscopy sensor in a harvester, with real-time measurements, and possible relationships with soil attributes that justify site-specific management based on product quality (Chapter 4).

References

ABASI, S.; MINAEI, S.; JAMSHIDI, B.; FATHI, D. Dedicated non-destructive devices for food quality measurement: A review. **Trends in Food Science and Technology**, vol. 78, p. 197–205, 1 Aug. 2018. https://doi.org/10.1016/j.tifs.2018.05.009.

BALUJA, J.; DIAGO, M. P.; GOOVAERTS, P.; TARDAGUILA, J. Assessment of the spatial variability of anthocyanins in grapes using a fluorescence sensor: Relationships with vine vigour and yield. **Precision Agriculture**, vol. 13, no. 4, p. 457–472, 17 Mar. 2012a. DOI 10.1007/s11119-012-9261-x. Available at: https://link.springer.com/article/10.1007/s11119-012-9261-x. Accessed on: 17 Aug. 2021.

BALUJA, J.; DIAGO, M. P.; GOOVAERTS, P.; TARDAGUILA, J. Spatio-temporal dynamics of grape anthocyanin accumulation in a Tempranillo vineyard monitored by proximal sensing. **Australian Journal of Grape and Wine Research**, vol. 18, no. 2, p. 173–182, 1 Jun. 2012b. DOI 10.1111/j.1755-0238.2012.00186.x. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1755-0238.2012.00186.x. Accessed on: 17 Aug. 2021.

BALUJA, J.; TARDAGUILA, J.; AYESTARAN, B.; DIAGO, M. P. Spatial variability of grape composition in a Tempranillo (Vitis vinifera L.) vineyard over a 3-year survey. **Precision Agriculture**, vol. 14, no. 1, p. 40–58, 2013. https://doi.org/10.1007/s11119-012-9282-5.

BORGOGNO-MONDINO, E.; LESSIO, A.; TARRICONE, L.; NOVELLO, V.; DE PALMA, L. A comparison between multispectral aerial and satellite imagery in precision viticulture. **Precision Agriculture**, vol. 19, no. 2, p. 195–217, 2018. DOI 10.1007/s11119-017-9510-0. Available at: https://doi.org/10.1007/s11119-017-9510-0. Accessed on: 12 Aug. 2019.

BRAMLEY, R. Mixed fortunes in crop quality sensing. 2012. **Australasian Precision Agriculture Symposium** [...]. [*S. l.*: *s. n.*], 2012. p. 4–7.

CASTRIGNANÒ, A.; BUTTAFUOCO, G.; QUARTO, R.; PARISI, D.; VISCARRA ROSSEL, R. A.; TERRIBILE, F.; LANGELLA, G.; VENEZIA, A. A geostatistical sensor data fusion approach for delineating homogeneous management zones in Precision Agriculture. **Catena**, vol. 167, p. 293–304, Aug. 2018. DOI 10.1016/j.catena.2018.05.011. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0341816218301772. Accessed on: 18 Jul. 2019.

COLAÇO, A. F.; RICHETTI, J.; BRAMLEY, R. G. V.; LAWES, R. A. How will the next-generation of sensor-based decision systems look in the context of intelligent agriculture? A case-study. **Field Crops Research**, vol. 270, p. 108205, 1 Aug. 2021. DOI 10.1016/j.fcr.2021.108205. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0378429021001519. Accessed on: 19 Jun. 2021.

CORRÊDO, L. de P.; CANATA, T. F.; MALDANER, L. F.; DE LIMA, J. de J. A.; MOLIN, J. P. Sugarcane Harvester for In-field Data Collection: State of the Art, Its Applicability and Future Perspectives. **Sugar Tech**, vol. 22, no. 4, p. 1–14, 10 Aug. 2020. DOI 10.1007/s12355-020-00874-3. Available at: http://link.springer.com/10.1007/s12355-020-00874-3. Accessed on: 10 Aug. 2020.

CORRÊDO, L. P.; WEI, M. C. F.; FERRAZ, M. N.; MOLIN, J. P. Near-infrared spectroscopy as a tool for monitoring the spatial variability of sugarcane quality in the fields. **Biosystems Engineering**, vol. 206, p. 150–161, 1 Jun. 2021. DOI 10.1016/j.biosystemseng.2021.04.001. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1537511021000854. Accessed on: 28 Apr. 2021.

CORTÉS, V.; BLASCO, J.; ALEIXOS, N.; CUBERO, S.; TALENS, P. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. **Trends in Food Science and Technology**, vol. 85, p. 138–148, 1 Mar. 2019. https://doi.org/10.1016/j.tifs.2019.01.015.

DIGMAN, M. F.; SHINNERS, K. J. Real-Time Moisture Measurement On A Forage Harvester Using Near-Infrared Reflectance Spectroscopy. **Transactions of the ASAE**, vol. 51, no. 5, p. 1801–1810, 2008. .

EL-MESERY, H. S.; MAO, H.; ABOMOHRA, A. E. F. Applications of non-destructive technologies for agricultural and food products quality inspection. **Sensors (Switzerland)**, vol. 19, no. 4, p. 846, 18 Feb. 2019. DOI 10.3390/s19040846. Available at: https://www.mdpi.com/1424-8220/19/4/846/htm. Accessed on: 18 Aug. 2021.

FERNÁNDEZ-NOVALES, J.; TARDÁGUILA, J.; GUTIÉRREZ, S.; PAZ DIAGO, M. On-The-Go VIS + SW - NIR Spectroscopy as a Reliable Monitoring Tool for Grape Composition within the Vineyard. **Molecules (Basel, Switzerland)**, vol. 24, no. 15, p. 2795, 31 Jul. 2019. DOI 10.3390/molecules24152795. Available at: www.mdpi.com/journal/molecules. Accessed on: 25 May 2021.

JIE, D.; XIE, L.; RAO, X.; YING, Y. Using visible and near infrared diffuse transmittance technique to predict soluble solids content of watermelon in an on-line detection system. **Postharvest Biology and Technology**, vol. 90, p. 1–6, 2014. DOI 10.1016/j.postharvbio.2013.11.009. Available at: http://dx.doi.org/10.1016/j.postharvbio.2013.11.009. Accessed on: 9 Jan. 2019.

LONG, D. S.; ENGEL, R. E.; SIEMENS, M. C. Measuring grain protein concentration with in-line near infrared reflectance spectroscopy. **Agronomy Journal**, vol. 100, no. 2, p. 247–252, 2008. https://doi.org/10.2134/agronj2007.0052.

LONG, Dan S.; MCCALLUM, J. D. Adapting a relatively low-cost reflectance spectrometer for on-combine sensing of grain protein concentration. **Computers and Electronics in Agriculture**, vol. 174, p. 105467, 1 Jul. 2020. https://doi.org/10.1016/j.compag.2020.105467.

MALDANER, L. F.; MOLIN, J. P. Data processing within rows for sugarcane yield mapping. **Scientia Agricola**, vol. 77, no. 5, p. 2020, 2020. DOI 10.1590/1678-992x-2018-0391. Available at: http://dx.doi.org/10.1590/1678-992X-2018-0391. Accessed on: 24 Dec. 2019.

MESAS-CARRASCOSA, F. J.; DE CASTRO, A. I.; TORRES-SÁNCHEZ, J.; TRIVIÑO-TARRADAS, P.; JIMÉNEZ-BRENES, F. M.; GARCÍA-FERRER, A.; LÓPEZ-GRANADOS, F. Classification of 3D point clouds using color vegetation indices for precision viticulture and digitizing applications. **Remote Sensing**, vol. 12, no. 2, p. 317, 18 Jan. 2020. DOI 10.3390/rs12020317. Available at: https://www.mdpi.com/2072-4292/12/2/317/htm. Accessed on: 17 Aug. 2021.

MOLIN, J. P.; BAZAME, H. C.; MALDANER, L.; CORREDO, L. D. P.; MARTELLO, M.; CANATA, T. F.; DE PAULA CORREDO, L.; MARTELLO, M.; CANATA, T. F. Precision agriculture and the digital contributions for site-specific management of the fields. **Revista Ciência Agronômica**, vol. 51, no. Special Agriculture 4.0, p. 1–10, 2020. DOI 10.5935/1806-6690.20200088. Available at: http://ccarevista.ufc.br/seer/index.php/ccarevista/article/view/7720. Accessed on: 5 Aug. 2021.

MONTES, J. M.; PAUL, C. Near Infrared Spectroscopy on Agricultural Harvesters: Technical Aspects. **NIR news**, vol. 19, no. 1, p. 10–12, 2008. DOI 10.1255/nirn.1059. Available at: http://nir2007.com/abstracts/sys-. Accessed on: 27 Mar. 2021.

NAWI, N. M.; CHEN, G.; JENSEN, T. In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: a review. **Precision Agriculture**, vol. 15, no. 6, p. 684–703, 2014. https://doi.org/10.1007/s11119-014-9362-9.

NAWI, N. M.; CHEN, G.; JENSEN, T. Visible and shortwave near infrared spectroscopy for predicting sugar content of sugarcane based on a cross-sectional scanning method. **Journal of Near Infrared Spectroscopy**, vol. 21, no. 4, p. 289–297, 2013. https://doi.org/10.1255/jnirs.1060.

NAWI, N. M.; CHEN, G.; JENSEN, T.; MEHDIZADEH, S. A. Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared. **Biosystems Engineering**, vol. 115, p. 154–161, 2013. DOI 10.1016/j.biosystemseng.2013.03.005. Available at: http://dx.doi.org/10.1016/j.biosystemseng.2013.03.005.

PALLOTTINO, F.; ANTONUCCI, F.; COSTA, C.; BISAGLIA, C.; FIGORILLI, S.; MENESATTI, P. Optoelectronic proximal sensing vehicle-mounted technologies in precision agriculture: A review. **Computers and Electronics in Agriculture**, vol. 162, p. 859–873, 1 Jul. 2019. https://doi.org/10.1016/j.compag.2019.05.034.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, vol. 1026, p. 8–36, 2018. https://doi.org/10.1016/j.aca.2018.04.004.

PAUL, C.; MONTES, J. M.; WILLIAMS, P.; PAUL, C. Near infrared spectroscopy on agricultural harvesters: the background to commercial development. **NIR news**, vol. 19, no. 8, p. 8, 2008. https://doi.org/10.1255/nirn.1103.

PHETPAN, K.; UDOMPETAIKUL, V.; SIRISOMBOON, P. An online visible and near-infrared spectroscopic technique for the real-time evaluation of the soluble solids content of sugarcane billets on an elevator conveyor. **Computers and Electronics in Agriculture**, vol. 154, p. 460–466, Nov. 2018. DOI 10.1016/j.compag.2018.09.033. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0168169918302096. Accessed on: 7 Jan. 2019.

POREP, J. U.; KAMMERER, D. R.; CARLE, R. On-line application of near infrared (NIR) spectroscopy in food production. **Trends in Food Science and Technology**, vol. 46, no. 2, p. 211–230, 1 Dec. 2015. https://doi.org/10.1016/j.tifs.2015.10.002.

RUIZ-ALTISENT, M.; RUIZ-GARCIA, L.; MOREDA, G. P.; LU, R.; HERNANDEZ-SANCHEZ, N.; CORREA, E. C.; DIEZMA, B.; NICOLAÏ, B.; GARCÍA-RAMOS, J. Sensors for product characterization and quality of specialty crops-A review. **Computers and Electronics in Agriculture**, vol. 74, no. 2, p. 176–194, 1 Nov. 2010. https://doi.org/10.1016/j.compag.2010.07.002.

SALGUERO-CHAPARRO, L.; BAETEN, V.; ABBAS, O.; PEÑA-RODRÍGUEZ, F. On-line analysis of intact olive fruits by vis-NIR spectroscopy: Optimisation of the acquisition parameters. **Journal of Food Engineering**, vol. 112, no. 3, p. 152–157, 1 Oct. 2012. https://doi.org/10.1016/j.jfoodeng.2012.03.034.

SALGUERO-CHAPARRO, L.; BAETEN, V.; FERNÁNDEZ-PIERNA, J. A.; PEÑA-RODRÍGUEZ, F. Near infrared spectroscopy (NIRS) for on-line determination of quality parameters in intact olives. **Food Chemistry**, vol. 139, no. 1–4, p. 1121–1126, 15 Aug. 2013. https://doi.org/10.1016/j.foodchem.2013.01.002.

SANTESTEBAN, L. G. Precision viticulture and advanced analytics. A short review. **Food Chemistry**, vol. 279, p. 58–62, 1 May 2019. https://doi.org/10.1016/j.foodchem.2018.11.140.

SEXTON, J.; EVERINGHAM, Y.; DONALD, D.; STAUNTON, S.; WHITE, R. Investigating the identification of atypical sugarcane using NIR analysis of online mill data. **Computers and Electronics in Agriculture**, vol. 168, 1 Jan. 2020. DOI 10.1016/j.compag.2019.105111. Available at: https://doi.org/10.1016/j.compag.2019.105111. Accessed on: 3 Jan. 2021.

SILVA, S. de A.; QUEIROZ, D. M.; PINTO, F. A. C.; SANTOS, N. T. Coffee quality and its relationship with Brix degree and colorimetric information of coffee cherries. **Precision Agriculture**, vol. 15, no. 5, p. 543–554, 23 Feb. 2014. DOI 10.1007/s11119-014-9352-y. Available at: https://link.springer.com/article/10.1007/s11119-014-9352-y. Accessed on: 17 Aug. 2021.

UDOMPETAIKUL, V.; PHETPAN, K.; SIRISOMBOON, P. Development of the partial least-squares model to determine the soluble solids content of sugarcane billets on an elevator conveyor. **Measurement**, vol. 167, p. 107898, 1 Jan. 2021. DOI 10.1016/j.measurement.2020.107898. Available at: https://linkinghub.elsevier.com/retrieve/pii/S026322412030436X. Accessed on: 17 Feb. 2021.

VEGA, A.; CÓRDOBA, M.; CASTRO-FRANCO, M.; BALZARINI, M. Protocol for automating error removal from yield maps. **Precision Agriculture**, vol. 20, no. 5, p. 1030–1044, 10 Jan. 2019. DOI 10.1007/s11119-018-09632-8. Available at: https://doi.org/10.1007/s11119-018-09632-8. Accessed on: 28 Aug. 2020.

WALSH, K. B.; MCGLONE, V. A.; HAN, D. H. The uses of near infra-red spectroscopy in postharvest decision support: A review. **Postharvest Biology and Technology**, vol. 163, p. 111139, 1 May 2020. https://doi.org/10.1016/j.postharvbio.2020.111139.

WALSH, Kerry B.; BLASCO, J.; ZUDE-SASSE, M.; SUN, X. Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. **Postharvest Biology and Technology**, vol. 168, p. 111246, 1 Oct. 2020. https://doi.org/10.1016/j.postharvbio.2020.111246.

XU, X.; XU, H.; XIE, L.; YING, Y. Effect of measurement position on prediction of apple soluble solids content (SSC) by an on-line near-infrared (NIR) system. **Journal of Food Measurement and Characterization**, vol. 13, no. 1, p. 506–512, 15 Mar. 2019. DOI 10.1007/s11694-018-9964-4. Available at: https://doi.org/10.1007/s11694-018-9964-4. Accessed on: 1 Jun. 2020.

ZAREEF, M.; CHEN, Q.; HASSAN, M. M.; ARSLAN, M.; HASHIM, M. M.; AHMAD, W.; KUTSANEDZIE, F. Y. H.; AGYEKUM, A. A. An Overview on the Applications of Typical Non-linear Algorithms Coupled With NIR Spectroscopy in Food Analysis. **Food Engineering Reviews**, vol. 12, no. 2, p. 173–190, 2020. https://doi.org/10.1007/s12393-020-09210-7.

2. NEAR-INFRARED SPECTROSCOPY AS A TOOL FOR MONITORING THE SPATIAL VARIABILITY OF SUGARCANE QUALITY IN THE FIELDS

This section was published in Biosystems Engineering Journal (CORRÊDO, L. P.; WEI, M. C. F.; FERRAZ, M. N.; MOLIN, J. P. Near-infrared spectroscopy as a tool for monitoring the spatial variability of sugarcane quality in the fields. **Biosystems Engineering**, vol. 206, p. 150–161, 1 Jun. 2021. DOI 10.1016/j.biosystemseng.2021.04.001).

ABSTRACT

It is known that Near-infrared spectroscopy (NIRS) is a reliable technique used in industrial laboratories to measure sugarcane quality. However, its use as a proximal sensing technology for monitoring the spatial variability of attributes in the fields has not yet been evaluated. The aim of this research was to examine the potential of NIRS for predicting and mapping Brix, Pol and Fibre content in a commercial sugarcane field. The quality attributes models were adjusted considering the spectral reflectance from the 1100-1800 nm wavelengths by using partial least squares regressions (PLSR). A total of 350 samples were collected in a sugar mill laboratory for calibration and cross-validation models development. For the external validation, 91 georeferenced samples were obtained from a commercial field. The results indicated that the developed models are capable of predicting Brix and Pol, with a coefficient of determination (R^2_P) of 0.71 for both parameters, and with a root mean square error of prediction ($RMSE_P$) of 0.80 % and 0.58 %, respectively. In contrast, the results for Fibre were unsatisfactory (R^2_P of 0.24 and $RMSE_P$ of 1.15 %). Predicted values showed spatial dependence of the sugarcane quality attributes. Predicted and observed values of Brix and Pol presented a coefficient of correlation of 0.85. Results showed that NIRS has potential to be applied as a proximal sensing method supporting crop management based on the spatial variability of the quality attributes.

Keywords: Precision agriculture; NIR spectroscopy; Chemometrics; *Saccharum officinarum*; Portable spectrometer.

2.1. Introduction

In-field yield monitoring of sugarcane fields represents the main approach for managing the spatial and temporal variability of the crop (MALDANER; MOLIN, 2020). However, Brazilian sugarcane producers rarely use yield monitors (SANCHES; GRAZIANO MAGALHÃES; JUNQUEIRA FRANCO, 2019) despite the existence of them. In addition to the importance of yield, quality analysis is also important for sugarcane crops since their combination is used to calculate the payment to sugarcane growers (NAWI; CHEN; JENSEN, 2014).

Pricing and trading between growers and sugar mills is performed as a function of the total recoverable sugar (TRS) produced (PHETPAN; UDOMPETAIKUL; SIRISOMBOON, 2018) and is determined by the average result of a sampling obtained at the sugar mill. The TRS is calculated by attributes related to the sugarcane composition which should meet the quality standards of the industrial process. These attributes are the percentage of water-insoluble matter of cane and bagasse (Fibre), soluble solids content (Brix), apparent sucrose in the juice, measured by the optical activity which causes polarised light to be rotated (Pol). Purity, Pol of cane (PC) and reducing sugars (RS), are quality attributes calculated based on the former parameters (Brix, Pol and Fibre) by empirical equations (RODRIGUES; MAGALHÃES; FRANCO, 2013). Then, the TRS value is obtained as a function of the estimated PC and RS of cane. In Brazil, these attributes are determined by conventional laboratory

analysis described by the National Council of Sugarcane Producers of São Paulo State (CONSECANA, 2015), which are in agreement with the international rules from the International Commission for Uniform Methods of Sugar Analysis (ICUMSA).

The analysis to evaluate sugarcane quality at the field level aiming to manage its variability must be different from the conventional laboratory routine. The conventional analysis includes high-cost, time-consuming and environmental unfriendly methods (CORRÊDO *et al.*, 2020). Therefore, the management of spatial variability of sugarcane quality to assist in decision making must be from a feasible method to obtain quality results of the raw material, i.e. development of a proximal sensing method that rapidly measures product quality with high accuracy, requires minimal sample preparation and is environmentally friendly (SANSEECHAN *et al.*, 2018).

Near-infrared spectroscopy (NIRS) has been used to determine the quality of many agricultural products (CORTÉS *et al.*, 2019), due the possibility in detecting signals from most of the major structures and functional groups of organic compounds (PASQUINI, 2018). The interaction between electromagnetic radiation corresponding to NIRS, in the wavelength range from 750 nm to 2500 nm, and the sample can be reflected, transmitted, or absorbed (OSBORNE, 2006). As a result of the interaction, the NIR spectrum can be obtained and used to extract information about the chemical composition of the sample (WANG *et al.*, 2015). Then, chemometric methods are applied to interpret the signal, relating them with other attributes of interest and to extract the responses (PASQUINI, 2018).

Several studies have been carried out using NIRS to predict sugarcane quality attributes (MARAPHUM *et al.*, 2018; NAWI *et al.*, 2013; NAWI; CHEN; JENSEN, 2013; PHUPHAPHUD; SAENGPRACHATANARUG; POSOM; MARAPHUM; *et al.*, 2019; SOROL *et al.*, 2010; TAIRA; UENO; KAWAMITSU, 2010; VALDERRAMA; BRAGA; POPPI, 2007), presenting promising indicative results. Phetpan, Udompetaikul and Sirisomboon (2018) used an approach to measure quality attributes by NIRS directly in the elevator conveyor of a sugarcane harvester. For this purpose, they developed a sugarcane harvester simulated elevator to measure the content of Brix in motion on the elevator conveyor resulting in a coefficient of determination of prediction (R^2_P) of 0.785 and root mean square error of prediction (RMSEP) of 0.30.

Although several results indicate the NIRS as the most viable technology for predicting the quality of sugarcane, including its potential as a PA tool, there are lack of reports available analysing georeferenced samples obtained in the field to evaluate the viability of using NIRS to monitor the spatial variability of sugarcane quality attributes. Thus, the objective of this study was to evaluate the potential of NIRS to predict sugarcane quality attributes (Fibre, Brix and Pol) and to characterise their spatial variability in a commercial field.

2.2. Material and Methods

This study was conducted in two stages. The first refers to the development of calibration models to predict sugarcane quality attributes. The second stage is the application of the fitted model to georeferenced sampled data obtained in a commercial field to evaluate its spatial variability by means of NIRS.

2.2.1. Laboratory sampling and conventional analysis

Calibration data set was composed of samples obtained from a sugar mill quality laboratory located in Pradópolis, SP, Brazil, in July of 2018. The conventional analysis protocol used in this research followed the one used by the sugar mill laboratory quality standard analysis, as proposed by the National Council of Sugar Cane Producers of São Paulo State (CONSECANA, 2015). These protocols agree with the international rules from International Commission for Uniform Methods of Sugar Analysis.

Conventional sugarcane quality analysis relies on a sampling procedure at the sugar mill. From each truck carrying harvested sugarcane billets, a sample of about 10 kg was extracted randomly from a truckload on the arrival of the harvested cargo at the sugar mill through an oblique probe and labelled (Figure 1a and 1b). The samples were milled in a mechanical knife crusher and homogenised in a concrete mixer. Then, 500 g and 50 g samples of the homogenised mixture of defibrated cane were collected to proceed with the conventional and NIRS analysis, respectively.

For the conventional sugarcane attributes analysis, each 500 g sample was pressed in a hydraulic press under constant pressure, at 24.5 MPa for 60 s to obtain the raw juice. An automatic refractometer (RX-5000a, ATAGO Co Ltd, Tokyo, Japan), with maximum resolution of 0.1 ºBrix, was used to determine the soluble solids content in the raw juice for each sample. After that, 14 g of a mixture composed of 4:2:1 pro- portion of Celite (mineral filtering agent), aluminium chloride and calcium hydroxide, respectively, was added to 200 mL of raw juice and homogenised through agitation process by a magnetic stirrer until the solution was well-mixed. Then, the solution was filtered through filter paper to obtain a trans- parent solution (Fig. 1c). A volume of 70 mL of this filtered solution was used for saccharimetric reading, measured by a digital polarimeter (Schmidt þ Haensch, Polartronic NHZ 8, Berlin, Germany). The result was obtained as percentage of apparent sucrose in the juice (Pol). The remaining defibrated cane from the press step (without juice), named wet bagasse, was transferred to metal baskets, with holes at the base. Then, the wet bagasse was weighed on a semi-analytical balance to determine the weight of the wet bagasse. The balance was previously tared to discount the mass of the baskets. Then, samples were held in a forced air circulation drying equipment at a temperature of 105 ºC, until constant weight. The dryer samples were weighed, and the Fibre content (insoluble solids) was determined.

2.2.2. NIR measurements

Spectral absorbance was measured with a portable spectrometer AgriNIR (Dynamica Generale S.p.A., Poggio Rusco, Mantova, Italy), mostly used for forage analysis, operating in the spectral range of 1100–1800 nm with spectral resolution of 10 nm.

The 50 g samples of defibrated cane were transferred to a sample holder (Figure 1d). The sample holder has a metal plate which compressed the sample on a glass surface reducing the number of pores and standardising the surface of the matrix to be scanned. Then, the sample holder was inserted in the dark chamber of the equipment (Figure 1e) with a halogen light source inside for spectral measurements of the sample surface through the glass surface. The equipment results are composed of an average of three spectral readings. In addition, the procedure was repeated five times for each sample.

All spectral readings and conventional measurements were conducted at the same time at the sugarcane quality laboratory under controlled temperature (20 ºC ± 5 ºC). Thus, sugar degradation through microbial activities and molecular vibrations caused by temperature fluctuations during measurements were considered negligible.

The sample characteristic spectrum was composed of the average of readings to be associated with the result obtained from the conventional analysis. Hence, 350 samples of different cultivars of sugarcane were obtained for the calibration step.

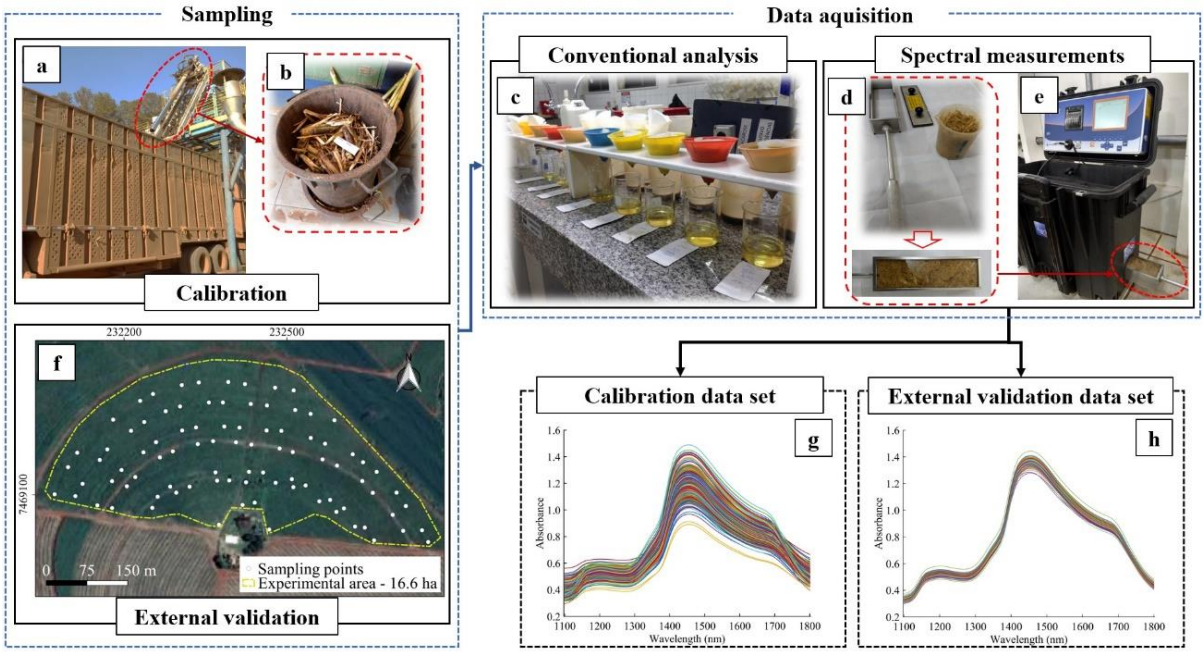

Figure 1. Process of sample obtaining to calibration (in the sugarcane quality laboratory) by an oblique probe (a) pulling out sample in the cargo truck (b); conventional laboratory analysis (c) after sample prepare; preparation of defibrated sample (d) to spectral measurements by NIRS (e); sampling points in the field to external validation (f); spectral data set of calibration samples (g) and external validation samples (h).

2.2.3. Study field and external validation data set

The validation data set used in this study was obtained from the spatial-analysis database of the Precision Agriculture Laboratory at the University of Sao Paulo. This data set was composed of in-field database obtained from 16.6 ha of cultivar RB855453 harvested in the 2015 season. This cultivar is characterised as having high yield and sucrose content, and medium Fibre content (RIDESA, 2015). The georeferenced samples were randomly collected at the edges of each harvested line (Figure 1f). A total of 91 samples were collected in the sugarcane field (approximately 5.5 samples ha^{-1}) composed of four stalks for each sampling. The whole stalks were manually cut at ground level and the growing point, and the leaves were removed. In-field samples were taken to the laboratory and milled to break the Fibre, homogenised and pressed, to replicate the sugar mill conventional laboratory procedure. All these samples were submitted to the conventional and NIRS analysis in the same day. Also, both procedures were performed in the same way as for samples in calibration step, as describe in topics 2.1 and 2.2.

2.2.4. Chemometrics

The multivariate analysis techniques applied to chemical compounds prediction are widely used in spectroscopic data analysis (MEHMOOD; AHMED, 2016). Generally, this procedure involves spectral pre-treatments, construction of calibration models and regression methods, and prediction of interest compounds by unknown samples (CORTÉS *et al.*, 2019; WANG *et al.*, 2015).

Initially, spectral data was used to build a matrix of independent variables (X) where the lines are composed of the 350 mean spectra of samples (Figure 1g) and the columns were composed of 71 wavelengths. Thus, the vectors of dependent variables (y) were obtained from each sugarcane quality attribute determined by the conventional analytical methods used at the sugar mill laboratory. The external validation data set (Figure 1h) was composed of the data obtained in-field.

Different pre-processing techniques were applied to the spectral data before calibrations, as baseline correction, standard normal variate (SNV), multiplicative scatter correction (MSC), 1st and 2nd derivatives by Savitzky-Golay algorithm (SAVITZKY; GOLAY, 1964), mean centering and autoscaling. The main objective of this step is to transform data into useful information to develop models with reliable performance to the subsequent multivariate analysis, improving the model accuracy (CORTÉS *et al.*, 2019; PHETPAN; UDOMPETAIKUL; SIRISOMBOON, 2018), and ensuring non-distortion of the chemical information (OLIVERI *et al.*, 2019). The pre-processing adjustment was evaluated by the lowest value of the root mean square error (RMSE) of cross-validation (Equation 2.1).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_i^N (y_i - \hat{y}_i)^2}{N}} \tag{2.1}$$

where y_i and $\hat{y}_i$ are measured and predicted values for each i sample, respectively. When internal cross-validation is used, N represents the number of samples in cross-validation set. Then, root mean square error is named as $RMSE_{CV.}$ For external validation, with prediction samples set, N represents the number of predicted samples, and, in this case, the statistical parameters are named root mean square error of prediction ($RMSE_P$).

A preliminary data overview through principal component analysis (PCA) was performed for both data set (calibration and validation), for examination of pre-processed spectral data to find the similarities, differences and outliers among all the samples. The outliers were removed from the data set according the Hotelling's T^2 and Q-residuals tests, at a significance level of 5% (ASTM, 2017). Then, the modeling was performed applying partial least squares regression (PLSR).

PLSR is one of the most applied multivariate regression methods used to predict food quality attributes. It is a chemometrics covariance-based method that seeks to optimise both the variance explained and correlation with the dependent variables (WOLD; SJÖSTRÖM; ERIKSSON, 2001). The spectral data are decomposed in latent variables (LV), which explain most of the variance of the raw data, modelling the linear relationship between the independent variables (matrix X) and dependent variables (vector y) using the scores of the most correlated features (MORELLOS *et al.*, 2016; PHETPAN; UDOMPETAIKUL; SIRISOMBOON, 2018). PLSR models determine new variables, which are estimates of the latent variables, named X-scores (predictors of the response ones). In addition, the deviations between measured and predicted values (y-residuals) are also obtained (Equation 2.2) (STEIDLE NETO; LOPES; *et al.*, 2017).

$$\mathbf{y} = \mathbf{bX} + \mathbf{e} \tag{2.2}$$

where **y** is the vector of dependent (response) variables (%), **b** indicates a vector of regression coefficients (dimensionless), **X** is the matrix of spectral variables of each sample (independent variables, %), and **e** is a vector that indicates the y-residuals. The model was performed by venetian blinds cross-validation, with 10 splits. The optimal number of LV selected was the value that minimised $RMSE_{CV}$ (Equation 2) the most. Then, the calibrated model was used to predict the quality attributes of the samples collected in-field.

To perform the spatial variability management of interest attributes, it is essential that differences from predicted and observed attribute present low error and the model presents a satisfactory performance. Thus, the regression models were evaluated by the RMSE of prediction (Equation 2.2) and coefficient of determination of prediction (Equation 2.3).

$$R^2 = 1 - \frac{\sum_i^N (y_i - \hat{y}_i)^2}{\sum_i^N (y_i - \bar{y}_i)^2} \tag{2.3}$$

where y_i and $\hat{y}_i$ are measured and predicted values for each i sample, respectively. The $\bar{y}_i$ is the mean y values. For external validation, with prediction samples set, N represents the number of predicted samples, and, in this case, the statistical parameter is named coefficient of determination of prediction (R^2_P).

The variable importance for the projection (VIP) was calculated to verify the wavelengths with greater impact on the external validate models (Equation 2.4).

$$VIP_j = \sqrt{p \sum_{k=1}^{h} \left[Z \left(\frac{w_{jk}}{||w_k||} \right)^2 \right] . \left(\sum_{k=1}^{h} Z \right)^{-1}} \tag{2.4}$$

where VIP is the variable importance for projection (dimensionless), j is a specific wavelength (nm), p is the number of wavelengths (dimensionless), h is the number of latent variables (dimensionless), Z is the fraction of variance in the prediction explained by the latent variable (dimensionless), and w is the loading weight (dimensionless). This parameter was analysed together with the regression vector generated by the modeling.

In this study, Brix, Pol and Fibre models were developed using PLSR, which was performed in Matlab framework (Matlab R2015a, The MathWorks Inc., Natick, USA) and PLS-Toolbox 8.8 (Eigenvector Research Inc., Wenatchee, USA).

2.2.5. Spatial variability analysis

Geostatistical analyses were performed on the georeferenced database containing observed and predicted values from the laboratory analysis and by means of the best models from the chemometrics analysis, respectively. Semi-variograms were fitted for each set of observed and predicted sugarcane qualitative parameter using VESPER software (MINASNY; MCBRATNEY; WHELAN, 2005) to estimate the structure of the spatial variation. The semi-variogram models that resulted in the lowest RMSE were selected. Spatial dependence was classified according to Cambardella *et al.* (1994). Maps were produced by interpolating the data using ordinary kriging at a spatial resolution of 1 m^2. Then, sugarcane quality attributes were mapping using the Quantum GIS Software 2.18 version (QGIS Development Team, 2016).

The similarity between measured and predicted maps for the sugarcane quality attributes was evaluated by visual comparison (TÜMSAVAŞ *et al.*, 2019) and using the coefficient of correlation (COLAÇO *et al.*, 2019).

2.3. Results and discussion

2.3.1. Overview of the statistic characteristics and spectral data

Results of sugarcane quality attributes obtained by conventional analysis for calibration step and field samples (external validation step) are shown in Table 1. The in-field values found for Brix and Pol ranged between the maximum and minimum values obtained for calibration samples. For Fibre, the minimum value observed in the external validation data set were lower than the minimum observed in the calibration samples.

After the pre-processing and applying PCA, only one sample in the external validation data set was identified by Hotelling T^2 and Q-residual statistic tests as an outlier, indicated as sample poorly modelled with high influence on the model, thus it was removed from the data set.

Table 1. Descriptive statistics of the samples for the calibration and prediction sets, after outlier removal.

Sample Set	---------- Calibration ----------			------ External validation ------		
Quality attribute	Brix	Pol	Fibre	Brix	Pol	Fibre
N	---------- 350 ----------			---------- 90 ----------		
Max (%)	23.8	20.9	19.1	19.2	18.0	11.9
Mean (%)	20.46	17.70	12.77	17.48	15.58	10.70
Min (%)	13.8	10.1	10.2	15.7	13.4	9.7
Median (%)	20.73	17.98	12.60	17.60	15.53	10.70
SD (%)	1.4746	1.6118	1.3153	0.69	0.98	0.5658
CV (%)	7.22	9.12	10.30	3.99	6.30	5.29

N: number of samples; Max: maximum; Min: minimum; SD: standard deviation; CV: coefficient of variation

The coefficient of variation (CV) presented in the calibration data set was higher than those in the validation data set independently of the attribute, characterizing the diversity induced by samples from different cultivars obtained in the calibration data set. It is noteworthy that at this stage, it was desirable to obtain samples from different cultivars to obtain greater diversity of quality attributes to build a robust prediction model (PHETPAN; UDOMPETAIKUL; SIRISOMBOON, 2018).

The mean spectra, as well the standard deviation, for calibration and external validation samples are shown in Figures 2a and 2b, respectively.

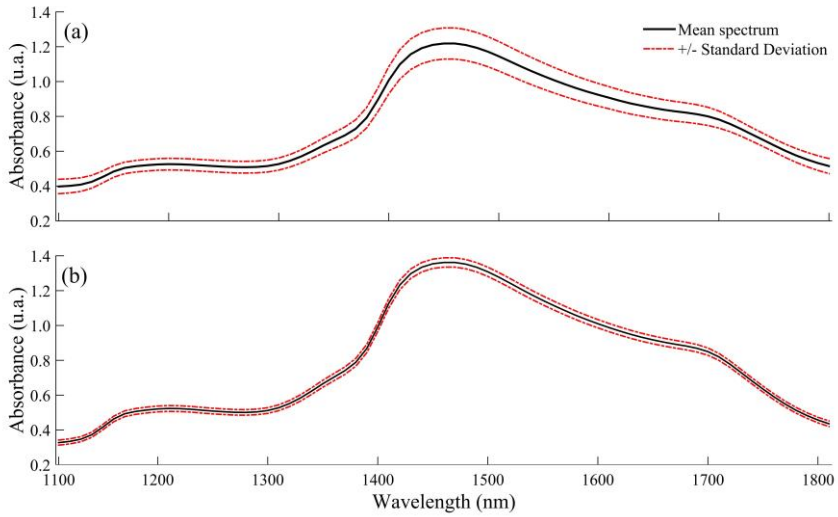

Figure 2. Mean (black line) spectra, with standard deviation above and below (dashed red lines), of spectral data of defibrated sugarcane samples of calibration (a) and external validation (b). u.a.: arbitrary units.

Looking at the quality attributes, the PLSR model was more accurate when it was pre-processed with the standard normalize variation method (SNV) followed by second derivative using Savitzky-Golay filter (SAVITZKY; GOLAY, 1964), with a window size of nine spectral bands and a polynomial degree of two. The SNV removes the path length variations on which the average value of a spectrum is subtracted from the original spectrum, and then the result is divided by the standard deviation (WANG *et al.*, 2015), and second derivative removes the offset and slope on the spectral data.

2.3.2. Performance of the calibration and prediction models

A proper model should have a low RMSE and a high R^2, for both calibration and prediction steps (NAWI; CHEN; JENSEN, 2013). The results obtained for the models predicting Brix and Pol were satisfactory (Figure 3a and 3b), presenting $RMSE_P$ of 0.80 and 0.58, respectively, and R^2_P of 0.71 for both models (Table 2). Using absorbance spectra, Nawi *et al.* (2013) predicted Brix content from stalk samples achieving $RMSE_P$ of 1.51 %, value higher than the one found in this study ($RMSE_P$ of 0.80 %). In a study predicting Pol, also from the whole stalk samples, Maraphum *et al.* (2018) evaluated the effect of skin wax on its estimation using NIRS. As a result, they found values of $RMSE_P$ ranging from 1.2% to 1.5% by scanning stalks without wax. These values were higher than the value of $RMSE_P$ (0.58 %) found in this study from defibrated samples. The use of defibrated samples in NIRS analysis present several benefits such: elimination of the negative effect that skin compounds, i.e. wax, causes to the reading; and promotion of a more homogenous stem sample, since the plant present uneven sucrose concentration along the stem, with accumulation in an upward direction (NAWI; CHEN; JENSEN, 2013; WANG *et al.*, 2013). On the other hand, the Fibre predictive model presented an unsatisfactory performance, compared to the other attributes, with values of R^2 below of 0.3 and RMSE above of 1% (Figure 3c).

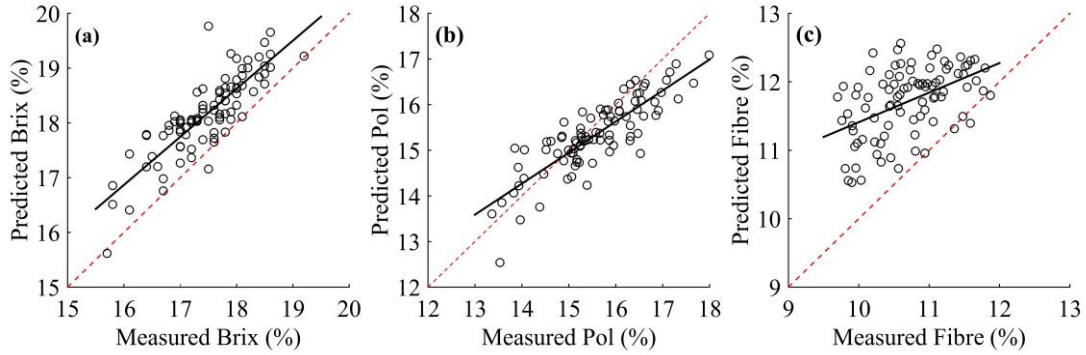

Figure 3. Scatter plots of measured versus predicted solids soluble content (Brix) (a), sucrose (Pol) (b) and Fibre content (c) of external validation data set.

Phuphaphud *et al.* (2019) obtained better results for predicting sugarcane Fibre content than those found in this study (Table 2), with $RMSE_P$ and R^2_P values of 0.81 and 0.63, respectively. However, the authors used the stalk measurement approach, divided into three sections, a reality that is difficult to implement for an embedded sensor on a sugarcane harvester due to its layout and mechanical characteristics not allowing to distinguish each section after passing through the chipper roll. Therefore, the viability of an embedded NIRS sensor depends on ways of obtaining homogeneous and representative samples of each portion harvested on the crop (CORRÊDO *et al.*, 2020).

As an overtone spectroscopy, NIRS incorporate abundant information concerning OH, CH and NH vibration absorptions (WANG *et al.*, 2015). In the modelling by PLSR, the observed values are approximated by the values of the predictors through a linear combination, and the coefficients of this combination are the regression coefficients (TÜMSAVAŞ *et al.*, 2019). Each coefficient is associated to a wavelength and represent its ability to affect the performance of the model to predict the chemical properties of interest (WANG *et al.*, 2015). In addition, the variables that are most important in determining the attributes of interest can be identified through the scores of the variable importance in projection (VIP) (STEIDLE NETO; TOLEDO; *et al.*, 2017).

Table 2. Model results for partial least squares regression (PLSR) in cross-validation (cv) and external validation (p) sets for the prediction of solids soluble content (Brix), sucrose (Pol) and Fibre content.

Quality attribute	LV	--- Cross-validation ---		--- External Validation ---	
		$RMSE_{CV}$ (%)	R^2_{CV}	$RMSE_P$ (%)	R^2_P
Brix	7	0.66	0.80	0.80	0.71
Pol	7	0.80	0.75	0.58	0.71
Fibre	8	1.13	0.26	1.15	0.24

LV: latent variables; RMSE: root mean square error; R^2: coefficient of determination; cv = cross-validation; p = prediction.

Comparing the regression coefficients and the VIP scores along the wavelength range for Brix and Pol, it reveals visual substantial similarities (Figure 4). Among soluble solids there are inorganic and organic compounds (sugars and non-sugars). Sugars are found in the form of glucose and fructose, called reducing sugars, and sucrose. As already mentioned, in the laboratory, sucrose is measured indirectly by Pol. Brix and Pol are highly correlated attributes (MANCINI *et al.*, 2012), probably due to the higher proportion of sucrose in the concentration of soluble solids. This fact explains the similarities found between the response of these attributes.

There are no differences between Brix and Pol for the VIP scores and the regression vector in relation to the wavelengths as expected, due to the compositional relationship between these attributes. Thus, it is possible to infer the important regions for modelling using these two metrics. The regression vector plots are represented in the Figure 4a for Brix and Figure 4b for Pol. Similarly, the more sensitive wavelengths to the response variable are represented by the VIP scores for Brix (Figure 4c) and Pol (Figure 4d). For VIP values higher than 1, the variables are highly influential (WOLD; SJÖSTRÖM; ERIKSSON, 2001). Absorption bands around 1100 nm and between 1410 and 1440 nm represent the first overtones of OH combinations and stretching, respectively (GOLIC; WALSH; LAWSON, 2003).

As documented by Golic, Walsh and Lawson (2003), in the wavelengths between 1100 nm and 1230 nm the second overtone of CH stretching vibration can occur. It is possible to observe a prominent peak at 1160 nm. Most evident in the middle of the VIP score plots, there is a wide region of absorption bands with greater influence to predict the quality attributes between 1360 nm and 1490 nm, which occurs combinations of vibrational modes involving CH and first overtone of OH and NH chemical bonds (OSBORNE, 2006). There is a peak with a higher VIP value, compared the wavelengths analysed, at 1390 nm for Pol (Figure 4d). On the other hand, its value is surpassed by the band at 1730 nm for Brix (Figure 4c). At 1690 nm and 1730 nm regions, absorbance arising from CH stretching first overtone (BÁZÁR *et al.*, 2015; GOLIC; WALSH; LAWSON, 2003). In the following bands, between 1720 nm and 1765 nm occurs a peak at 1730 nm. This absorption range is arising from CH_2 stretching first overtone of molecules vibration (GOLIC; WALSH; LAWSON, 2003).

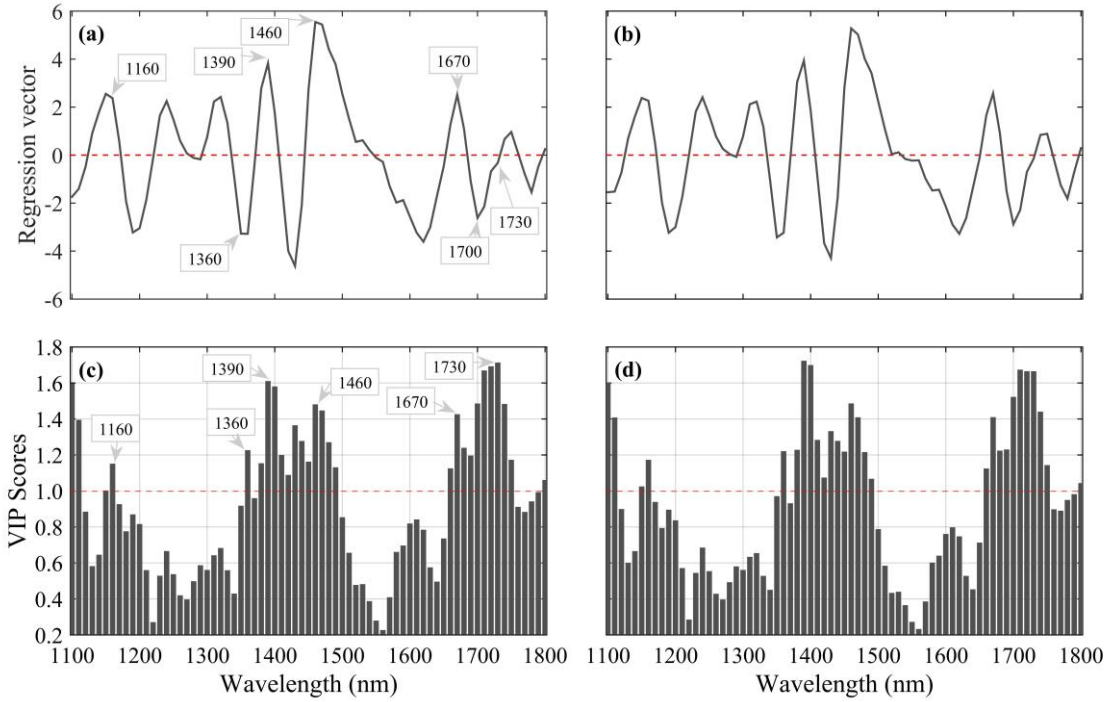

Figure 4. Variation of regression coefficients with wavelength, obtained with partial least squares regression (PLSR) analysis of Brix (a) and Pol (b) of sugarcane, and Variable Importance in Projection (VIP) for models used to predict Brix (c) and Pol (d).

VIP values lower than 0.8 is insignificant for forecasting (STEIDLE NETO; TOLEDO; *et al.*, 2017). There were two absorption intervals in which this occurred with greater intensity: 1210 nm to 1340 nm and 1510 nm and 1650 nm. For VIP values between 0.8 and 1.0, there is moderately influential variables. Indeed, when

considering the analysis of soluble solids content (Brix), there is a low concentration of inorganic, organic non-sugars compounds and reducing sugars (glucose and fructose), besides saccharose content (Pol). The presence of these compounds in the samples is probably the reason for the minimum differences observed between the graphs of these two attributes.

2.3.3. Sugarcane quality mapping

It is evident that NIRS is a suitable tool to provide predictor variables to be used in sugarcane quality forecast models. This fact is corroborated by several studies in the literature (NAWI *et al.*, 2013; PHETPAN; UDOMPETAIKUL; SIRISOMBOON, 2018; PHUPHAPHUD *et al.*, 2020; PHUPHAPHUD; SAENGPRACHATANARUG; POSOM; MARAPHUM; *et al.*, 2019; TAIRA; UENO; KAWAMITSU, 2010). In view of this and aware that Brazilian sugarcane farmers are paid based on their quantitative (yield – Mg ha-1) and for the quality of the raw material, it is fairer to farmers to be paid according to the results of each area considering their variability and not being paid by the average value. Thus, sugarcane quality mapping is justified, not only because it provides a more reasonable approach to the buyer and seller, but also, it supports farmers decision-making regarding their crop (FARID *et al.*, 2016), providing an important layer of information to site-specific management of the crops to meet the precision agriculture objectives (ISPA, 2019).

However, a method to obtain high-density data for attributes related to sugarcane quality is not developed, despite the existence of methods to obtain quantitative data (MAGALHÃES; CERRI, 2007; MOLIN, J. P.; MENEGATTI, 2004). Thus, the use of NIRS provides a handy, fast, accurate and reliable method to forecast sugarcane quality attributes. We believe that minimum preparation should solve the issue of variability of contents along the cane stem, as it provides a more uniform and homogeneous sampling. Although, sugarcane harvest is a particular tough environment and the absence of sample prepare is an aspect desirable for an on-board sensor development. Some researchers have achieved satisfactory results for solids soluble content prediction in controlled conditions for sugarcane billets measurements (PHETPAN *et al.*, 2018; UDOMPETAIKUL; PHETPAN; SIRISOMBOON, 2021). However, it is needed advance for in-field applications.

The maps resulting of laboratory measurements and the maps obtained by NIR measurements show visual spatial similarity for both sugarcane attributes as shown in Figure 5. These results are expected since the prediction model presented low RMSEP and high R^2p for both Pol and Brix content as already seen in Table 2. Initially, there is a region distinguished as having a low potential for Brix and Pol when compared to other regions of the field, extending practically for all the east corner towards the upper central area. On the other hand, there is a region, in the lower central region of the field, that presents high Brix and Pol content in relation to the other areas of the field. Figure 5 highlights the importance for mapping sugarcane quality parameters since there is an uneven distribution for Pol and Brix values along the area. This uneven distribution is expected for both qualitative and quantitative results of the crop as they are the result of a complex interaction among soil properties, weather conditions (macro and microclimate), weed infestations and other factors that present spatial peculiarities (JOHNSON; RICHARD, 2005; SANCHES; GRAZIANO MAGALHÃES; JUNQUEIRA FRANCO, 2019).

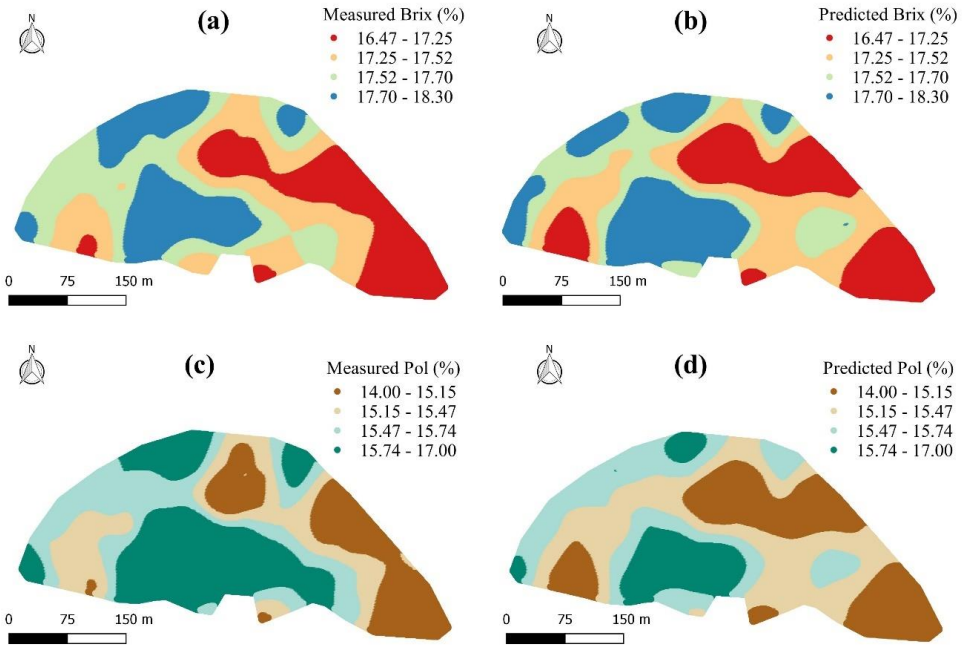

Figure 5. Maps of Brix measured (a) and predicted (b), and Pol measured (c) and predicted (d), calculated with partial least squares (PLS) regression.

All geostatistical models presented moderate spatial dependence (Table 3), according to the classification proposed by Cambardella *et al.* (1994). These results were similar to those obtained by Rodrigues, Magalhães and Franco (2013), when analysing the spatial variability of quality attributes using physical and chemical soil attributes and leaf nitrogen as predictors to fit mathematical models to predict Brix, Pol and Fibre. Looking at the range values obtained from predicted and measured models for both Brix and Pol content, the lowest value among all variables was 122.5 m indicating that it is necessary to collect 2.67 samples ha^{-1} to represent faithfully both Brix and Pol content. Thus, comparing the calculated sampling grid (2.67 samples ha^{-1}) to the applied grid (5.50 samples ha^{-1}), it can be concluded that the applied sampling grid was effectively well-designed. In addition to that, in the studied conditions, the results show that there is no need for high density sampling for sugarcane quality analysis indicating that three samples ha-1 could be enough. Thus, it opens an opportunity for farmers to adopt sugarcane quality sampling and submit the samples to the NIRS method to predict its qualitative attributes since there is not a developed method to be applied in large scale.

Table 3. Variogram parameters for Brix and Pol measured and predicted values.

		Model	C	C_0	($C+C_0$)	A (m)	SDI (%)
Brix	Measured	Spherical	0.1762	0.2982	0.4744	132.4	62.9
	Predicted	Spherical	0.2511	0.279	0.5301	122.5	52.6
Pol	Measured	Spherical	0.5114	0.4872	0.9986	150.8	48.8
	Predicted	Spherical	0.2723	0.3542	0.6265	124.1	56.5

C: Partial sill, C_0: Nugget, ($C+ C_0$): Sill, A: Range, SDI: spatial dependence index.

The spatial distribution of the model residuals for Brix and Pol by PLSR are show in Figure 6. The models tested did not show a similar spatial pattern of residuals across the two spatial parameters. There is a tendency to overestimate (residuals between 0.21 - 0.42%) in areas with low Brix values (Figure 5a and 5b, Figure

6a). On the other hand, for Pol, overestimation in the area with the greatest potential for this attribute is more evident (Figure 5c and 5d, Figure 6b). This overestimation and underestimation are the result of the bias-variance tradeoff (MEHTA *et al.*, 2019). However, both presented a satisfactory performance with residuals values lower than 0.77%, for under or overestimation. This is easily recognised when looking at the transition areas, ranging from -0.22 to 0.21% for Brix, and -0.26 to 0.42 % for Pol, occupying considerably the largest area of the maps.

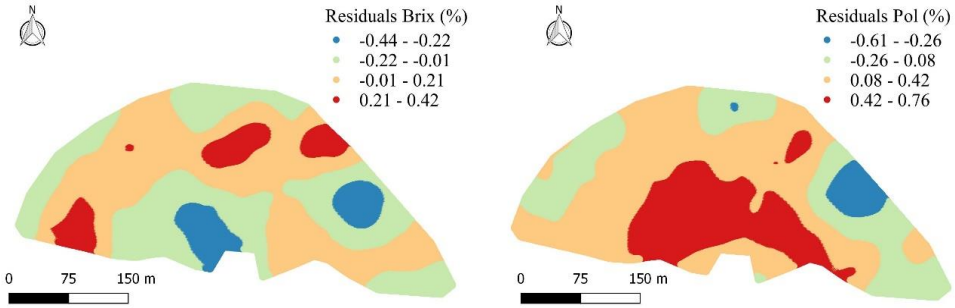

Figure 6. Maps of residuals calculated with partial least squares regression (PLSR) for prediction of Brix (a) and Pol (b) of sugarcane. Dark hues show overestimation and light hues show underestimation of the parameter's value.

Given the importance of TRS for sugar producers and sugar mills to price and market the product, and the future perspective of obtaining maps as a tool for managing cause and effect relationships and industrial planning, we analysed the possibility of using maps from base attributes (Brix and Pol) to understand the spatial variability of the sugarcane quality. Figure 7 shows the total recoverable sugar (TRS) map obtained by conventional measurements, in kilograms of sugar recoverable by the industrial process per megagram of sugarcane.

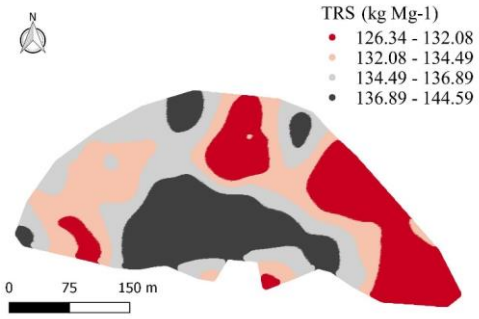

Figure 7. Total sugar recoverable map.

On a first view, there is a visual spatial similarity between TRS (Figure 7) and measured Pol maps (Figure 5c). This was expected due to the presence of a high correlation between these attributes (FERRAZ *et al.*, 2019). The correlation analysis among TRS, measured and predicted Brix and Pol are shown in the Table 4. The correlation between the measured Brix and Pol, and measured TRS were 0.89 and 0.99, respectively. The TRS is obtained by an empirical equation as a function of the pol of cane and reducing sugars of cane, according to CONSECANA (2015). Both attributes are obtained as a function of the pol of juice and reducing sugars of juice, respectively, and as a function inversely proportional to the Fibre content. Despite these relations, in this study, it was not possible to accurately predict Fibre content, therefore, making unfeasible to use this data to predict TRS by NIRS. Possibly, this

is because the sample range used for calibration does not include minimum values observed in the sample set used in the external validation (Table 1). A viable alternative for future works should pass for data processing, evaluating the use of non-linear methods for modeling the Fibre content (SEXTON *et al.*, 2018). However, TRS presents high correlation with Pol, which is an indirect measure of sucrose content, as mentioned earlier, and presents a low inversely correlation with Fibre content (MANCINI *et al.*, 2012).

From the models developed by means of NIRS, the correlation obtained between the interpolated predicted values of Brix and Pol by NIRS and the measured TRS, determined by conventional methods, was 0.85 for both attributes (Table 4, bold numbers). The high correlation presented among attributes, indicate the possibility of applying NIRs technology to measure the quality of sugarcane in the field.

Table 4. Pearson's correlation matrix among sugarcane quality maps.

	$Brix_{measured}$	$Brix_{predicted}$	$Pol_{measured}$	$Pol_{predicted}$	$TRS_{measured}$
$Brix_{measured}$	1.00				
$Brix_{predicted}$	0.91*	1.00			
$Pol_{measured}$	0.89*	0.88*	1.00		
$Pol_{predicted}$	0.90*	1.00*	0.88*	1.00	
$TRS_{measured}$	0.89*	**0.85***	0.99*	**0.85***	1.00

* Significant at 5%.

2.3.4. Future works

Future studies with online measurements may assist in advances regarding the application of this technology aiming to obtain high density data. In addition, minimal sample preparation, such as measurements from the stalk, can assist in the implementation of technology (NAWI *et al.*, 2014). Some studies pointed to promising results in this direction (NAWI *et al.*, 2013; PHETPAN; UDOMPETAIKUL; SIRISOMBOON, 2018; PHUPHAPHUD *et al.*, 2020). However, the performance of these systems in field conditions has not yet been evaluated. In addition, the achievement of TRS results by NIRS depends on the development of robust forecast models for Fibre content, in addition to Pol and Brix. Future studies using non-linear methods for modelling multivariate data should help in this regard. Future perspectives on obtaining data related with crop quality will assist in the management of agricultural inputs or even in studies focusing the production of as-apply maps having both layers of information, yield and quality, as a tool to support decision making to evaluate the cause-effect relation between TRS and inputs used in the crop (CORRÊDO *et al.*, 2020). Finally, future study with a NIR system embedded in a sugarcane harvester would help to improve the robustness of the technology, enabling the collection of high-density data in sugarcane fields improving the geostatistical analysis to evaluate the spatial variability. Therefore, supporting for studies to overcome the intrinsic adversities of the sugarcane harvesting process.

2.4. Conclusions

NIRS models were developed from parallel with conventional laboratory analyses for determining the quality of sugarcane in a sugar mill. The models developed for Brix and Pol were sufficiently accurate to achieve satisfactory results for the prediction of external validation data, with RMSEP of 0.80 and 0.58, respectively, and R^2P

of 0.71 for both. However, results obtained for Fibre by NIRS models were not satisfactory (R^2P of 0.24 and RMSEP of 1.15).

These data were composed of georeferenced samples obtained in a commercial field of sugarcane to characterise the spatial variability of the attributes through the NIRS. The results obtained in the laboratory and through prediction via NIRS showed spatial dependence. In addition, after data interpolation, the maps generated for the attributes from results obtained in the laboratory by conventional analysis obtained a high correlation with the maps generated from the NIRS results (0.91 for Brix and 0.88 for Pol). In addition, both showed a strong correlation with the TRS map generated from the laboratory results (0.85 for both). This study demonstrates that the estimation of sugarcane quality attributes by NIRS allows the analysis of the spatial variability of these attributes in the field. Moreover, future studies should be allowed the requirement for sample preparation be overcome.

References

BÁZÁR, G.; KOVACS, Z.; TANAKA, M.; FURUKAWA, A.; NAGAI, A.; OSAWA, M.; ITAKURA, Y.; SUGIYAMA, H.; TSENKOVA, R. Water revealed as molecular mirror when measuring low concentrations of sugar with near infrared light. **Analytica Chimica Acta**, vol. 896, p. 52–62, 8 Oct. 2015. https://doi.org/10.1016/j.aca.2015.09.014.

CAMBARDELLA, C. A.; MOORMAN, T. B.; NOVAK, J. M.; PARKIN, T. B.; KARLEN, D. L.; TURCO, R. F.; KONOPKA, A. E. Field-Scale Variability of Soil Properties in Central Iowa Soils. **Soil Science Society of America Journal**, vol. 58, no. 5, p. 1501–1511, 1994. https://doi.org/10.2136/sssaj1994.03615995005800050033x.

COLAÇO, A. F.; MOLIN, J. P.; ROSELL-POLO, J. R.; ESCOLÀ, A. Spatial variability in commercial orange groves. Part 2: relating canopy geometry to soil attributes and historical yield. **Precision Agriculture**, vol. 20, no. 4, p. 805–822, 2019. DOI 10.1007/s11119-018-9615-0. Available at: https://doi.org/10.1007/s11119-018-9615-0. Accessed on: 17 May 2020.

CONSECANA. National Concill of Sugarcane Producers of São Paulo State. vol. 6, p. 81, 2015. .

CORRÊDO, L. de P.; CANATA, T. F.; MALDANER, L. F.; DE LIMA, J. de J. A.; MOLIN, J. P. Sugarcane Harvester for In-field Data Collection: State of the Art, Its Applicability and Future Perspectives. **Sugar Tech**, vol. 22, no. 4, p. 1–14, 10 Aug. 2020. DOI 10.1007/s12355-020-00874-3. Available at: http://link.springer.com/10.1007/s12355-020-00874-3. Accessed on: 10 Aug. 2020.

CORTÉS, V.; BLASCO, J.; ALEIXOS, N.; CUBERO, S.; TALENS, P. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. **Trends in Food Science and Technology**, vol. 85, p. 138–148, 1 Mar. 2019. https://doi.org/10.1016/j.tifs.2019.01.015.

FARID, H. U.; BAKHSH, A.; AHMAD, N.; AHMAD, A.; MAHMOOD-KHAN, Z. Delineating site-specific management zones for precision agriculture. **Journal of Agricultural Science**, vol. 154, no. 2, p. 273–286, 2016. DOI 10.1017/S0021859615000143. Available at: https://doi.org/10.1017/S0021859615000143. Accessed on: 21 May 2020.

FERRAZ, M. N.; CORRÊDO, L. P.; WEI, M. C. F.; MOLIN, J. P. Spatial variability mapping of sugarcane qualitative attributes. **Engenharia Agrícola**, vol. 39, no. spe, p. 109–117, 2019. DOI 10.1590/1809-4430-eng.agric.v39nep109-117/2019. Available at: http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v39nep109-117/2019. Accessed on: 31 Oct. 2019.

GOLIC, M.; WALSH, K.; LAWSON, P. Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature. **Applied Spectroscopy**, vol. 57, no. 2, p. 139–145, 2003. DOI 10.1366/000370203321535033. Available at: https://journals.sagepub.com/doi/pdf/10.1366/000370203321535033. Accessed on: 9 Jan. 2019.

JOHNSON, R. M.; RICHARD, E. P. Sugarcane yield, sugarcane quality, and soil variability in Louisiana. **Agronomy Journal**, vol. 97, no. 3, p. 760–771, 2005. https://doi.org/10.2134/agronj2004.0184.

MAGALHÃES, P. S. G. G.; CERRI, D. G. P. P. Yield Monitoring of Sugar Cane. **Biosystems Engineering**, vol. 96, no. 1, p. 1–6, Jan. 2007. DOI 10.1016/j.biosystemseng.2006.10.002.

MALDANER, L. F.; MOLIN, J. P. Data processing within rows for sugarcane yield mapping. **Scientia Agricola**, vol. 77, no. 5, p. 2020, 2020. DOI 10.1590/1678-992x-2018-0391. Available at: http://dx.doi.org/10.1590/1678-992X-2018-0391. Accessed on: 24 Dec. 2019.

MANCINI, M. C.; LEITE, D. C.; PERECIN, D.; BIDÓIA, M. A. P.; XAVIER, M. A.; LANDELL, M. G. A.; PINTO, L. R. Characterization of the Genetic Variability of a Sugarcane Commercial Cross Through Yield Components and Quality Parameters. **Sugar Tech**, vol. 14, no. 2, p. 119–125, 2012. https://doi.org/10.1007/s12355-012-0141-5.

MARAPHUM, K.; CHUAN-UDOM, S.; SAENGPRACHATANARUG, K.; WONGPICHET, S.; POSOM, J.; PHUPHAPHUD, A.; TAIRA, E. 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**, vol. 26, no. 5, p. 287–296, 2018. DOI 10.1177/0967033518795810. Available at: https://journals.sagepub.com/doi/pdf/10.1177/0967033518795810. Accessed on: 11 Mar. 2019.

MEHMOOD, T.; AHMED, B. The diversity in the applications of partial least squares: An overview. **Journal of Chemometrics**, vol. 30, no. 1, p. 4–17, Jan. 2016. DOI 10.1002/cem.2762. Available at: http://doi.wiley.com/10.1002/cem.2762. Accessed on: 5 Feb. 2020.

MEHTA, P.; BUKOV, M.; WANG, C. H.; DAY, A. G. R.; RICHARDSON, C.; FISHER, C. K.; SCHWAB, D. J. A high-bias, low-variance introduction to Machine Learning for physicists. **Physics Reports**, vol. 810, p. 1–124, 30 May 2019. https://doi.org/10.1016/j.physrep.2019.03.001.

MINASNY, B.; MCBRATNEY, A. B.; WHELAN, B. M. **VESPER 1.5 - Spatial Prediction Software for Precision Agriculture**. [*S. l.*: *s. n.*], 2005.

MOLIN, J. P.; MENEGATTI, L. A. A. Field-testing of a sugar cane yield monitor in Brazil. **ASAE Annual International Meeting 2004**, vol. 0300, no. 04, p. 733–744, 2004. https://doi.org/10.13031/2013.16159.

MORELLOS, A.; PANTAZI, X. E.; MOSHOU, D.; ALEXANDRIDIS, T.; WHETTON, R.; TZIOTZIOS, G.; WIEBENSOHN, J.; BILL, R.; MOUAZEN, A. M. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. **Biosystems Engineering**, vol. 152, p. 104–116, 2016. DOI 10.1016/j.biosystemseng.2016.04.018. Available at: https://ac.els-cdn.com/S1537511015304165/1-s2.0-S1537511015304165-main.pdf?_tid=9bd2a49e-74fd-4d85-879e-0462d1e3e1e5&acdnat=1546893219_bfe1381c3b989b935bfd4e7d9069837f. Accessed on: 7 Jan. 2019.

NAWI, N. M.; CHEN, G.; JENSEN, T. In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: a review. **Precision Agriculture**, vol. 15, no. 6, p. 684–703, 2014. https://doi.org/10.1007/s11119-014-9362-9.

NAWI, N. M.; CHEN, G.; JENSEN, T. Visible and shortwave near infrared spectroscopy for predicting sugar content of sugarcane based on a cross-sectional scanning method. **Journal of Near Infrared Spectroscopy**, vol. 21, no. 4, p. 289–297, 2013. https://doi.org/10.1255/jnirs.1060.

NAWI, N. M.; CHEN, G.; JENSEN, T.; MEHDIZADEH, S. A. Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared. **Biosystems Engineering**, vol. 115, p. 154–161, 2013. DOI 10.1016/j.biosystemseng.2013.03.005. Available at: http://dx.doi.org/10.1016/j.biosystemseng.2013.03.005.

NAWI, N. M.; ROWSHON, K. M.; CHEN, G.; TROY, J.; MAT, N. N.; ROWSHON, K. M.; GUANGNAN, C.; TROY, J. Prediction of Sugarcane Quality Parameters Using Visible-shortwave Near Infrared Spectroradiometer. **Agriculture and Agricultural Science Procedia**, vol. 2, p. 136–143, 1 Jan. 2014. DOI 10.1016/j.aaspro.2014.11.020. Available at: www.sciencedirect.com. Accessed on: 27 Aug. 2018.

OLIVERI, P.; MALEGORI, C.; SIMONETTI, R.; CASALE, M. The impact of signal pre-processing on the final interpretation of analytical outcomes – A tutorial. **Analytica Chimica Acta**, vol. 1058, p. 9–17, 13 Jun. 2019. https://doi.org/10.1016/j.aca.2018.10.055.

OSBORNE, B. G. Near-Infrared Spectroscopy in Food Analysis. **Encyclopedia of Analytical Chemistry**, Chichester, UK, , p. 14, 30 Oct. 2006. DOI 10.1002/9780470027318.a1018. Available at: http://doi.wiley.com/10.1002/9780470027318.a1018. Accessed on: 11 Dec. 2018.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, vol. 1026, p. 8–36, 2018. https://doi.org/10.1016/j.aca.2018.04.004.

PHETPAN, K.; UDOMPETAIKUL, V.; SIRISOMBOON, P. An online visible and near-infrared spectroscopic technique for the real-time evaluation of the soluble solids content of sugarcane billets on an elevator conveyor. **Computers and Electronics in Agriculture**, vol. 154, p. 460–466, Nov. 2018. DOI 10.1016/j.compag.2018.09.033. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0168169918302096. Accessed on: 7 Jan. 2019.

PHUPHAPHUD, A.; SAENGPRACHATANARUG, K.; POSOM, J.; MARAPHUM, K.; TAIRA, E. Non-destructive and rapid measurement of sugar content in growing cane stalks for breeding programmes using visible-near infrared spectroscopy. **Biosystems Engineering**, vol. 197, p. 76–90, 1 Sep. 2020. https://doi.org/10.1016/j.biosystemseng.2020.06.012.

PHUPHAPHUD, A.; SAENGPRACHATANARUG, K.; POSOM, J.; MARAPHUM, K.; TAIRA, E. Prediction of the fibre content of sugarcane stalk by direct scanning using visible-shortwave near infrared spectroscopy. **Vibrational Spectroscopy**, vol. 101, p. 71–80, Mar. 2019. DOI 10.1016/j.vibspec.2019.02.005. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0924203118302613. Accessed on: 3 Jun. 2019.

RIDESA. **Liberação nacional de variedades RB de cana-de-açúcar**. [*S. l.*: *s. n.*], 2015.

RODRIGUES, F. A.; MAGALHÃES, P. S. G.; FRANCO, H. C. J. Soil attributes and leaf nitrogen estimating sugar cane quality parameters: Brix, pol and fibre. **Precision Agriculture**, vol. 14, no. 3, p. 270–289, 2013. https://doi.org/10.1007/s11119-012-9294-1.

SANCHES, G. M.; GRAZIANO MAGALHÃES, P. S.; JUNQUEIRA FRANCO, H. C. Site-specific assessment of spatial and temporal variability of sugarcane yield related to soil attributes. **Geoderma**, vol. 334, p. 90–98, 15 Jan. 2019. DOI 10.1016/j.geoderma.2018.07.051. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0016706118303513. Accessed on: 27 Dec. 2018.

SANSEECHAN, P.; PANDUANGNATE, L.; SAENGPRACHATANARUG, K.; WONGPICHET, S.; TAIRA, E.; POSOM, J. A portable near infrared spectrometer as a non-destructive tool for rapid screening of solid density stalk in a sugarcane breeding program. **Sensing and Bio-Sensing Research**, vol. 20, p. 34–40, Sep. 2018. DOI 10.1016/j.sbsr.2018.07.001. Available at: https://linkinghub.elsevier.com/retrieve/pii/S2214180418300515. Accessed on: 7 Jan. 2019.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, vol. 36, no. 8, p. 1627–1639, 1964. DOI 10.1021/ac60214a047. Available at: https://pubs.acs.org/sharingguidelines. Accessed on: 28 Mar. 2020.

SEXTON, J.; EVERINGHAM, Y.; DONALD, D.; STAUNTON, S.; WHITE, R. A comparison of non-linear regression methods for improved on-line near infrared spectroscopic analysis of a sugarcane quality measure. **Journal of Near Infrared Spectroscopy**, 2018. DOI 10.1177/0967033518802448. Available at: https://journals.sagepub.com/doi/pdf/10.1177/0967033518802448. Accessed on: 11 Mar. 2019.

SILVA, C. B.; DE MORAES, M. A. F. D.; MOLIN, J. P. Adoption and use of precision agriculture technologies in the sugarcane industry of São Paulo state, Brazil. **Precision Agriculture**, vol. 12, no. 1, p. 67–81, 6 Jan. 2011. DOI 10.1007/s11119-009-9155-8. Available at: https://link.springer.com/article/10.1007/s11119-009-9155-8. Accessed on: 11 Aug. 2021.

SOROL, N.; ARANCIBIA, E.; BORTOLATO, S. A.; OLIVIERI, A. C. Visible/near infrared-partial least-squares analysis of Brix in sugar cane juice. A test field for variable selection methods. **Chemometrics and Intelligent Laboratory Systems**, vol. 102, no. 2, p. 100–109, 2010. DOI 10.1016/j.chemolab.2010.04.009. Available at: https://ac.els-cdn.com/S0169743910000584/1-s2.0-S0169743910000584-main.pdf?_tid=9ea5d650-a1e2-40fd-8b20-2089be31da78&acdnat=1549390123_c9d156b7ef74f98af5764f4ba293261e. Accessed on: 2 Jul. 2018.

STEIDLE NETO, A. J.; LOPES, D. C.; PINTO, F. A. C.; ZOLNIER, S. Vis/NIR spectroscopy and chemometrics for non-destructive estimation of water and chlorophyll status in sunflower leaves. **Biosystems Engineering**, vol. 155, p. 124–133, Mar. 2017. DOI 10.1016/j.biosystemseng.2016.12.008. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1537511015304505. Accessed on: 26 Mar. 2019.

STEIDLE NETO, A. J.; TOLEDO, J. V.; ZOLNIER, S.; LOPES, D. de C.; PIRES, C. V.; SILVA, T. G. F. d. Prediction of mineral contents in sugarcane cultivated under saline conditions based on stalk scanning by Vis/NIR spectral reflectance. **Biosystems Engineering**, vol. 156, p. 17–26, Apr. 2017. DOI 10.1016/j.biosystemseng.2017.01.003. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1537511016302513. Accessed on: 7 Jan. 2019.

TAIRA, E.; UENO, M.; KAWAMITSU, Y. Automated quality evaluation system for net and gross sugarcane samples using near infrared spectroscopy. **Journal of Near Infrared Spectroscopy**, vol. 18, no. 3, p. 209–215, 2010. DOI 10.1255/jnirs.884. Available at: https://journals.sagepub.com/doi/pdf/10.1255/jnirs.884. Accessed on: 31 Aug. 2019.

TÜMSAVAŞ, Z.; TEKIN, Y.; ULUSOY, Y.; MOUAZEN, A. M. Prediction and mapping of soil clay and sand contents using visible and near-infrared spectroscopy. **Biosystems Engineering**, vol. 177, p. 90–100, 1 Jan. 2019. https://doi.org/10.1016/j.biosystemseng.2018.06.008.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Validation of multivariate calibration models in the determination of sugar cane quality parameters by near infrared spectroscopy. **Journal of the Brazilian Chemical Society**, vol. 18, no. 2, p. 259–266, 2007. DOI 10.1590/S0103-50532007000200003. Available at: http://www.scielo.br/pdf/jbchs/v18n2/a03v18n2.pdf. Accessed on: 3 Jul. 2018.

WANG, H.; PENG, J.; XIE, C.; BAO, Y.; HE, Y. Fruit quality evaluation using spectroscopy technology: A review. **Sensors**, vol. 15, no. 5, p. 11889–11927, 21 May 2015. DOI 10.3390/s150511889. Available at: http://www.mdpi.com/1424-8220/15/5/11889. Accessed on: 12 Aug. 2019.

WANG, J.; NAYAK, S.; KOCH, K.; MING, R. Carbon partitioning in sugarcane (Saccharum species). **Frontiers in Plant Science**, vol. 4, p. 201, 18 Jun. 2013. DOI 10.3389/fpls.2013.00201. Available at: http://journal.frontiersin.org/article/10.3389/fpls.2013.00201/abstract. Accessed on: 13 Apr. 2020.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: A basic tool of chemometrics. 58., 2001. **Chemometrics and Intelligent Laboratory Systems** [...]. [*S. l.*: *s. n.*], 2001. vol. 58, p. 109–130. DOI 10.1016/S0169-7439(01)00155-1. Available at: www.elsevier.comrlocaterchemometrics. Accessed on: 4 Jan. 2019.

3. EVALUATION OF MINIMUM PREPARATION SAMPLING STRATEGIES FOR SUGARCANE QUALITY PREDICTION BY VIS-NIR SPECTROSCOPY

Part of this section was published in Sensors Journal (CORRÊDO, L. de P.; MALDANER, L. F.; BAZAME, H. C.; MOLIN, J. P. Evaluation of Minimum Preparation Sampling Strategies for Sugarcane Quality Prediction by vis-NIR Spectroscopy. **Sensors**, vol. 21, no. 6, p. 2195, 21 Mar. 2021. DOI 10.3390/s21062195).

ABSTRACT

Proximal sensing for assessing sugarcane quality information during harvest can be affected by various factors, including the type of sample preparation. The objective of this study was to determine the best sugarcane sample type and analyze the spectral response for the prediction of quality parameters of sugarcane from visible and near-infrared (vis-NIR) spectroscopy. The sampling and spectral data acquisition were performed during the analysis of samples by conventional methods in a sugar mill laboratory. Samples of billets were collected, and four modes of scanning and sample preparation were evaluated: outer surface ('skin') (SS), cross-sectional scanning (CSS), defibrated cane (DF), and raw juice (RJ) to analyze the parameters soluble solids content (Brix), apparent saccharose in the juice (Pol), Fibre, Pol of cane and total recoverable sugars (TRS). Predictive models based on Partial Least Square Regression (PLSR) were built with the vis-NIR spectral measurements. There was no difference (p-value > 0.05) between the accuracy SS and CSS samples compared to DF and RJ samples for all prediction models. However, DF samples presented the best predictive performance values for the main sugarcane quality parameters and required only minimal sample preparation. The results contribute to advancing the development of on-board quality monitoring in sugarcane, indicating better sampling strategies.

Keywords: chemometrics; proximal sensing; precision agriculture

3.1. Introduction

Near-Infrared (NIR) spectroscopy is a well-established technique to monitor the quality of raw sugarcane received by sugar mills (SEXTON *et al.*, 2020), and consequently, for pricing and trading with producers and growers (PHETPAN; UDOMPETAIKUL; SIRISOMBOON, 2018). Crop quality is estimated based on physicochemical parameters related to physiological composition, such as soluble solids content (Brix), water-insoluble solids (Fibre), and the apparent sucrose in the juice (Pol). Furthermore, all other parameters (purity, Pol of cane, reducing sugars, and total recoverable sugars) are calculated based on the former parameters (RODRIGUES; MAGALHÃES; FRANCO, 2013), from which total recoverable sugars (TRS) are used for the pricing and trading of the raw material. Sugarcane quality parameters are determined by analytical methods and empirical equations described by the National Council of Sugarcane Producers (CONSECANA), which are based on the International Commission for Uniform Methods of Sugar Analysis (ICUMSA). Using calibration methods, it is now possible to obtain some crop quality parameters by NIR spectroscopy (RAMÍREZ-MORALES *et al.*, 2016).

Different wavelength regions of the electromagnetic spectrum can be used in spectroscopy, such as visible (400 to 750 nm), near-infrared (NIR, 750 to 2500 nm), shortwave near-infrared (SWNIR, 750 to 1100 nm), and visible and near-infrared (vis-NIR, 400 to 2500 nm) (CORTÉS *et al.*, 2019; PASQUINI, 2018; WALSH *et al.*, 2020; WORKMAN JR.; WEYER, 2012). However, there is no consensus on the limits between these regions. Interaction

between electromagnetic radiation and matter causes molecular vibrations involving heavy atoms (C, N, O, and S) attached to a hydrogen atom (PASQUINI, 2018). This basic principle has allowed substantial scientific advances to predict organic compounds of agricultural products associated with its quality using vis-NIR spectroscopy as a nondestructive and environmentally friendly analysis technique (CORTÉS *et al.*, 2019). Moreover, several studies have shown promising results when using the technique to predict sugar cane quality in the sugar mill (SOROL *et al.*, 2010; TAIRA; UENO; FURUKAWA; *et al.*, 2013; TAIRA; UENO; KAWAMITSU, 2010; VALDERRAMA; BRAGA; POPPI, 2007a) and for breeding programs (PHUPHAPHUD *et al.*, 2020; PHUPHAPHUD; SAENGPRACHATANARUG; POSOM; MARAPHUM; *et al.*, 2019; TAIRA; UENO; SAENGPRACHATANARUG; *et al.*, 2013).

Despite the advances in industrial sugarcane quality monitoring, spectral methods are still a distant reality for in-field measurements in line with precision agriculture practices. Current proximal sensing technologies applied at the canopy level allow only monitoring crop yield (BRAMLEY, R.G.V., 2009; NAWI; CHEN; JENSEN, 2014). However, some studies have indicated that vis-NIR could also be a viable technology for acquiring quality data of harvested products in real-time during mechanical harvesting (NAWI; CHEN; JENSEN, 2014). The monitoring of crop quality parameters across the field is important to adopt precision agriculture (PA) practices, in which quality maps would show the variability of the crop and help guide site-specific management (BRAMLEY *et al.*, 2013). In this context, mechanical harvest opens a way to obtain a high sampling rate to analyze the quality variability across the field (CORRÊDO *et al.*, 2020). However, some requirements need to be satisfied to use vis-NIR as proximal sensing technology for this purpose: (i) the location of adaptation in the harvester, (ii) development of a sampler system, and (iii) the type of sampling required for analysis. The first two requirements are fundamentally dependent on the last one.

Nawi et al. (2014) indicated that the ideal place for implementing an on-board sugarcane quality monitor would be in the elevator of the harvester, where the sugarcane material is partially cleaned after being processed in the form of billets. In this context, some studies have reported promising results on the prediction of sugarcane Brix from sensor readings made on the outer-surface ('skin') (NAWI *et al.*, 2013) or on cross-sections (NAWI; CHEN; JENSEN, 2013) of sugarcane billets. Furthermore, more recent studies have advanced with on-board vis-NIR spectroscopy sensor applications on the elevator of a sugarcane harvester simulator, i.e., analyzing samples at a distance and in motion (PHETPAN; UDOMPETAIKUL; SIRISOMBOON, 2018; UDOMPETAIKUL; PHETPAN; SIRISOMBOON, 2021). However, Maraphum et al. (2018) and Phuphaphud et al. (2019) reported that the waxy material should be removed from the cane surface for maximum accuracy in the spectral data condition, even though this may be impractical for an embedded system. Associated with this fact, Phetpan et al. (2018) reported on the need to evaluate the potential of the vis-NIR spectroscopy technique with data sets consisting of a larger number of sugarcane varieties. In addition, despite the advantages of nondestructive measurement, there is a lack of basic studies comparatively evaluating various forms of sampling without and with minimal processing, using extensive numbers of samples obtained over the course of a harvest. Thus, the objective of this study was to compare different sugarcane sample types, including billets, defibrated cane, and raw juice, and to analyze the spectral response of each sampling type for the prediction of quality parameters of sugarcane from vis-NIR spectroscopy.

3.2. Material and Methods

3.2.1. Sampling

Variability of sugarcane quality parameters throughout a harvest occurs due to diverse environmental conditions, mainly temperature and precipitation, during the harvest season (CARDOZO; SENTELHAS, 2013). Based on this, the data collection procedure occurred on random periods over six months of the 2019 harvest (June to November) in the same sugar mill, for several varieties of sugarcane. We sought to obtain higher variability of the sugarcane quality parameters from this collection procedure throughout the harvest.

The data collection was carried out in the quality laboratory of a sugar mill. Three hundred and two samples were collected, and different levels of processing were applied. Also, all data acquisition was performed in a controlled temperature environment (20 ± 5 °C), minimizing the effects of sugar degradation by microorganisms. In addition, we collected the daily results of conventional analysis performed by sugar mill quality laboratory over the same months in which the samples were collected for spectral analysis.

The sampling procedure for vis-NIR analysis was carried out simultaneously to the sample preparation for conventional analysis, as described:

1. An oblique probe collected a sample of sugarcane billets in each truckload to proceed with the conventional analyzes of the sugarcane transported to the sugar mill (Figure 8a). Before the laboratory processes the sample, we took a subsample composed of three random sugarcane billets (Figure 8b);
2. The remaining sample of billets was milled in a mechanical knife crusher and homogenized in a mixer (Figure 8c). Then, samples with 500 g and 50 g of the homogenized defibrated sugarcane were collected (Figure 8d) to proceed with the conventional and spectral analysis, respectively;
3. The 500 g sample was pressed in a hydraulic press under constant pressure, at 24.5 MPa (250 kgf cm^{-2}) for 60 s to obtain the raw juice for conventional analysis (Figure 8e). At the same moment, the third sample composed of extracted raw juice was collected for spectral measurements (Figure 8f).

The sample types composed of billets, defibrated cane, and raw juice were prepared (Figure 8g,h) and immediately used for spectral measurements at the sugar mill laboratory (Figure 8i). The remains of raw juice and the bagasse without juice (after pressed) were used for conventional analysis (Figure 8j).

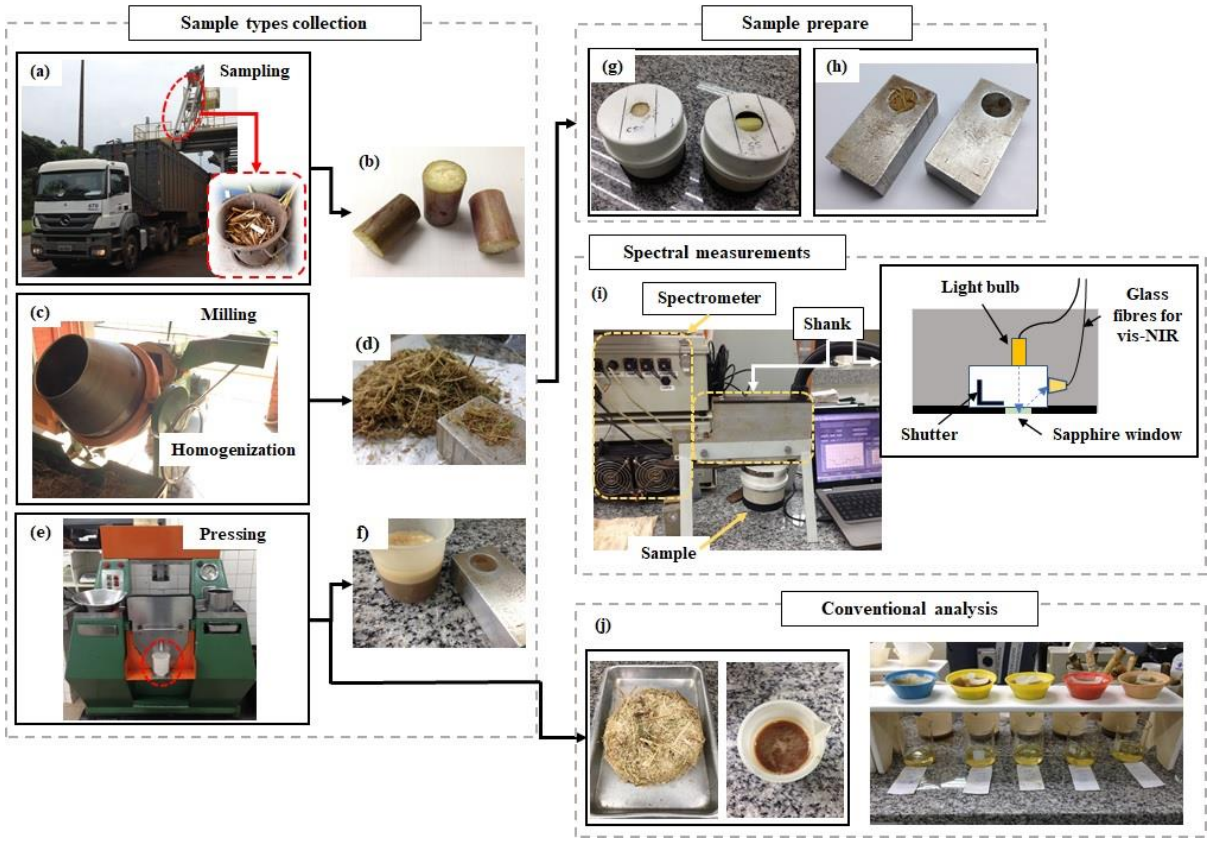

Figure 8. The sequence samples collection and spectral measurements during the preparation steps of samples for conventional analysis. Sampling of sugarcane billets by an oblique probe in the cargo truck (a); sugarcane billets for skin and cross-sectional scanning measurements (b); milling and homogenization of sugarcane to defibrated sample (c); defibration sample (d); pressing of defibrated sample to extracting of juice (e); extracted raw juice (f); prepared samples for vis-NIR spectral measurements: cross-sectional and skin of billets inside pipeline chambers (g), defibrated cane and raw juice (h); Veris vis-NIR spectrometer and internal configuration scheme of the measurement shank (i); fibrous cane residue and extracted raw juice for conventional analysis(j).

3.2.2. Sugarcane quality analysis

All the procedures and equations for sugarcane quality parameters determination followed the standard protocol proposed by the CONSECANA (2015). These protocols agree with the international rules from ICUMSA.

Initially, the soluble solids content (Brix) was determined by pouring raw juice into a refractometer probe (RX-5000α, ATAGO Co Ltd., Tokyo, Japan) with a maximum resolution of 0.1 Brix. Then, 14 g of a mixture composed of a 4:2:1 proportion of Celite (mineral filtering agent), aluminum chloride, and calcium hydroxide, respectively, was added to 200 mL of raw juice homogenized by a magnetic stirrer until the solution was well-mixed. The solution was filtered through filter paper to obtaining clarified juice. A volume of 70 mL was added in a digital polarimeter (Schmidt + Haensch, Polartronic NHZ 8, Berlin, Germany) to the saccharimetric reading. The result was obtained as percentage of apparent sucrose in the juice (Pol).

After the juice extraction, the remaining fibrous cane residue (Figure 1j) was transferred to metal baskets, with holes at the base. The fibrous residue was weighed on a semianalytical balance. Then, samples were maintained in a forced air circulation dryer at a temperature of 105 °C, until constant weight was achieved. The dryer samples were weighed, and the fibre content (insoluble solids) was determined.

The Pol of cane and TRS were calculated for each sample from the previous parameters as described by CONSECANA (2015). The triplicate values obtained from each sample were averaged.

It is important to mention that reducing sugars (fructose and glucose) and purity (apparent sucrose in the soluble solids content) were also determined in the laboratory. However, while these parameters are not the subject of the present study, they were used to calculate TRS by the reference method.

3.2.3. Equipment and sensibility conditions

The spectral measurements were performed with a Veris vis-NIR spectrometer (Veris Technologies Inc., Salina, KS, USA). This equipment was developed for on-the-go soil measurements mounted on a platform, connected to a three-point hitch, and pulled by a tractor [27]. However, it may also be used in bench mode. The equipment consists of a CCD array spectrometer (USB4000, Ocean optics, Largo, FL, USA), measuring wavelengths between 373 and 1011 nm, and an InGaAs photodiode-array spectrometer (C9914GB, Hamamatsu Photonics, Hamamatsu, Japan), with a spectral range between 1170 and 2222 nm. The system presents a resolution of around 5 nm. Spectral measurement acquisitions were performed via a sapphire window in the lower of the shank using a tungsten halogen lamp as an electromagnetic energy source (Figure 1i). Each spectrum recorded by the equipment software (Veris spectrophotometer software V1.79) corresponded to the average of 20 spectral readings. The inside shutter is operated automatically to obtain dark and reference spectra before each analysis. Four external references with different grey levels provided by the manufacturer were used for spectral calibration of the spectrometer before the analysis. The spectral data were stored as absorbance units.

In order to identify trends and regions with importance to the prediction as well to evaluate the sensitivity of the equipment, optimal prediction conditions of the compounds of interest were simulated. For that purpose, sucrose solutions were prepared at different concentrations for analysis by the equipment. For the preparation of the solutions, sucrose P.A. was used. (Synth® Brasil Ltda, São Paulo, Brazil). The sucrose was weighed on an analytical balance with accuracy of three decimal places and diluted in 100 mL of distilled water with the aid of a magnetic stirrer until complete dilution. After dilution, a pocket digital refractometer (Pocket PAL-1, ATAGO, Japan) was used to read the soluble solids content of the solution (Brix). The prism of the equipment was filled with solution (approximately 1 mL) and the reading was taken in triplicate, for 5 subsamples. The average result of the readings was recorded as the Brix value (%) for each solution. The same samples were analyzed with the Veris vis-NIR equipment, with readings in triplicate. For each subsample, the average spectrum was recorded. The procedures were repeated 5 times for each of the 13 concentrations prepared, for a total of 65 spectra obtained. Table 5 shows the concentrations of each prepared solution and the respective Brix values.

Table 5. Sucrose solution concentration and respective mean Brix values.

Concentration [g (100 mL)$^{-1}$]	Brix (%)	Concentration [g (100 mL)$^{-1}$]	Brix (%)
0	0	35	26,4
5	5,0	40	29,2
10	9,3	45	31,3
15	13,4	50	33,4
20	16,9	55	36,0
25	20,1	60	37,9
30	23,8		

3.2.4. Acquisition of spectral data

The three billets of each sugarcane sample were cut transversely at both extremities, and their skin was lightly wiped with paper to remove residues from harvest (Figure 8b). An ad hoc dark chamber was constructed with PVC pipes and foam to accommodate the billets, and its inside was painted matte black. Also, magnets were placed on the cover of the chamber, next to the hole through which the spectral scans were performed, aiming to fix the chamber on the reading shank of the spectrometer. This device standardizes the distance between the sapphire window of the spectrometer and the sugarcane billets and removes interference from external lighting (Figure 8g).

The spectral scan method on billets was adapted from Nawi et al. (2013) and Nawi, Chen, and Jensen (2013) and Phuphaphud et al. (2019). The spectral scans were performed at three equidistant points (around 120°) on the skin of each billet. Furthermore, the cross-sectional scanning of billets was performed in triplicate in each cross-sectional surface of each billet, only changing the position after each reading. Therefore, each sample type measurement, skin scanning (SS) and cross-sectional scanning (CSS) of billets, was represented by an average of nine successive scans.

The spectral measurements of defibrated cane and raw juice were performed in the same manner. A recipient available from the equipment itself with a volume of around 3 mL was used. The recipient was filled with sample (defibrated cane or extracted raw juice, Figure 8h), and the spectral measurements were performed in triplicate. Three replicates were performed for each sample. Thus, the average of nine spectral readings of defibrated cane (DF) and nine spectral readings of raw juice (RJ) were recorded.

3.2.5. Spectral preprocessing

Data preprocessing steps were performed to remove or minimize the sources of spectral variabilities, such as noise present in the dataset, which was not related to the analytical signal (OLIVERI *et al.*, 2019; PASQUINI, 2018).

Firstly, the spectra were preprocessed using standard normal variate (SNV) (BARNES; DHANOA; LISTER, 1989) to eliminate the deviations caused by particle size and scattering, which centers each spectrum on its mean and then scales it by its standard deviation. Also, the second derivative based on the Savitzky-Golay algorithm (SAVITZKY; GOLAY, 1964) was applied, with a window size of 11 points and second-order polynomial fitting to minimize hurdles such as baseline shifts drifts and to remove high-frequency noise from a spectrum and improve the signal-to-noise ratio (OLIVERI *et al.*, 2019). After the preprocessing of the spectral data, Pareto scaling (PS) was applied to variables, which is the most applied scaling method in infrared data (GERRETZEN *et al.*, 2015). The method centered all variables at their means, and then divided them by the square root of the standard deviation.

3.2.6. Multivariate analysis

Firstly, the spectral data of the four sampling conditions were concatenated. Then, the data was divided into calibration (75%, 227 samples) and external validation (25%, 75 samples) data sets, based on the Kennard-Stone method (KENNARD; STONE, 1969). This procedure allowed to obtain the same samples for calibration and

external validation data set for both sample types evaluated. The spectral measurements were used to build predictive models for sugarcane quality parameters based on Partial least square regression (PLS) (NAWI *et al.*, 2013).

The models were calibrated using the venetian blinds cross-validation method with 10 splits. The optimal PLS models were determined based on the lowest number of latent variables (LV), in which the root mean square error of cross-validation (RMSECV) was not significantly higher than the minimum RMSECV (MARAPHUM *et al.*, 2018). The root mean square error (RMSE) was calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})2}{n}} \quad (3.1)$$

where n is the number of samples, y_i is the reference measurement of sample i, and $\widehat{y_i}$ is the estimated result for sample i.

For the sensibility evaluation described in the section 3.3.3, the sucrose solution spectra data set was segmented into a calibration set (43 samples) and a validation set (22 samples) using the Kennard-Stone and calibrated using leave-one-out cross validation.

The outliers were evaluated during the calibration step for the reference lab values and spectral data. The presence of outliers in the spectral data was evaluated by the 'influence plots' based on high leverage and unmodeled residuals by Hotelling T^2 and Q statistics, respectively (PASQUINI, 2018). Samples with high values in both cases, at 5% of significance level, were considered outliers and removed from the spectral data set. On the other hand, outliers in reference data were evaluated by the root mean square error in calibration (RMSEC) values. Samples that presented errors in prediction greater than ± 3 × RMSEC were considered outliers and removed from the data set (VALDERRAMA; BRAGA; POPPI, 2007b). The external validation samples were considered unknown samples. In this way, the outliers were evaluated only for the spectral data set. The process was carried out at most three times in the calibration step, as recommended by ASTM E1655-7 (ASTM E1655-17, 2017).

The model accuracy was evaluated based on the RMSE for calibration, cross-validation, and prediction (RMSEC, RMSECV, RMSEP, respectively). Prediction performance was evaluated based on the determination coefficient (R^2) for calibration and prediction (R^2c and R^2p, respectively), and the ratio of performance to the interquartile range (RPIQ), which is calculated by the ratio between the interquartile difference and the RMSEP. Also, a randomization test (VAN DER VOET, 1994) with 0.05 significance level of probability was performed. The aim was to compare the accuracy of regression models using different sugarcane sample types in the validation set. The hypothesis evaluated were:

- Null hypothesis (H_0): $\mathrm{RMSEP}_{\text{sample type 1}} = \mathrm{RMSEP}_{\text{sample type 2}}$ (accuracy is similar);
- Alternative hypothesis (H_1): $\mathrm{RMSEP}_{\text{sample type 1}} \neq \mathrm{RMSEP}_{\text{sample type 2}}$ (accuracy is not similar).

An advantage of this test is its simplicity and the fact that assumptions about normality or homoscedasticity of the data are not required (distribution-free) (SANTANA *et al.*, 2019). More details about this test, included an algorithm script, can be found in Olivieri (2015).

Moreover, the variable importance for the projection (VIP) was calculated to verify the wavelengths with a more significant impact on the external validate models (FRANCESCHINI *et al.*, 2018) for each sample type. The VIP was calculated as follows:

$$\mathrm{VIP_j} = \sqrt{p \sum_{k=1}^{h}\left[Z\left(\frac{w_{jk}}{||w_k||}\right)^2\right] . \left(\sum_{k=1}^{h} Z\right)^{-1}} \quad (3.2)$$

where VIP is the variable importance for projection (dimensionless), j is a specific wavelength (nm), *p* is the number of wavelengths (dimensionless), h is the number of latent variables (dimensionless), Z is the fraction of

variance in the prediction explained by the latent variable (dimensionless), and w is the loading weight (dimensionless).

All models, routines, and data processing were performed in Matlab R2015a (The MathWorks, Natick, MA, USA) and PLS Toolbox 8.9 (R8.9.1; Eigenvector Research, Wenatchee, WA, USA).

3.3. Results and Discussion

3.3.1. Exploratory data analysis

There was no obvious pattern differentiating sucrose concentrations along the spectra. The mean spectral behavior of sucrose solutions can be seen in Figure 9a.

As sucrose was the only soluble solid available in the prepared samples, the model with only five latent variables allowed to obtain an RMSEp of 0.86 % and R^2p of 0.99 was obtained (Figure 9b), evidencing the sensitivity of the equipment to identify the compound of interest in the samples. The first two latent variables were responsible for capturing 95.14 % of the data variance (Figure 9c). Moreover, the calculation of VIP scores allowed the identification of the spectrum region with the greatest influence on the modeling between 943 and 1324 nm (Figure 9d).

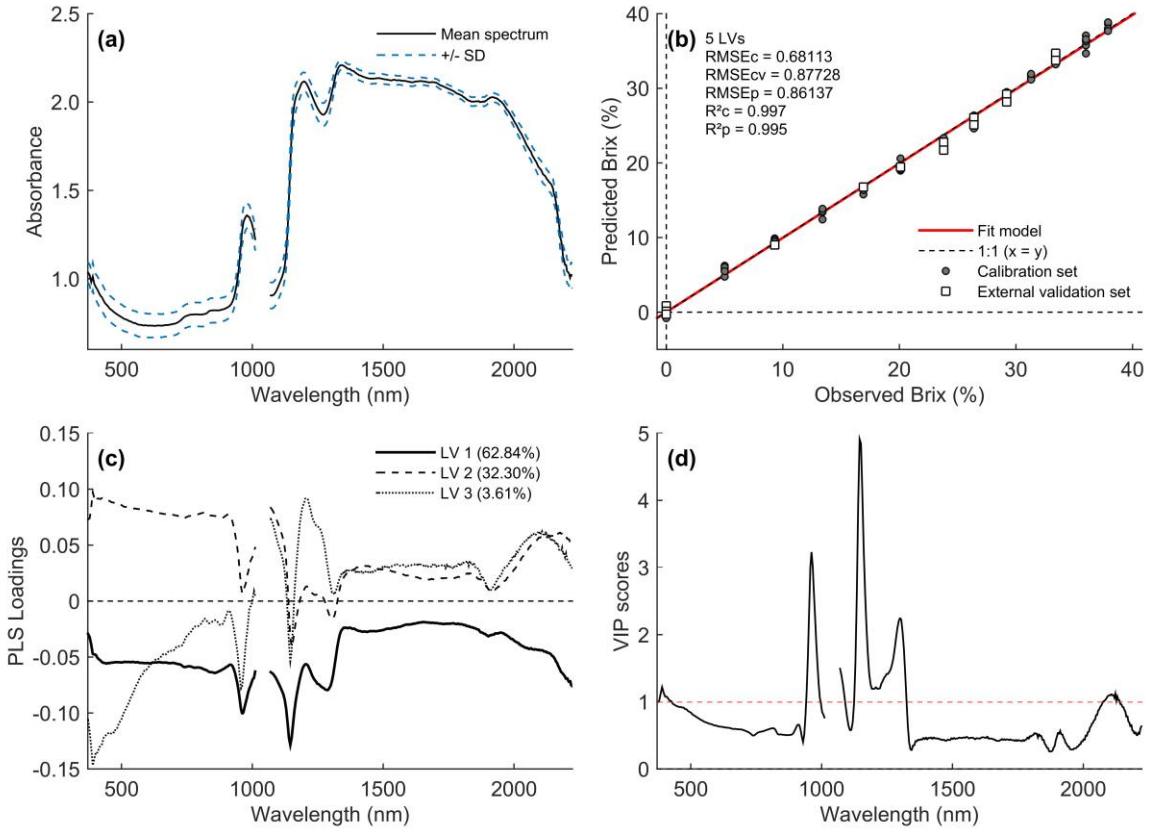

Figure 9. (a) Average vis-NIR spectral behavior of different concentrations of sucrose solutions. (b) Plots of observed versus predicted sucrose values from vis-NIR PLS model. (c) Three first Partial least squares loadings for Brix prediction. (d) Variable Importance in Projection (VIP) of Brix prediction model.

3.3.2. Overview of Sugarcane Quality Reference Data and vis-NIR Spectral Measurements of Different Sample Types

From the daily results of analyses performed by conventional methods at the mill, it was possible to characterize the variation of the main parameters determined analytically (Brix, Pol, and Fibre), as well as for TRS, throughout the months in which the experiment was performed (Figure 10).

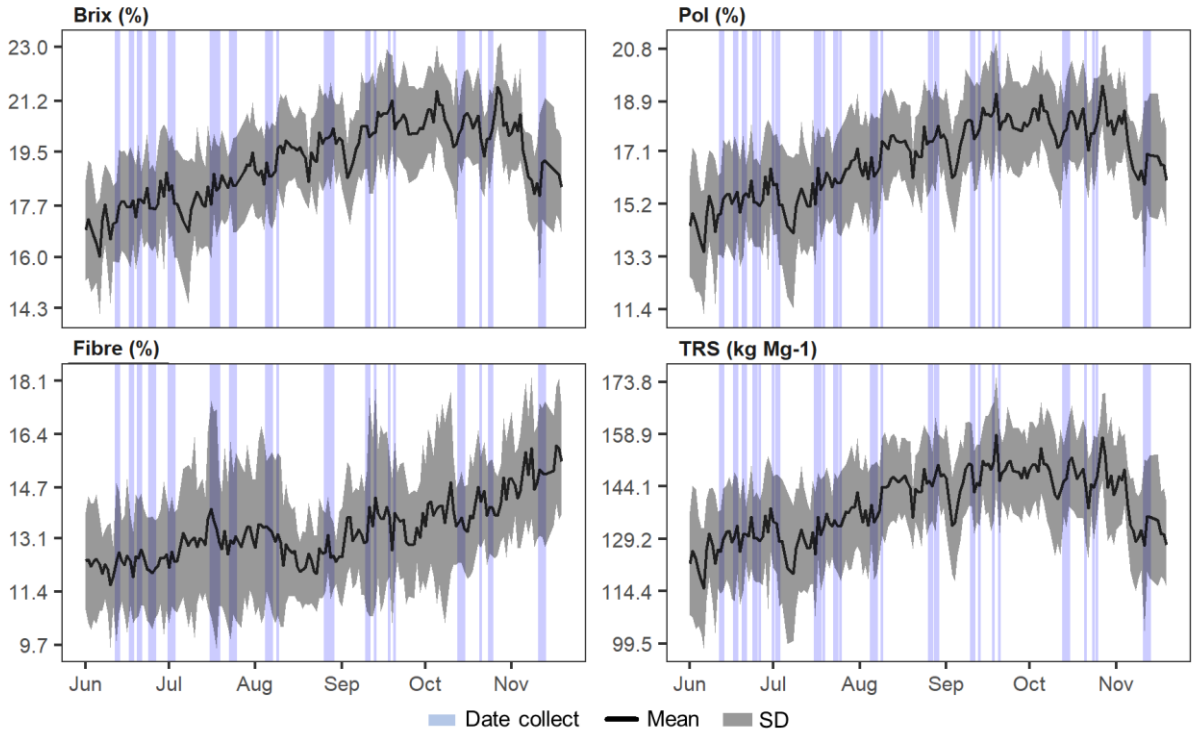

Figure 10. Mean and standard deviation (SD) of annual variation of sugarcane quality parameters, Brix, Pol, Fibre, and total recoverable sugars (TRS); and spectral data collection periods (vertical bars).

It is possible to observe an increasing trend in all parameters from June to October. Afterward, there is a tendency to decrease, except for Fibre. Weather is highly influential on sucrose storage (CARDOZO; SENTELHAS, 2013). In the months corresponding to autumn (June) and winter (June to September), water stress and cooler temperatures contribute to the reduction of vegetative crop growth and favor sucrose storage (PAGANI *et al.*, 2017). With the beginning of spring (September/October) and the beginning of the rainy season, the vegetative growth of the crop is resumed, and the reserves are consumed. The sample acquisition on different periods (vertical bars in Figure 10) throughout the harvest allowed us to obtain data including different stages in this variation. The effect of this variability was reflected in the range of all samples collected during the experiment (Table 6).

On a first view, the Kennard-Stone method provides a representative calibration data set, with external validation data set between its range (Table 6). The sample acquisition method provided a satisfactory variability of data, as expected; TRS varied from 86.94 to 173.80 kg of sugar per Mg of cane.

The distribution of all quality parameter values had wide distribution (Figure 11). Fibre content did not correlate with any other parameters analyzed, with values varying from −0.16 to 0.13 ($p < 0.05$). On the other hand, the other parameters showed a positive correlation higher than 0.94 ($p < 0.05$). The highest correlation was observed

between Pol of cane and TRS, close to 1.00 ($p < 0.05$). Higher correlation values are observed between the TRS with parameters analytically determined such as Brix and Pol (0.94 and 0.96, respectively, $p < 0.05$). The correlation values for these attributes are firstly explained by the composition of the soluble solids content of sugarcane, measured by Brix, in which the largest proportion corresponds to sucrose (about 15–18%), measured by Pol (MANCINI *et al.*, 2012; WANG, Jungang *et al.*, 2017), and reducing sugars (fructose and glucose) in a smaller proportion (about 0.5%) (SOLOMON, 2009); note that the determination of reducing sugars was not the objective of the present study.

Table 6. Descriptive statistics of the reference results for the sugarcane quality attributes of all samples, calibration, and external validation data sets.

Parameter	unit	Mean ± SD	Median	Range	p_{25}	p_{75}	SEL
All samples (n = 302)							
Brix	%	18.95 ± 1.71	18.99	13.08–23.42	17.80	20.01	0.03
Pol	%	16.67 ± 1.90	16.66	10.78–21.20	15.41	17.95	0.01
Fibre	%	13.29 ± 1.79	12.90	7.22–20.08	12.07	14.33	0.07
Pol of cane	%	13.80 ± 1.56	13.91	8.40–17.56	12.92	14.78	0.01
TRS	kg Mg^{-1}	137.66 ± 14.48	138.66	86.94–173.80	129.75	146.84	1.12
Calibration set (n = 227)							
Brix	%	18.86 ± 1.66	18.80	13.08–23.42	17.79	19.98	-
Pol	%	16.54 ± 1.86	16.55	10.78–21.20	15.38	17.78	-
Fibre	%	13.31 ± 1.89	12.83	7.22–20.08	12.05	14.41	-
Pol of cane	%	13.69 ± 1.52	13.79	8.40–17.56	12.83	14.60	-
TRS	kg Mg^{-1}	136.65 ± 14.07	137.01	86.94–173.80	128.99	145.19	-
Validation set (n = 75)							
Brix	%	19.24 ± 1.85	19.59	13.55–23.05	18.06	20.61	-
Pol	%	17.06 ± 1.98	17.33	11.24–20.90	15.73	18.54	-
Fibre	%	13.23 ± 1.44	13.02	10.49–17.15	12.24	14.16	-
Pol of cane	%	14.14 ± 1.64	14.34	8.96–17.14	13.27	15.47	-
TRS	kg Mg^{-1}	140.76 ± 15.36	142.59	92.16–169.02	132.35	152.44	-

SD: standard deviation; p_{25}: lower quartile; p_{75}: upper quartile; SEL: standard error of laboratory; RS: reducing sugars; TRS: total recoverable sugar.

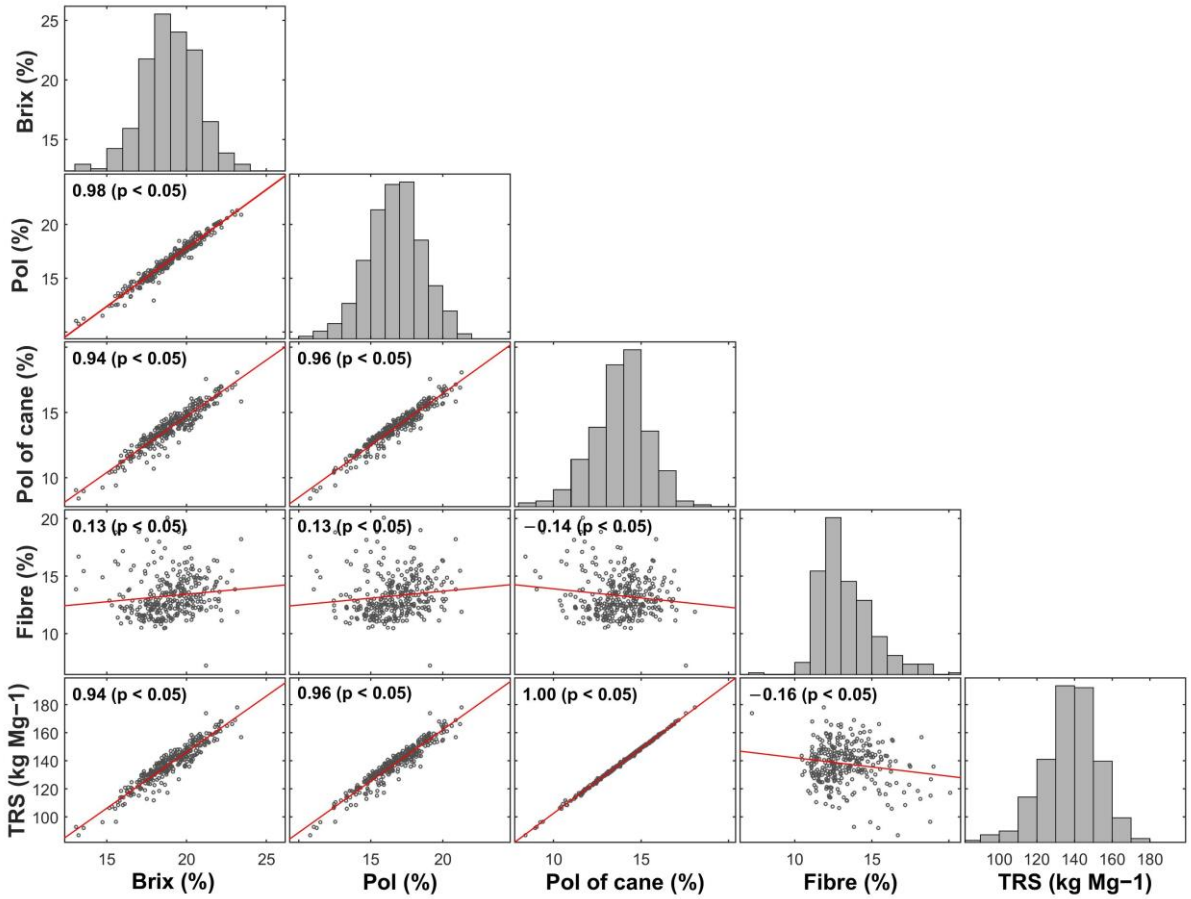

Figure 11. Correlogram of sugarcane quality parameters with frequency distributions on the principal diagonal, Pearson's correlation coefficient and respective *p*-values, and the correlation trend line.

The vis-NIR raw spectral data obtained for 302 samples of each sugarcane sample type are shown in Figure 12. The spectral data were evaluated to identify possible spectral errors (BAZAR; KOVACS; TSENKOVA, 2016). We observed a noisy aspect in the region corresponding to the visible spectrum (400 to 698 nm), mainly for SS samples. This effect may have been attributed to the influence of skin colors of billet samples, which were obtained for several different sugarcane varieties (Annex 1, Table 13), or soil residues from the harvest present in the RJ samples. Therefore, this spectral region was removed from the data set. Phuphaphud et al. (2020) observed the same effect due to the skin color of sugarcane billets. Also, based on the evaluation of the coefficient of variation (CV) obtained for each spectral band, the last spectral bands showed high CV concerning their neighbors and were also removed from the dataset, similar to the procedure performed by Franceschini et al. (2018) in a study on the external effects on the spectral reading of vis-NIR of soil samples using the same equipment. Thus, only bands in the spectral range between 699 and 1010 nm and between 1070 and 2153 nm (303 spectral bands) were retained.

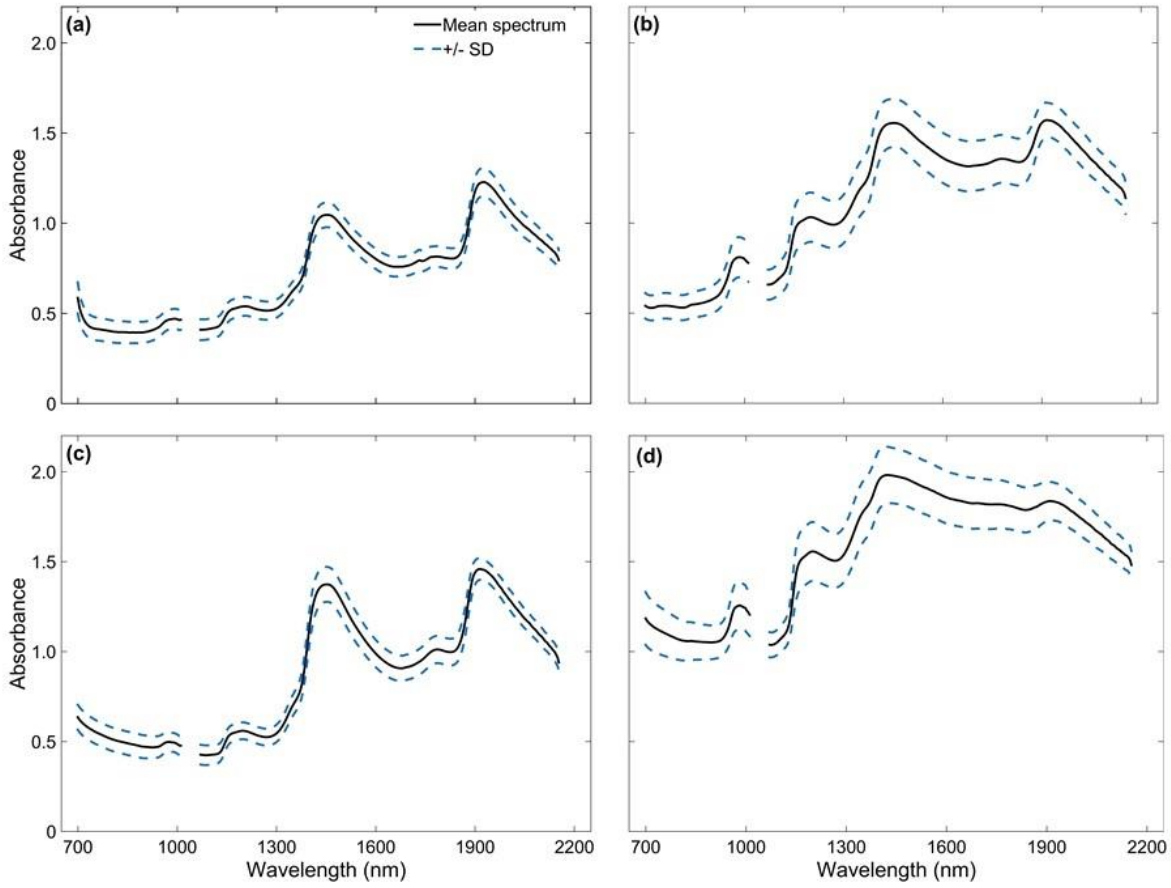

Figure 12. vis-NIR mean spectra and standard deviation (SD) of all 302 sugarcane samples for (a) skin and (b) cross-sectional scanning of billets, (c) defibrated, and (d) raw juice samples.

A PCA analysis performed an exploratory overview of the data structure. The spectral data were only mean-centered, and the classes were identified by sample type. Two principal components, PC1, and PC2, explained 98.6% of the data variance (Figure 13). The first component explained 96.8% of the data variance. The data structure was different for each sample type, as can be seen from ellipses illustrating the majority of samples. However, a first overview allowed us to verify the greater difficulty in explaining the variance of less processed samples, such as samples obtained by spectral readings in the skin (SS) and cross-sectional (CSS) of the billets, than processed samples, such as raw juice samples (RJ).

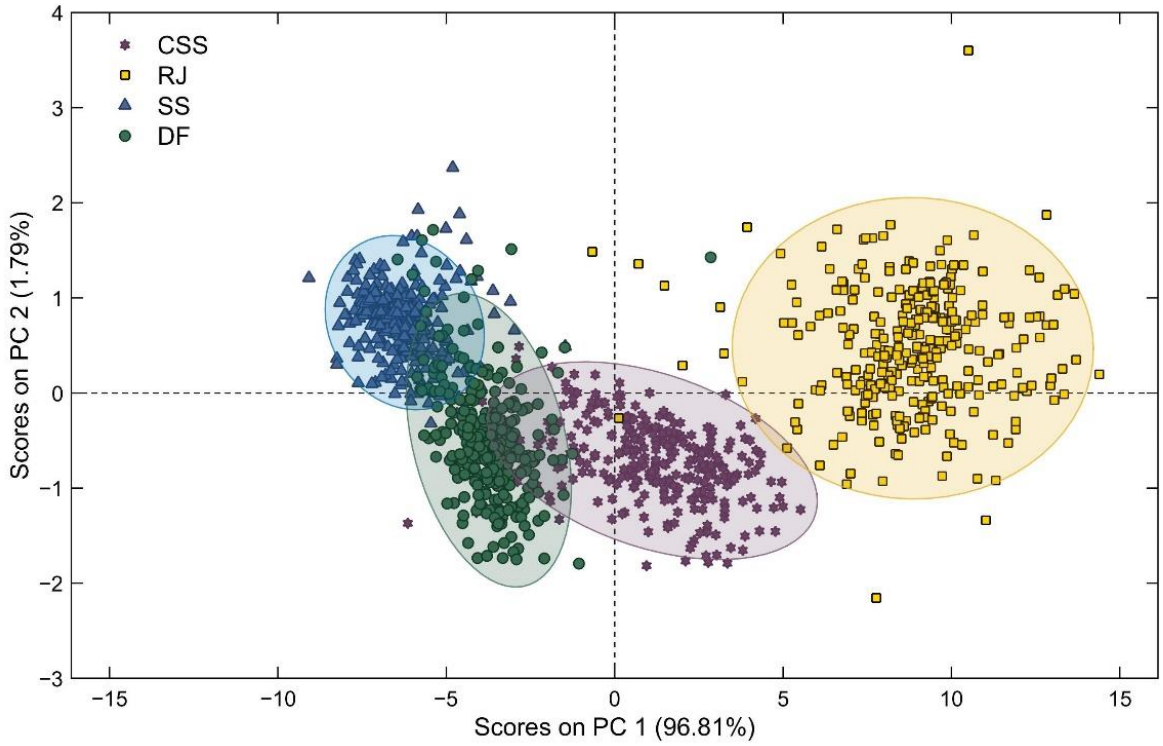

Figure 13. PCA score plot for the sugarcane sample types analyzed. SS-skin scanning of billets; CSS-cross-sectional scanning of billets; DF-defibrated samples; RJ-raw juice samples.

Vis-NIR spectroscopy may be used in several applications, including the classification of sugarcane varieties, with promising results (STEIDLE NETO *et al.*, 2018). The same authors showed that the spectral regions between 650 and 750 nm, corresponding to the visible spectrum, was the most suitable for sugarcane discrimination. The principal component analysis for the four sample types individually showed that the scatter plots were not categorized into groups based on sugarcane varieties (Annex 1, Figure 30). These results indicated that the sample set composed for many different varieties did not affect the spectral characteristics between each variety for both sample types. A similar effect was observed by Phuphaphud et al. (2020) when evaluating the classification of three varieties. Therefore, the present study was conducted for all varieties combined.

The vis-NIR technique principle is based on the detection of compounds and molecules through their molecular vibration states (PASQUINI, 2018). Different varieties naturally have different concentrations of parameters such as sucrose and Fibre according to genetics. Furthermore, for all of them, the plant matrix is essentially composed of water (75–82%), insoluble solids content (Fibre, 10–18%), and soluble solids (Brix, 18–25%), which are composed of non-sugars (1–2%), sucrose (14–24%), and reducing sugars (0–1.5%) (COSTA *et al.*, 2021). However, the prediction of quality parameters related to chemical compounds of interest should be independent of sugarcane varieties.

3.3.3. Prediction Performance of Models Based on Different Sugarcane Sample Types

Figure 14 presents scatter plots showing reference versus predicted values of sugarcane quality parameters. There was an underestimation of high values and overestimation of lower values for all attributes and sample types evaluated. However, this effect was more intense for the less prepared sampling condition, such as SS and CSS. Also, overall, the residuals showed no trend (Annex 1, Figure 31).

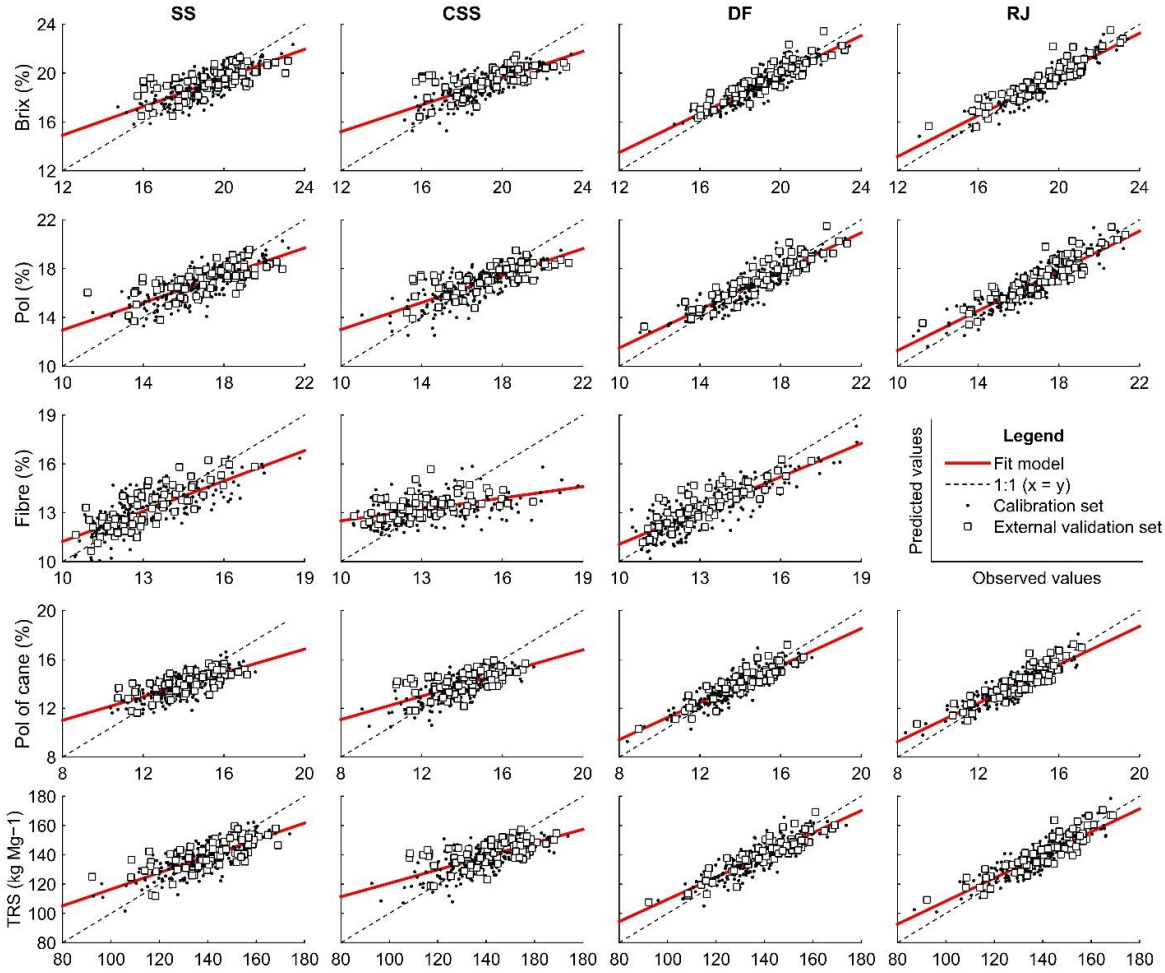

Figure 14. Plots of observed values versus predicted sugarcane quality values from vis-NIR by skin (SS) and cross-sectional (CSS) scanning of billets, defibrated (DF), and raw juice (RJ) samples. Brix, Pol, Fibre, and Pol of cane are in percentage, and TRS values are in kg Mg^{-1}.

More LVs were necessary to explain the variance of the data for models constructed from SS samples (between 7 and 10) than those obtained to predict the same parameters from other sample types (Table 7). Also, it could be observed that SS and CSS did not show similar accuracy (p-value < 0.05) for Fibre, Pol of cane, and TRS (Table 8). Moreover, the prediction performance results for these parameters by these sample types were worse than the performance results obtained by DF and RJ samples (Table 8). The RPIQ values for TRS were 40% higher on average than those for SS and CSS samples, for example. Furthermore, the model accuracy observed between sampling methods for all the other conditions was statistically nonsignificant (p-value > 0.05).

The model performance for DF and RJ samples was equivalent for practically all parameters evaluated. There was no significant difference between the model's accuracy (p-value > 0.05) and very close values of R^2p and RPIQ. Moreover, from DF samples, it was possible to obtain a satisfactory performance to predict Fibre content; this was not possible for RJ samples. On the other hand, the models performed for SS samples presented a higher number of LV than for a prepared sample. The model performance for predicting parameters related to sucrose (Brix, Pol, Pol of cane, and TRS) was not satisfactory, with R^2p and RPIQ below 0.5 and 2.0, respectively, except for Fibre prediction. The prediction results from SS samples for Fibre were close to those obtained for DF samples, as shown by the values of R^2p and RPIQ. However, the results were less promising than those obtained by Phuphaphud et al. (2019), which obtained the following results: maxima of 0.81 for R^2p and 0.63 for RMSEP. Although Fibre content is an important attribute for sugarcane quality determination, it is not essential for sucrose estimation. Fibre content has no relation with some important attributes, such as Brix and Pol, and only minimally impacts TRS calculation. The prediction of this parameter is important for producing energy cane and breeding programs, as in work developed by Phuphaphud et al. (2019).

Table 7. Figures of merit for the PLSR models for all studied sugarcane quality attributes and sample types.

Attribute	Sample Type	LV	RMSEC [a]	RMSECV [a]	RMSEP [a]	R^2c	R^2p	RPIQ
Brix	SS	9	0.92	1.10	1.29	0.64	0.48	1.98
	CSS	6	0.95	1.04	1.38	0.62	0.41	1.85
	DF	7	0.67	0.75	0.84	0.81	0.80	3.05
	RJ	8	0.64	0.83	0.75	0.85	0.85	3.39
Pol	SS	8	1.09	1.26	1.42	0.60	0.48	1.98
	CSS	6	1.09	1.19	1.44	0.61	0.44	1.95
	DF	7	0.82	0.93	0.87	0.79	0.83	3.24
	RJ	7	0.80	0.97	0.90	0.82	0.81	3.12
Fibre	SS	10	1.02	1.29	0.87	0.59	0.65	2.22
	CSS	4	1.45	1.50	1.27	0.24	0.23	1.51
	DF	5	0.93	1.04	0.82	0.69	0.69	2.36
	RJ[b]	-	-	-	-	-	-	-
Pol of cane	SS	7	0.95	1.07	1.13	0.52	0.46	1.94
	CSS	5	1.01	1.09	1.27	0.52	0.31	1.73
	DF	7	0.73	0.84	0.72	0.76	0.81	3.04
	RJ	7	0.71	0.85	0.72	0.78	0.81	3.07
TRS	SS	9	8.57	10.27	10.86	0.60	0.50	1.85
	CSS	5	9.49	10.17	11.86	0.50	0.32	1.69
	DF	7	6.50	7.51	6.71	0.76	0.82	2.99
	RJ	7	6.38	7.95	6.79	0.78	0.81	2.96

SS: skin scanning of billets samples; CSS: cross-sectional scanning of billets samples; DF: defibrated samples; RJ: raw juice samples. LV: latent variable. RMSEC: root mean square error of calibration. RMSECV: Root Mean Square Error of Cross-Validation. RMSEP: Root Mean Square Error of Prediction. R^2c: calibration coefficient of determination. R^2p: prediction coefficient of determination. RPIQ: Ratio of performance to interquartile distance. [a] values for Brix, Pol, Fibre, and Pol of cane are in percentage and TRS in kg Mg^{-1}. [b] the fibre content was not determined from raw juice samples.

Table 8. *p*-Values of randomization test of external validation set for all compared sugarcane sample types.

Binary Combination (Sample Types)	Sugarcane Quality Parameters				
	Brix	Pol	Fibre	Pol of cane	TRS
SS vs. CSS	0.104	0.116	<0.001	0.036	0.008
SS vs. DF	1.00	1.00	0.667	1.00	1.00
SS vs. RJ	1.00	1.00	-	1.00	1.00
CSS vs. DF	1.00	1.00	1.00	1.00	1.00
CSS vs. RJ	1.00	1.00	-	1.00	1.00
DF vs. RJ	0.879	0.344	-	0.606	0.502

SS: skin scanning of billets; CSS: cross-sectional scanning of billets; DF: defibrated samples; RJ: raw juice samples; TRS: total recoverable sugar.

Some models developed for CSS samples were similar to those developed for SS, as for Brix and Pol prediction. However, its predictive performance was lower than those obtained for Fibre, Pol of cane, and TRS predicted by SS samples, with worse results for R^2p and RPIQ. In a first investigation, Nawi et al. (2013) obtained values of 0.87 for R^2p. The excellent performance of this index can be explained by the method of data acquisition adopted by the authors, with individualized samples according to the stem portion (lower, middle, and upper portion) and only three varieties of cane. Sucrose accumulation occurs in an ascending manner, with more accumulation in the lower portion and less in the internodes of the upper portion, close to the leaves (UYS *et al.*, 2007; WANG *et al.*, 2013; WANG *et al.*, 2017). Therefore, samples composed of different sections resulted in more variability in quality parameters. However, if we analyze the characteristics of a sugar cane harvester, after the stems pass through the chopper roll system, the distinction between portions of the sugarcane stem is not viable.

The RPIQ values for the SS method were higher than those obtained by the CSS method. The SS method on billet samples on the harvester conveyor would be the most practical method, due to the better operability of sample acquisition in that portion of the harvester (NAWI; CHEN; JENSEN, 2014). However, several external factors must be considered to measure quality attributes by the SS method. A critical one is the constitution of the sugarcane skin itself, as various colors depending on the variety, black and white waxy material, and organic compounds may be present (INARKAR; LELE, 2012; PHUPHAPHUD; SAENGPRACHATANARUG; POSOM; WONGPICHET; *et al.*, 2019).

There are common waxy materials on the cane surface that affect vis-NIR measurements by the SS method. Maraphum *et al.* (2018) evaluated the effect of the waxy material on the cane surface to eliminate or avoid getting low accuracy of the models for Pol measurements. They obtained RMSEP values were around 1.20 to 1.50%, i.e., close to those found by the present study. The authors concluded that spectra acquisition by removed-wax samples was convenient for the measurement of Pol. However, other compounds could affect vis-NIR spectroscopy measurements, such as cellulose and lignin (WORKMAN JR.; WEYER, 2012).

Overall, the predictive performance results of the models (based on the R^2p and RPIQ) indicate that DF and RJ samples presented similar performance and provide the best results. However, models built from DF samples require less preparation, i.e., by avoiding juice extraction, making them more attractive for an on-board system. Moreover, CSS samples presented worse performance than all other samples. SS samples presented higher values of R^2p and RPIQ than CSS samples for all quality parameters. Furthermore, SS samples not showed difference (*p*-value > 0.05) of accuracy (RMSEP) with models built from DF samples, but worse results for performance (R^2p and RPIQ), except for Fibre. Possibly DF results were satisfactory due to the exposure of the internal constituents, which overlapped concerning the waxy material that becomes visually negligible. On the other hand, the organic compounds in the sugarcane outer-surface may have interfered in the prediction models. Future studies using

nonlinear processing methods (SEXTON *et al.*, 2018) or advanced filtering methods, with orthogonalization of unwanted signals concerning the compounds of interest (ROGER; BOULET, 2018), may help in increasing the predictive performance of the models, which is more interesting for an on-board system.

3.3.4. Variable Influence on the Models

VIP scores were used to describe the importance of each wavelength to the prediction of the main sugarcane quality parameters (Figure 15). The identification of the most important bands in the prediction is fundamental for the development of specific and more cost-effective equipment to measure a given chemical compound (PASQUINI, 2018).

As a vibrational spectroscopy technique, the interaction between the vis-NIR electromagnetic radiation and the matter of the sample could be interpreted mainly by overtones and combinations of vibrational modes involving C-H, O-H, and N-H chemical bonds (OSBORNE, 2006; PASQUINI, 2018). VIP values greater than 1.0 indicate variables with greater influence on the models, and VIP values between 0.8 and 1.0 indicate the moderately influential variables. All variables with VIP smaller than 0.8 are insignificant to the predictions (STEIDLE NETO; TOLEDO; *et al.*, 2017). Some substantial similarities could be observed for different samples and quality parameters on a first overview.

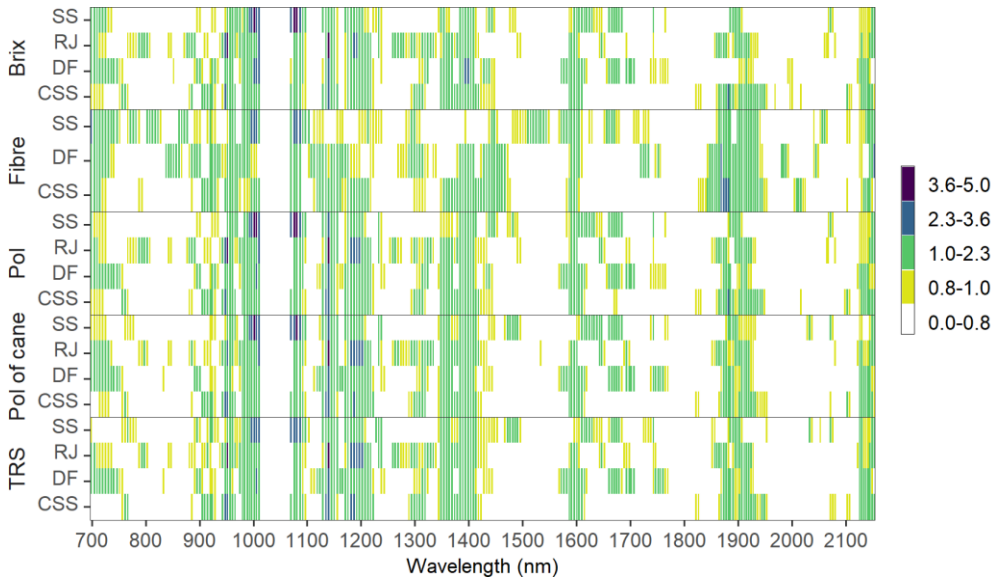

Figure 15. Heatmap of Variable Importance in Projection (VIP) for models used to predict Brix, Pol, Fibre, Pol of cane, and TRS based on different spectral sample types datasets. SS: skin scanning of billets; CSS: cross-sectional scanning of billets; DF: defibrated samples; RJ: raw juice samples; TRS: total recoverable sugars.

At around 960 nm, there is a small interval with high values of VIP (higher than 2.3), especially related to the prediction of Brix, Pol, Pol of cane, and TRS by CSS and RJ samples, corresponding to the second and third overtone of O-H and C-H stretching, respectively (GOLIC; WALSH; LAWSON, 2003). The region between 980 and 1030 nm can be regarded as an important contributor to quality predictions (VIP higher than 1.0). Between 972 and 1009 nm there is a characteristic signal related to saccharides and the third overtone of O-H (WORKMAN JR.; WEYER, 2012). This signal is more expressive for parameters determined for SS samples. This spectral range could

be associated with cellulosic fibres, which explained the higher VIP values to SS samples. Similar observations were found by Phuphaphud *et al.* (2020) to predict commercial cane sugar from growing cane stalks for breeding programs using vis-NIR spectroscopy.

At 1139 nm, there is a small band with VIP values higher than 1.0 for all parameters predicted for four sample types, except for Fibre content. On the interval between 1100 and 1230 nm occurs the second vibrational frequency overtones associated with C-H stretching (GOLIC; WALSH; LAWSON, 2003). Also, at around 1170–1180 nm, there are VIP values higher than 1 for Fibre predicting, mainly by DF and CSS samples. In this region, the third overtone of C-H and unsaturated C=C double bonds are typically associated with Fibre, such as lignin (WORKMAN JR.; WEYER, 2012).

At 1360 nm, there is another expressive region with high VIP values, possibly related to C-H combinations and the O-H first overtone, respectively (OSBORNE, 2006). Then, at 1600 nm, there are highly similar VIP values possibly associated with to first vibrational frequency overtone of C-H stretching (GOLIC; WALSH; LAWSON, 2003). Another region shows a high contribution, with VIP values higher than 1, for Fibre predicting between 1850 and 1900 nm, mainly for CSS and DF samples. Around 1820 nm occurs the effects of O-H stretching associated with two combinations of C-O stretching commonly associated with Fibre as cellulosic (WORKMAN JR.; WEYER, 2012). This interval is lower and with lower VIP values for SS samples, possibly due to the waxy effect on the near-infrared signal (MARAPHUM *et al.*, 2018). Finally, in the last bands of the spectra, after 2100 nm, the intensity of VIP values is similar for all predicted attributes from any sample type due to O-H bending and C-O stretching combination (WORKMAN JR.; WEYER, 2012).

The scores of the models for all measured quality parameters are displayed by their first PLS loadings (Annex 1, Figures 32–36), accounting for more than 95% of the data variance. Overall, the most considerable variation occurred in the spectral region between 1300 and 1500 nm and between 1800 and 1950 nm. Other authors have found similar response in these spectral regions for prediction of sugars in other products (BÁZÁR *et al.*, 2016; POMARES-VICIANA *et al.*, 2018). This effect was similar for all sample types and all parameters evaluated. Therefore, this fact proved the relationships identified by the VIP scores and the key molecular bonds related to the parameters of interest described earlier.

The gap between two spectrometers, starting at 1011 nm until 1070 nm, is not related to significative known vibrational frequency overtones associated with some bands related to sugars or fibres (WORKMAN JR.; WEYER, 2012). Therefore, the absence of information in this range would not have significantly affected the development of the models.

Processed samples allowed a more significant interaction of electromagnetic radiation corresponding to vis-NIR bands with matter constituents. This physical effect resulted in more prominent signals from specific vibrational frequency bands related to the chemical constitution of sugarcane quality parameters. Overall, defibrated samples (DF) showed performance prediction results that were close to raw juice samples (RJ). Also, the DF sample allowed us to predict Fibre content as well as other parameters, which is not possible with RJ samples. The prediction of sugarcane quality parameters from less processed samples is a desirable characteristic for mechanization of on-the-go measurements of crops, thus promoting spatial information of crops based on quality. DF samples may partially satisfy this requirement; however, this is a destructive sampling technique.

Improving the predicting performance of sugarcane quality parameters from billets for on-the-go systems may be possible (NAWI *et al.*, 2013). Some effects, such as waxy and skin organic compounds, need to be considered and minimized (PHUPHAPHUD; SAENGPRACHATANARUG; POSOM; WONGPICHET; *et al.*, 2019). Other

data processing techniques such as nonlinear models (SEXTON *et al.*, 2018) or advanced filtering methods such as orthogonalization (ROGER; BOULET, 2018), could be investigated to improve the performance aiming to develop reliable models for measuring sugarcane quality using billets of cane.

3.4. Conclusions

This study demonstrates that vis-NIR spectroscopy could be used as a quick method to assess the abundance of chemical compounds of sugarcane related to its quality. There was no significant difference (*p*-value > 0.05) in the accuracy (RMSEP) of prediction of whole cane samples when compared to processed samples, such as de-fibrated cane (DF) and extracted raw juice (RJ), for all evaluated quality parameters. Also, outer-surface measurements of sugarcane billets presented a better accuracy (RMSEP, *p*-value > 0.05) and performance (R^2p and RPIQ) than measurements on the cross-section.

Despite the similar accuracy (p-value > 0.05), DF and RJ sampling presented better performance than outer-surface measurements of sugarcane billets. Moreover, the performance of the models from DF and RJ samples were similar, but DF samples involve less preparation, as they do not require juice extraction of the sample.

The results showed that DF sampling could be used to predict the main sugarcane quality parameters, such as soluble solids content (Brix), saccharose (Pol), Fibre, Pol of cane, and total recoverable sugars (TRS), all of which are used for pricing and trading be-tween mills and sugarcane producers. The DF models presented RMSEP varying between 0.72% and 0.87% for Brix, Pol, Fibre, and Pol of cane, and 6.71 kg Mg−1 for TRS.

The results in this study contribute to advancing the development of on-board quality monitoring in sugarcane. This information shows the spatial variability of crop quality and helps guide site-specific management of sugarcane fields.

References

ASTM E1655-17. **Standard Practices for Infrared Multivariate Quantitative Analysis**. [*S. l.*: *s. n.*], 2017. https://doi.org/10.1520/E1655-05R12.2.

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. **Applied Spectroscopy**, vol. 43, no. 5, p. 772–777, 1989. https://doi.org/10.1366/0003702894202201.

BAZAR, G.; KOVACS, Z.; TSENKOVA, R. Evaluating spectral signals to identify spectral error. **PLoS ONE**, vol. 11, no. 1, 5 Jan. 2016. https://doi.org/10.1371/journal.pone.0146249.

BÁZÁR, G.; ROMVÁRI, R.; SZABÓ, A.; SOMOGYI, T.; ÉLES, V.; TSENKOVA, R. NIR detection of honey adulteration reveals differences in water spectral pattern. **Food Chemistry**, vol. 194, p. 873–880, 2016. https://doi.org/10.1016/j.foodchem.2015.08.092.

BRAMLEY, R. G. V. Lessons from nearly 20 years of Precision Agriculture research, development, and adoption as a guide to its appropriate application. **Crop and Pasture Science**, vol. 60, no. 3, p. 197–217, 2009. https://doi.org/10.1071/CP08304.

BRAMLEY, R. G. V; PANITZ, J. H.; JENSEN, T. A.; BAILLIE, C. P. Within block spatial variation in CCS - Another potentially important consideration in the application of precision agriculture to sugarcane production. **International Sugar Journal**, vol. 115, no. 1374, p. 1–8, 2013. DOI http://www.scopus.com/inward/record.url?eid=2-s2.0-84916888448&partnerID=40&md5=265e29eda23d2e6a3a9affa0ccceffcc. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84916888448&partnerID=40&md5=265e29eda23d2e6a3a9affa0ccceffcc.

CARDOZO, N. P.; SENTELHAS, P. C. Climatic effects on sugarcane ripening under the influence of cultivars and crop age. **Scientia Agricola**, vol. 70, no. 6, p. 449–456, Nov. 2013. DOI 10.1590/S0103-90162013000600011. Available at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-90162013000600011&lng=en&nrm=iso&tlng=en. Accessed on: 20 Feb. 2021.

CONSECANA. National Concill of Sugarcane Producers of São Paulo State. vol. 6, p. 81, 2015. .

CORRÊDO, L. de P.; CANATA, T. F.; MALDANER, L. F.; DE LIMA, J. de J. A.; MOLIN, J. P. Sugarcane Harvester for In-field Data Collection: State of the Art, Its Applicability and Future Perspectives. **Sugar Tech**, vol. 22, no. 4, p. 1–14, 10 Aug. 2020. DOI 10.1007/s12355-020-00874-3. Available at: http://link.springer.com/10.1007/s12355-020-00874-3. Accessed on: 10 Aug. 2020.

CORTÉS, V.; BLASCO, J.; ALEIXOS, N.; CUBERO, S.; TALENS, P. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. **Trends in Food Science and Technology**, vol. 85, p. 138–148, 1 Mar. 2019. https://doi.org/10.1016/j.tifs.2019.01.015.

COSTA, M. V. A. da; FONTES, C. H.; CARVALHO, G.; JÚNIOR, E. C. de M. UltraBrix: A Device for Measuring the Soluble Solids Content in Sugarcane. **Sustainability**, vol. 13, no. 3, p. 1227, 25 Jan. 2021. DOI 10.3390/su13031227. Available at: https://www.mdpi.com/2071-1050/13/3/1227. Accessed on: 27 Jan. 2021.

FRANCESCHINI, M. H. D.; DEMATTÊ, J. A. M.; KOOISTRA, L.; BARTHOLOMEUS, H.; RIZZO, R.; FONGARO, C. T.; MOLIN, J. P. Effects of external factors on soil reflectance measured on-the-go and assessment of potential spectral correction through orthogonalisation and standardisation procedures. **Soil and Tillage Research**, vol. 177, p. 19–36, 1 Apr. 2018. https://doi.org/10.1016/j.still.2017.10.004.

GERRETZEN, J.; SZYMAŃSKA, E.; JANSEN, J. J.; BART, J.; VAN MANEN, H. J.; VAN DEN HEUVEL, E. R.; BUYDENS, L. M. C. Simple and Effective Way for Data Preprocessing Selection Based on Design of Experiments. **Analytical Chemistry**, vol. 87, no. 24, p. 12096–12103, 15 Dec. 2015. DOI 10.1021/acs.analchem.5b02832. Available at: https://pubs.acs.org/sharingguidelines. Accessed on: 13 Feb. 2020.

GOLIC, M.; WALSH, K.; LAWSON, P. Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature. **Applied Spectroscopy**, vol. 57, no. 2, p. 139–145, 2003. DOI 10.1366/000370203321535033. Available at: https://journals.sagepub.com/doi/pdf/10.1366/000370203321535033. Accessed on: 9 Jan. 2019.

INARKAR, M. B.; LELE, S. S. Extraction and Characterization of Sugarcane Peel Wax. **ISRN Agronomy**, vol. 2012, p. 1–6, 2012. https://doi.org/10.5402/2012/340158.

KENNARD, R. W.; STONE, L. A. Computer Aided Design of Experiments. **Technometrics**, vol. 11, no. 1, p. 137–148, 1969. https://doi.org/10.1080/00401706.1969.10490666.

MANCINI, M. C.; LEITE, D. C.; PERECIN, D.; BIDÓIA, M. A. P.; XAVIER, M. A.; LANDELL, M. G. A.; PINTO, L. R. Characterization of the Genetic Variability of a Sugarcane Commercial Cross Through Yield Components and Quality Parameters. **Sugar Tech**, vol. 14, no. 2, p. 119–125, 2012. https://doi.org/10.1007/s12355-012-0141-5.

MARAPHUM, K.; CHUAN-UDOM, S.; SAENGPRACHATANARUG, K.; WONGPICHET, S.; POSOM, J.; PHUPHAPHUD, A.; TAIRA, E. 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**, vol. 26, no. 5, p. 287–296, 2018. DOI 10.1177/0967033518795810. Available at: https://journals.sagepub.com/doi/pdf/10.1177/0967033518795810. Accessed on: 11 Mar. 2019.

NAWI, N. M.; CHEN, G.; JENSEN, T. In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: a review. **Precision Agriculture**, vol. 15, no. 6, p. 684–703, 2014. https://doi.org/10.1007/s11119-014-9362-9.

NAWI, N. M.; CHEN, G.; JENSEN, T. Visible and shortwave near infrared spectroscopy for predicting sugar content of sugarcane based on a cross-sectional scanning method. **Journal of Near Infrared Spectroscopy**, vol. 21, no. 4, p. 289–297, 2013. https://doi.org/10.1255/jnirs.1060.

NAWI, N. M.; CHEN, G.; JENSEN, T.; MEHDIZADEH, S. A. Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared. **Biosystems Engineering**, vol. 115, p. 154–161, 2013. DOI 10.1016/j.biosystemseng.2013.03.005. Available at: http://dx.doi.org/10.1016/j.biosystemseng.2013.03.005.

OLIVERI, P.; MALEGORI, C.; SIMONETTI, R.; CASALE, M. The impact of signal pre-processing on the final interpretation of analytical outcomes – A tutorial. **Analytica Chimica Acta**, vol. 1058, p. 9–17, 13 Jun. 2019. https://doi.org/10.1016/j.aca.2018.10.055.

OLIVIERI, A. C. Practical guidelines for reporting results in single- and multi-component analytical calibration: A tutorial. **Analytica Chimica Acta**, vol. 868, p. 10–22, 8 Apr. 2015. https://doi.org/10.1016/j.aca.2015.01.017.

OSBORNE, B. G. Near-Infrared Spectroscopy in Food Analysis. **Encyclopedia of Analytical Chemistry**, Chichester, UK, , p. 14, 30 Oct. 2006. DOI 10.1002/9780470027318.a1018. Available at: http://doi.wiley.com/10.1002/9780470027318.a1018. Accessed on: 11 Dec. 2018.

PAGANI, V.; STELLA, T.; GUARNERI, T.; FINOTTO, G.; VAN DEN BERG, M.; MARIN, F. R.; ACUTIS, M.; CONFALONIERI, R. Forecasting sugarcane yields using agro-climatic indicators and Canegro model: A case study in the main production region in Brazil. **Agricultural Systems**, vol. 154, p. 45–52, 1 Jun. 2017. https://doi.org/10.1016/j.agsy.2017.03.002.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, vol. 1026, p. 8–36, 2018. https://doi.org/10.1016/j.aca.2018.04.004.

PHETPAN, K.; UDOMPETAIKUL, V.; SIRISOMBOON, P. An online visible and near-infrared spectroscopic technique for the real-time evaluation of the soluble solids content of sugarcane billets on an elevator conveyor. **Computers and Electronics in Agriculture**, vol. 154, p. 460–466, Nov. 2018. DOI 10.1016/j.compag.2018.09.033. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0168169918302096. Accessed on: 7 Jan. 2019.

PHUPHAPHUD, A.; SAENGPRACHATANARUG, K.; POSOM, J.; MARAPHUM, K.; TAIRA, E. Non-destructive and rapid measurement of sugar content in growing cane stalks for breeding programmes using visible-near infrared spectroscopy. **Biosystems Engineering**, vol. 197, p. 76–90, 1 Sep. 2020. https://doi.org/10.1016/j.biosystemseng.2020.06.012.

PHUPHAPHUD, A.; SAENGPRACHATANARUG, K.; POSOM, J.; MARAPHUM, K.; TAIRA, E. Prediction of the fibre content of sugarcane stalk by direct scanning using visible-shortwave near infrared spectroscopy. **Vibrational Spectroscopy**, vol. 101, p. 71–80, Mar. 2019. DOI 10.1016/j.vibspec.2019.02.005. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0924203118302613. Accessed on: 3 Jun. 2019.

PHUPHAPHUD, A.; SAENGPRACHATANARUG, K.; POSOM, J.; WONGPICHET, S.; MARAPHUM, K.; TAIRA, E. Effects of waxy types of a sugarcane stalk surface on the spectral characteristics of visible-shortwave near infrared measurement. **Engineering Journal**, vol. 23, no. 1, p. 13–24, 2019. https://doi.org/10.4186/ej.2019.23.1.13.

POMARES-VICIANA, T.; MARTÍNEZ-VALDIVIESO, D.; FONT, R.; GÓMEZ, P.; DEL RÍO-CELESTINO, M. Characterisation and prediction of carbohydrate content in zucchini fruit using near infrared spectroscopy. **Journal of the Science of Food and Agriculture**, vol. 98, no. 5, p. 1703–1711, 30 Mar. 2018. https://doi.org/10.1002/jsfa.8642.

RAMÍREZ-MORALES, I.; RIVERO, D.; FERNÁNDEZ-BLANCO, E.; PAZOS, A. Optimization of NIR calibration models for multiple processes in the sugar industry. **Chemometrics and Intelligent Laboratory Systems**, vol. 159, p. 45–57, Dec. 2016. DOI 10.1016/j.chemolab.2016.10.003. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0169743916303604. Accessed on: 2 Sep. 2019.

RODRIGUES, F. A.; MAGALHÃES, P. S. G.; FRANCO, H. C. J. Soil attributes and leaf nitrogen estimating sugar cane quality parameters: Brix, pol and fibre. **Precision Agriculture**, vol. 14, no. 3, p. 270–289, 2013. https://doi.org/10.1007/s11119-012-9294-1.

ROGER, J. M.; BOULET, J. C. A review of orthogonal projections for calibration. **Journal of Chemometrics**, vol. 32, p. 1–16, 1 Sep. 2018. DOI 10.1002/cem.3045. Available at: http://doi.wiley.com/10.1002/cem.3045. Accessed on: 8 Apr. 2020.

SANTANA, F. B. de; GIUSEPPE, L. O. de; SOUZA, A. M. de; POPPI, R. J. Removing the moisture effect in soil organic matter determination using NIR spectroscopy and PLSR with external parameter orthogonalization. **Microchemical Journal**, vol. 145, p. 1094–1101, 1 Mar. 2019. https://doi.org/10.1016/j.microc.2018.12.027.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, vol. 36, no. 8, p. 1627–1639, 1964. DOI 10.1021/ac60214a047. Available at: https://pubs.acs.org/sharingguidelines. Accessed on: 28 Mar. 2020.

SEXTON, J.; EVERINGHAM, Y.; DONALD, D.; STAUNTON, S.; WHITE, R. A comparison of non-linear regression methods for improved on-line near infrared spectroscopic analysis of a sugarcane quality measure. **Journal of Near Infrared Spectroscopy**, 2018. DOI 10.1177/0967033518802448. Available at: https://journals.sagepub.com/doi/pdf/10.1177/0967033518802448. Accessed on: 11 Mar. 2019.

SEXTON, J.; EVERINGHAM, Y.; DONALD, D.; STAUNTON, S.; WHITE, R. Investigating the identification of atypical sugarcane using NIR analysis of online mill data. **Computers and Electronics in Agriculture**, vol. 168, 1 Jan. 2020. DOI 10.1016/j.compag.2019.105111. Available at: https://doi.org/10.1016/j.compag.2019.105111. Accessed on: 3 Jan. 2021.

SOLOMON, S. Post-harvest deterioration of sugarcane. **Sugar Tech**, vol. 11, no. 2, p. 109–123, 2009. https://doi.org/10.1007/s12355-009-0018-4.

SOROL, N.; ARANCIBIA, E.; BORTOLATO, S. A.; OLIVIERI, A. C. Visible/near infrared-partial least-squares analysis of Brix in sugar cane juice. A test field for variable selection methods. **Chemometrics and Intelligent Laboratory Systems**, vol. 102, no. 2, p. 100–109, 2010. DOI 10.1016/j.chemolab.2010.04.009. Available at: https://ac.els-cdn.com/S0169743910000584/1-s2.0-S0169743910000584-main.pdf?_tid=9ea5d650-a1e2-40fd-8b20-2089be31da78&acdnat=1549390123_c9d156b7ef74f98af5764f4ba293261e. Accessed on: 2 Jul. 2018.

STEIDLE NETO, A. J.; LOPES, D. C.; TOLEDO, J. V; ZOLNIER, S.; SILVA, T. G. F. Classification of sugarcane varieties using visible/near infrared spectral reflectance of stalks and multivariate methods. **Journal of Agricultural Science**, vol. 156, no. 4, p. 537–546, 2018. DOI 10.1017/S0021859618000539. Available at: https://doi.org/10.1017/S0021859618000539. Accessed on: 7 Jan. 2019.

STEIDLE NETO, Antonio José; TOLEDO, J. V.; ZOLNIER, S.; LOPES, D. de C.; PIRES, C. V.; SILVA, T. G. F. d. Prediction of mineral contents in sugarcane cultivated under saline conditions based on stalk scanning by Vis/NIR spectral reflectance. **Biosystems Engineering**, vol. 156, p. 17–26, Apr. 2017. DOI 10.1016/j.biosystemseng.2017.01.003. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1537511016302513. Accessed on: 7 Jan. 2019.

TAIRA, E.; UENO, M.; FURUKAWA, N.; TASAKI, A.; KOMAKI, Y.; NAGAI, J. I.; SAENGPRACHATANARUG, K. Networking system employing near infrared spectroscopy for sugarcane payment in Japan. **Journal of Near Infrared Spectroscopy**, vol. 21, no. 6, p. 477–483, 2013. DOI 10.1255/jnirs.1081. Available at: www.anton-paar.com/. Accessed on: 26 Nov. 2020.

TAIRA, E.; UENO, M.; KAWAMITSU, Y. Automated quality evaluation system for net and gross sugarcane samples using near infrared spectroscopy. **Journal of Near Infrared Spectroscopy**, vol. 18, no. 3, p. 209–215, 2010. DOI 10.1255/jnirs.884. Available at: https://journals.sagepub.com/doi/pdf/10.1255/jnirs.884. Accessed on: 31 Aug. 2019.

TAIRA, E.; UENO, M.; SAENGPRACHATANARUG, K.; KAWAMITSU, Y. Direct sugar content analysis for whole stalk sugarcane using a portable near infrared instrument. **Journal of Near Infrared Spectroscopy**, vol. 21, no. 4, p. 281–287, 2013. DOI 10.1255/jnirs.1064. Available at: https://journals.sagepub.com/doi/pdf/10.1255/jnirs.1064. Accessed on: 24 May 2019.

UDOMPETAIKUL, V.; PHETPAN, K.; SIRISOMBOON, P. Development of the partial least-squares model to determine the soluble solids content of sugarcane billets on an elevator conveyor. **Measurement**, vol. 167, p. 107898, 1 Jan. 2021. DOI 10.1016/j.measurement.2020.107898. Available at: https://linkinghub.elsevier.com/retrieve/pii/S026322412030436X. Accessed on: 17 Feb. 2021.

UYS, L.; BOTHA, F. C.; HOFMEYR, J. H. S.; ROHWER, J. M. Kinetic model of sucrose accumulation in maturing sugarcane culm tissue. **Phytochemistry**, vol. 68, no. 16–18, p. 2375–2392, 1 Aug. 2007. https://doi.org/10.1016/j.phytochem.2007.04.023.

VALDERRAMA, Patricia; BRAGA, J. W. B.; POPPI, R. J. Validation of multivariate calibration models in the determination of sugar cane quality parameters by near infrared spectroscopy. **Journal of the Brazilian Chemical Society**, vol. 18, no. 2, p. 259–266, 2007a. DOI 10.1590/S0103-50532007000200003. Available at: http://www.scielo.br/pdf/jbchs/v18n2/a03v18n2.pdf. Accessed on: 3 Jul. 2018.

VALDERRAMA, Patrícia; BRAGA, J. W. B.; POPPI, R. J. Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy. **Journal of Agricultural and Food Chemistry**, vol. 55, no. 21, p. 8331–8338, 2007b. DOI 10.1021/jf071538s. Available at: https://pubs.acs.org/doi/pdf/10.1021/jf071538s. Accessed on: 3 Jul. 2018.

VAN DER VOET, H. Comparing the predictive accuracy of models using a simple randomization test. **Chemometrics and Intelligent Laboratory Systems**, vol. 25, no. 2, p. 313–323, 1 Nov. 1994. https://doi.org/10.1016/0169-7439(94)85050-X.

WALSH, K. B.; BLASCO, J.; ZUDE-SASSE, M.; SUN, X. Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. **Postharvest Biology and Technology**, vol. 168, p. 111246, 1 Oct. 2020. https://doi.org/10.1016/j.postharvbio.2020.111246.

WANG, Jianping; NAYAK, S.; KOCH, K.; MING, R. Carbon partitioning in sugarcane (Saccharum species). **Frontiers in Plant Science**, vol. 4, p. 201, 18 Jun. 2013. DOI 10.3389/fpls.2013.00201. Available at: http://journal.frontiersin.org/article/10.3389/fpls.2013.00201/abstract. Accessed on: 13 Apr. 2020.

WANG, Jungang; ZHAO, T.; YANG, B.; ZHANG, S. Sucrose Metabolism and Regulation in Sugarcane. **Journal of Plant Physiology & Pathology**, vol. 05, no. 04, 7 Feb. 2017. https://doi.org/10.4172/2329-955x.1000167.

WORKMAN JR., J.; WEYER, L. **Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy**. [*S. l.*: *s. n.*], 2012.

4. ON BOARD SUGARCANE QUALITY MEASUREMENT ON THE HARVESTER

ABSTRACT

On-board quality measurements have been made on harvesters or combine harvesters in some studies with substantial evolution towards the development of commercial sensors, mainly based on near-infrared spectroscopy (NIR). However, initiatives applied to sugarcane quality need to advance to field level stages. This study aimed to propose a method to measure sugarcane quality with an on-board NIR spectroscopy sensor. A platform was built to support the system composed of the NIR sensor and external lighting on the elevator of a sugarcane harvester. Real-time data collections were performed with this layout in three areas of a commercial field. Georeferenced samples were collected for calibration, validation, and adjustment of the multivariate models by partial least squares (PLS) regression. After harvest, soil samples were collected for evaluation of cause-and-effect relationships. The samples were analyzed in the laboratory by conventional methods. In addition, subsamples of defibrated cane were collected for development of calibration transfer models by piecewise direct standardization (PDS). The method allowed the adjustment of the spectra collected in real time to predict the quality properties soluble solids content (Brix), apparent sucrose in juice (Pol), Fibre, cane Pol, and total recoverable sugar (TRS). The results of relative mean square error of prediction (RRMSEP) were from 1.80 to 2.14%, and ratio of interquartile performance (RPIQ) were from 1.79 to 2.46. The PLS-PDS models were applied to data collected in real time, allowing estimation of quality properties and identification of the existence of spatial variability in quality. Moreover, the evaluation of cause-and-effect relationships, based on results estimated by the method, allowed the assessment of relationships between sugarcane quality properties and soil physical attributes, corroborating other studies. The conclusions showed that it is possible to perform the measurement of sugarcane quality in real time by means of multivariate methods for treating spectral data. The findings of the present study represent an advance towards the development of an on-board sensor for this purpose, providing a new layer for the evaluation of the spatial variability of sugarcane fields and site-specific crop management.

Keywords: On-board NIR sensor; technological quality spatial variability

4.1. Introduction

Proximal sensing is the main tool for high density data acquisition to analyze the inherent in-field spatial variation and to allow the implementation of precision agriculture (PA) management concepts. In this context, yield data is the most robust information for the initiation of management strategies, as it translates the effects that occurred throughout the crop cycle, such as inputs applied, climate, and soil variability (MOLIN *et al.*, 2020). However, many crops have their added value tied to quality parameters (CORTÉS *et al.*, 2019).

Sugarcane is one of the tropical crops with the greatest implementation and application of technological resources in its cultivation (SILVA; DE MORAES; MOLIN, 2011). The monitoring of sugarcane yields is a feasible reality; on the other hand, its application is incipient (MALDANER *et al.*, 2021). The low adoption of the technology may be related to applicability due to the level of information that sugar mills and sugarcane growers wish to obtain. In this way, the combination of yield and quality information could provide in-field information of the product whose value is directly related to the yield of sugars (CORRÊDO *et al.*, 2020). Also, the current technologies have not capacity to monitor product quality (BRAMLEY, 2009).

Sugarcane quality is monitored by means of sampling in trucks transporting the cargo harvested in the field for sugar mill processing. This monitoring is done for purposes of payment to suppliers and industrial control.

The quality properties are determined in the laboratory by conventional methods of analysis, such as refractometry of the raw juice, to determine the soluble solids content (Brix), juice clarification by chemical reagents and measurement of apparent sucrose (Pol) by polarimetry, and determination of insoluble solids (Fibre) by gravimetric methods, after extraction of the juice from defibrated cane (CONSECANA, 2015). Based on these parameters others are estimated, as total recoverable sugars (TRS). For in-field conditions the analysis must be different from the conventional laboratory routine due high-cost, time-consuming and environmental unfriendly.

Currently the designing of new instruments for agricultural products quality measurement by non-destructive way has been the aim of several researchers mainly for monitoring at industry (ABASI *et al.*, 2018). Technologies with diversified physical principles have been employed in research evaluating the potential application for this purpose, such as acoustic, electronic noses, machine vision, electrical, and spectroscopic (WANG *et al.*, 2015). Among these, visible and near-infrared spectroscopy (VNIR) and near-infrared (NIR) only has received additional attention mainly for in-line inspection agricultural products (CORTÉS *et al.*, 2019).

The interaction between electromagnetic energy, as absorbance, reflectance and transmittance, and the matter produce vibration of molecular bonds containing relatively heavy atom (C, N, O, and S) attached to a hydrogen atom, associated with water and storage reserves of agricultural products (PASQUINI, 2018; WALSH *et al.*, 2020). Thus, robust analytical equipment previously used on benches, with static and standardized samples (off-line analysis), as for monitoring Brix and Pol on the sugar mill laboratory (SEXTON *et al.*, 2020), has been miniaturized and implemented for monitoring moving samples (in-line), with diversified applications, such as for fertilizer industry (RUANGRATANAKORN *et al.*, 2020) and fruit quality monitoring (POREP; KAMMERER; CARLE, 2015; SALGUERO-CHAPARRO; PEÑA-RODRÍGUEZ, 2014; XU; MO; *et al.*, 2019). Also, has been implemented as proximal sensors from which the sensor moving, as on-board applications in agricultural implements for characterization of soil attributes with substantial advances and satisfactory results (MUNNAF; MOUAZEN, 2021; NAWAR; MUNNAF; MOUAZEN, 2020; TÜMSAVAŞ *et al.*, 2019).

For in-line measurements in-field, i.e., by on-board sensing, in addition to accuracy limitations that may occur due to the sensor being in motion, there are the specificities conditions of agricultural fields to which the sensor will be subjected, such as dirt, moisture, vibrations, irregularity in the sensor sample distance, and lighting variation. In this way, there are a substantial number of methods to overcome environmental effects and potentialize prediction accuracy as advance filtering methods by orthogonalization (FRANCESCHINI *et al.*, 2018), machine learning (KUANG; TEKIN; MOUAZEN, 2015; MORELLOS *et al.*, 2016) and calibration model transfer (JI; VISCARRA ROSSEL; SHI, 2015a, b)

Model transfer is a robust multivariate method based on the limitation of multivariate calibration applied to measurements on new environmental conditions or another separate instrument from those in which they were initially performed (FEUDALE *et al.*, 2002). Thus, from a master model, the goal is to obtain similar accuracy in the new measurement conditions (PASQUINI, 2018). It has been used for transfer between equipment with different wavelength ranges (PISSARD *et al.*, 2021), equipment with different analytical-physical principles (PANCHUK *et al.*, 2017), and between measurements under different environmental conditions (FRANCESCHINI *et al.*, 2018; JI; VISCARRA ROSSEL; SHI, 2015a; NAWAR; MUNNAF; MOUAZEN, 2020).

Ji, Viscarra Rossel and Shi (2015b) investigated the calibration transfer method piecewise direct standardization (PDS) to overcome water and other environmental effects from on-the-go VNIR proximal sensing of soil organic carbon (OC). Based on a spectroscopy database of dried soil spectra recorded in bench, they used PDS to transfer the models to predict OC by in-line measurement in the field by an on-board sensor pulled by

agricultural implement. There are a substantially improvement in the accuracy results compared with original field spectra from a fewer transfer sample dataset. Thus, the method allows to overcome the negative effects of on-board measurement.

Also, about the use of NIR spectroscopy for on-board proximal sensing in the field, other studies have been conducted for monitoring grain quality in combines by means of on-board NIR proximal sensing (RISIUS *et al.*, 2015; YANG *et al.*, 2017). In this way, Long and McCallum (2020) adapted and evaluated a moderately priced reflectance spectrometer shipped in the combine for the prediction of grain protein concentration (GPC). Furthermore, they compared the low-cost equipment with commercial analytical equipment. The calibration models, performed by commercial equipment on static conditions, presented a standard error of cross-validation of 0.43 g kg^{-1} GPC. Also, the in-house equipment presented a correlation of 0.88 between protein maps performed for both equipment.

Studies with applications of VNIR and NIR spectroscopy techniques for the prediction of sugarcane quality attributes, aiming implementation as an on-board harvester proximal sensing technique, have been evolving. The first promising results were reported by Nawi *et al.* (2013), and Nawi, Chen and Jensen (2013) evaluating on the bench different ways of measuring billet samples by taking readings on the outer surface and cross-section. Also, in a review of different feasible technologies for this purpose, the authors propose sampler models for measurements at the elevator of the harvester (NAWI; CHEN; JENSEN, 2014).

Subsequently, Phetpan, Udompetaikul and Sirisomboon (2018) built a replica of sugarcane elevator, for study under controlled conditions. Using a measurement chamber on the elevator housing an optical fibre spectrometer and four lamps as external radiation sources, the authors evaluated different pre-processing techniques and spectral analysis to differentiate the target (sugarcane billets) from the elevator belt components. The results were promising, with an accuracy of 0.30% for Brix measurement and a coefficient of determination of 0.79. Moreover, as a continuation of this study, Udompetaikul, Phetpan and Sirisomboon (2021) aimed to improve the prediction of the system by evaluating the need of the visible region of the spectrum in the prediction and to determine if the development of Brix prediction models from the measurement of billets in the elevator, neglecting the influence of the different levels of cane in the elevator is feasible. The authors concluded that different cane levels affected the prediction of the models. However, the development of a combined model to cover variations in this difference was used to predict two outer sets, combined with full and half delivery and unknown with samples of different growing seasons, obtaining RMSEP of 0.42% for both conditions.

Using an extensive database collected at a sugar mill's quality lab, Corrêdo *et al.* (2021b) developed prediction models for Brix and Pol from samples of shredded cane. Based on these models, the authors evaluated the presence of spatial variability of these attributes from samples collected in a commercial field. The study shows the presence of spatial variability of these attributes, and correlation higher than 0.80 between maps interpolated from the predicted results and maps generated from the results obtained with conventional methods.

As recent studies have shown, there is an opportunity to use NIR spectroscopy to predict sugarcane quality in the field. Sugarcane crop quality data acquisition is an opportunity for more effective site-specific management (MOLIN *et al.*, 2020). However, the performance of the technique under the challenging conditions to which the sensor will be subjected for field measurement has not yet been evaluated (CORRÊDO *et al.*, 2020). One possible way to overcome these effects may be by implementing advanced data processing techniques with calibration transfer. Therefore, the goal of this study was to develop and validate a nondestructive method for sugarcane quality assessment during harvest with on-board NIR spectroscopy proximal sensing.

4.2. Material and Methods

4.2.1. Instrumentation

The instrument used in this study was the microNIR 1700 spectrometer (Viavi Solutions, JDSU Corporation, Milpitas, CA). It is a compact and portable device entirely powered with 5 Volts by an USB port of a portable computer. Also, the device is tolerant to vibrations and contains two tungsten light bulbs used as the radiation source allowing measurements of samples placed close to its window (PASQUINI, 2018). It operates in the spectral range from 908 to 1676 nm, with a nominal spectral resolution of 6.19 nm. The wavelength selection consists of the use of a linearly variable interference filter (LVF) deposited over a chip containing a 128-pixel InGaAs photodiode array detector.

Some adaptations were performed allowing the data acquisition as an on-board sensor in a sugarcane harvester. The microNIR was connected to a HUB module responsible for converting the signal transmitted via USB cable to Ethernet (Figure 16a), which allows the signal to be sent over longer distances in a more stable way (POREP; KAMMERER; CARLE, 2015). The assembly was housed in a cast iron enclosure in which the sensor was housed over a hole with a sapphire window (Figure 16b). In addition, the system was connected to a voltage regulator module responsible for reducing the voltage of the harvester's battery from 12 to 5 Volts.

The location in the sugarcane harvester defined for embed the sensor was the elevator, immediately before the secondary extractor, in which the billets are partially cleaned of coarser impurities, as recommended by Nawi, Chen, and Jensen (2014). A platform was built to support the sensor on the elevator belt with 200 mm between the sapphire window and the conveyor belt bottom and 40 mm over the conveyor slats (Figure 16c). The connection of the platform with the elevator was performed with four rubber vibration dampers. The electromagnetic radiation source consisted of four halogen lamps (AR111 type) with voltage of 12 V and power of 50 W (PHETPAN; UDOMPETAIKUL; SIRISOMBOON, 2018). The lamps were mounted at 420 mm and at 45º away over the conveyor belt bottom to irradiate the sample under the sensor window. The entire on-board system, including sensor and electromagnetic radiation sources, were enclosed in a galvanized steel box (800 x 800 x 600 mm), with the inside painted matte black, and the outside painted white, to minimize interference from external lighting. Also, the sensor signal was received in device specific software in a computer in the harvester cab and concatenated with the positional signal from a GNSS receiver SMART6-L™ (Novatel Inc., Calgary, Alberta, Canada) with TerraStar-C correction that allows the accuracy of ±0.09 m (Figure 16d). The GNSS was located above the harvester cab in the direction of the base cutter. The entire system was powered by one of two batteries of the harvester.

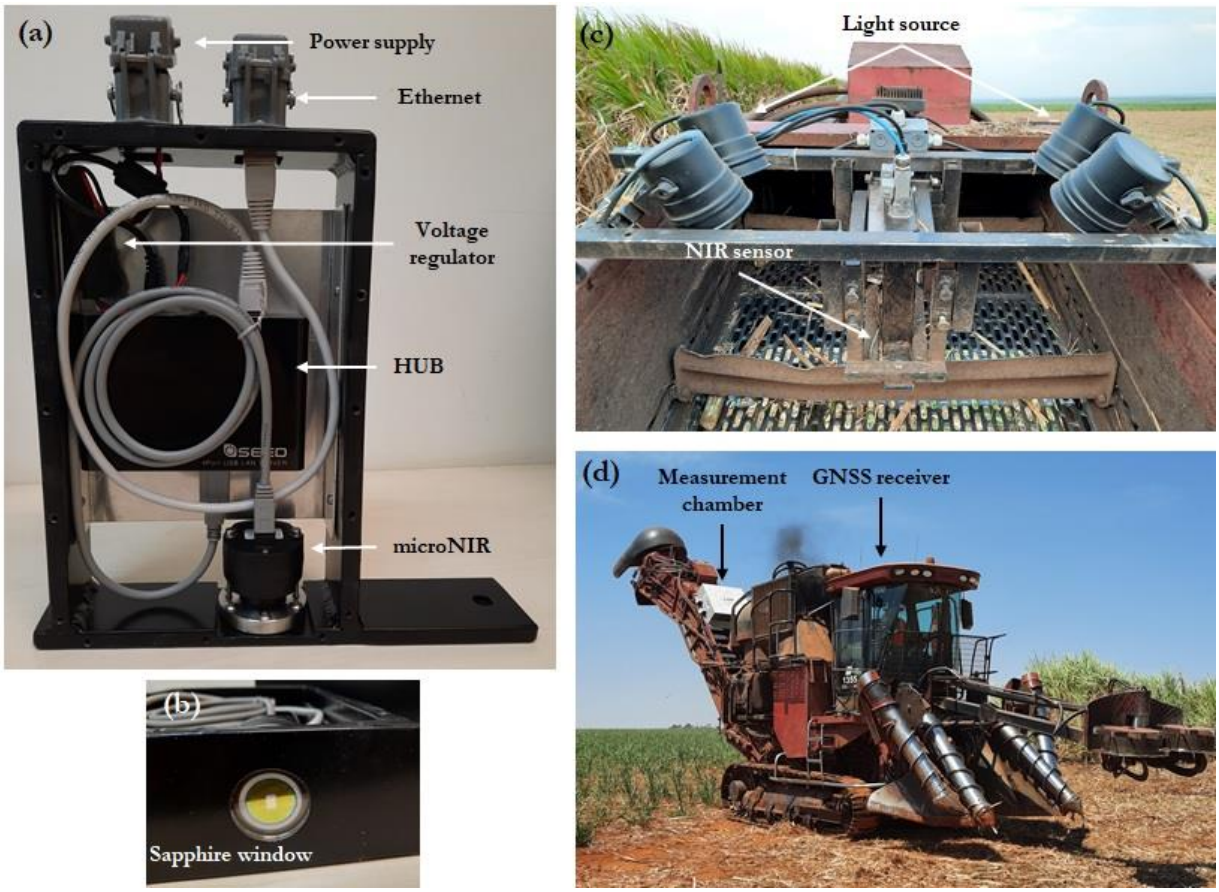

Figure 16. MicroNIR spectrometer and device for embed sensor (a); device bottom view (b); arrangement of electromagnetic radiation sources and sensor on the measurement platform on board of the harvester elevator (c); detail of the measurement chamber position and GNSS receiver in the sugarcane harvester (d).

4.2.2. Data acquisition: on-the-go measurements and reference samples

The experiment was performed in a commercial sugarcane growing field on three consecutive days of 2020/21 harvest season. The field was planted with the CTC 4 cultivar planted in the second year of the ratoon cane, with 1.50 m between rows, in a low slope area (Figure 17).

Prior the on-board measurements, it was performed the calibration of the NIR sensor under environmental field conditions. An increase in the operating distance may cause a drop in spectrum signal-to-noise ratio due to a decrease in light intensity reaching the sample. For that, the spectrometer integration time needs to properly optimize to account for the loss in photometric intensity. A squared white barium sulfate plate ($BaSO_4$), with 0.40 m, was used as the reference standard to calibrate the sensor, assuming it had a uniform reflectance equal to one (100%) (MAYRINK *et al.*, 2019), while a dark reference was obtained from the chamber measurement environment. The lights were turned on for white calibration and off for the dark reference (chamber environment). The white plate was positioned on the elevator belt, and the calibration was optimized to 50 scans and integration time of 35 ms (0.7 ms per scan), resulting in a measurement time of 1.75 s to obtain at around 50000 to 60000 raw counts of photometric intensity, as recommended by the manufacturer. The calibration of the device was performed on each collection day in the field, prior to the data acquisition and checked at noon.

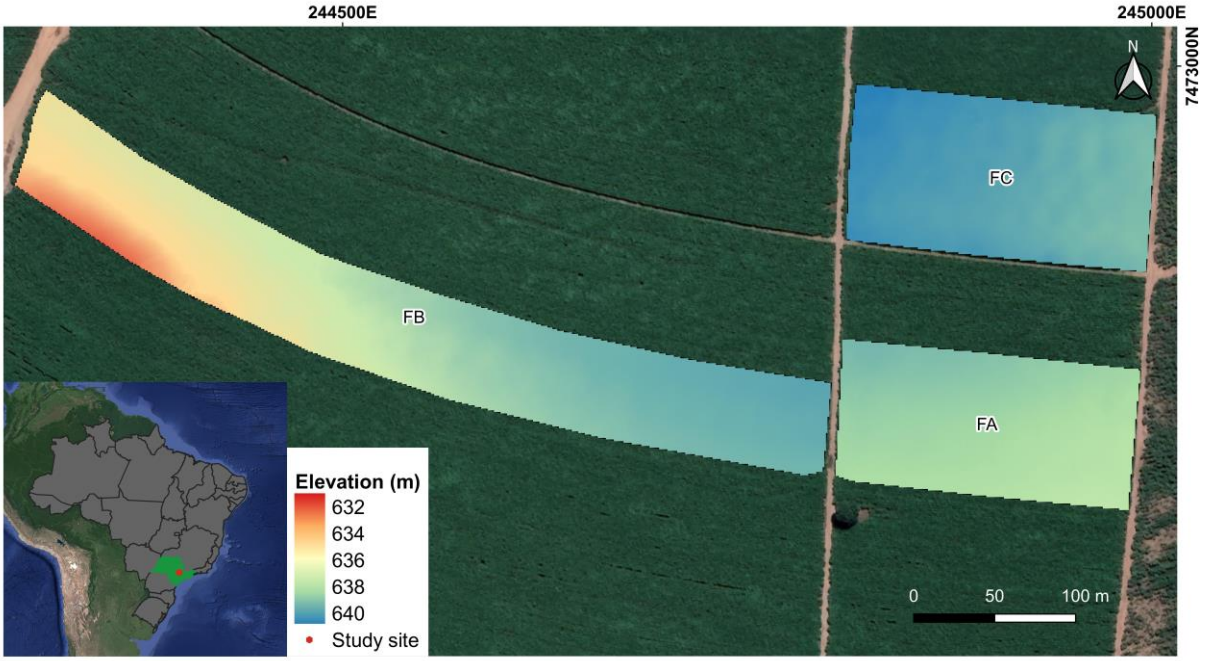

Figure 17. Location of the three study fields with relief details.

The volume of sugarcane billets carried by the elevator slats may be variable throughout the harvest, and therefore the distance between the sensor and the sample may be variable. However, Udompetaikul *et al.* (2021) evaluated the performance of calibration models developed for data collected at level of full, half cane delivery of the elevator slats, and global models composite for both conditions in a trial under static and controlled conditions. The authors showed that different levels affected the soluble solids content predictive accuracy of the model. However, combined model allowed to cover variations of this difference and was used to predict combined and unknown external sets resulting metrics close to those obtained for individual models. Therefore, in this study, the calibration models were built independently of the cane level on the conveyor, as well as in real field conditions.

Once the equipment software used in this study was designed for industrial purposes, its operation is programmed for batch times. The minimum batch time accepted by the software is 10 s. This condition was used in this study. Therefore, five to six spectra were recorded for one batch. Then, the software performs internally a principal component analysis, and Hotteling's T^2 test at 95% of confidence level. Spectra outside the confidence limit were automatically considered outliers and excluded of the data set. Then, the average spectrum is calculated from the remaining spectra, and this was concatenated with the respective geographic coordinate recorded at the end of this analysis.

The harvester mean speed was 1 m s^{-1}, and the conveyor belt speed was 2 m s^{-1}, common speed for the harvester elevator conveyor under real conditions and performed in another studies (PHETPAN; UDOMPETAIKUL; SIRISOMBOON, 2018; UDOMPETAIKUL; PHETPAN; SIRISOMBOON, 2021). Thus, it was necessary to perform an offset for all on-board measurement points, based on the sensor batch time (10 s) and the feeding time of the harvester (between the cut and the measurement at the elevator). The time between cutting the sugarcane, by the harvester base cutter, and the passage in the elevator, below the measurement platform, was assessed visually, equal to 10 s (feeding time). This time agrees with that reported in the literature for harvesters of the same manufacturer (MAGALHÃES; CERRI, 2007).

To perform the offset to the centroid of the harvested line it was sought to know the distance between each two consecutive points and the difference in collection time. Furthermore, we consider the total measurement time to be equal to the sum of the feed and batch time. Therefore, the initial position minus the horizontal offset along the track could be found and allocated as the corrected coordinate. The correction was based on the constrained method (DELEFORTRIE *et al.*, 2016).

4.2.3. Reference analyses and soil sampling

Samples were randomly collected throughout the fields during the on-board measurements for perform modelling and validation. At random times, the transshipment trailer moved ahead of the harvester and the sample was dropped directly to a canvas in the ground for collection (Figure 18a). Then, the collection point was recorded with a handheld GNSS receiver (Garmin 62 s, accuracy of 5 m, Figure 18b). The sampling was performed with a minimum spacing of two rows between each collection.

Figure 18. Sample collection procedure for modeling and validation. Sugarcane billets samples being dump on a canvas on the ground (a); georeferencing sample collection (b).

The composited billets samples were packed in plastic bags and sent to the laboratory, which it was stored in a freezer at 2ºC to minimize the effects of degradation of organic compounds by microorganisms and analyzed in the following day to determinate properties used to evaluate sugarcane industrial quality. The parameters evaluate were insoluble solids (Fibre), soluble solids content (Brix), apparent sucrose in the juice (Pol), estimated pol of cane and total recoverable sugars (TRS). All procedures for determining sugarcane industrial quality were performed as described by CONSECANA (CONSECANA, 2015), which agree with the international rules from International Commission for Uniform Methods of Sugar Analysis (ICUMSA). Also, after the sample milling in the laboratory, sub-samples of defibrated cane were collected for analysis with the NIR sensor under controlled bench conditions.

After the harvest of the three experimental fields, soil samples were collected in the superficial layer (0.00 to 0.20 m), with density of 1 sample per hectare in irregular grid on the crop line. Sample location was performed using the same GNSS used for on-board measurements and were submitted for physic-chemical wet laboratory analysis. The soil parameters determined were organic matter (OM), pH, extractable nutrients as phosphorus (P), potassium (K), calcium (Ca) and magnesium (Mg), hydrogen + aluminum (H+Al), cation exchange capacity (CEC),

sum of bases (SB), base saturation (V), sulfur (S), boron (B), copper (Cu), iron (Fe), manganese (Mn), zinc (Zn), clay, silt and sand. Details of the database collected in the present study are shown in Figure 19 and Table 9.

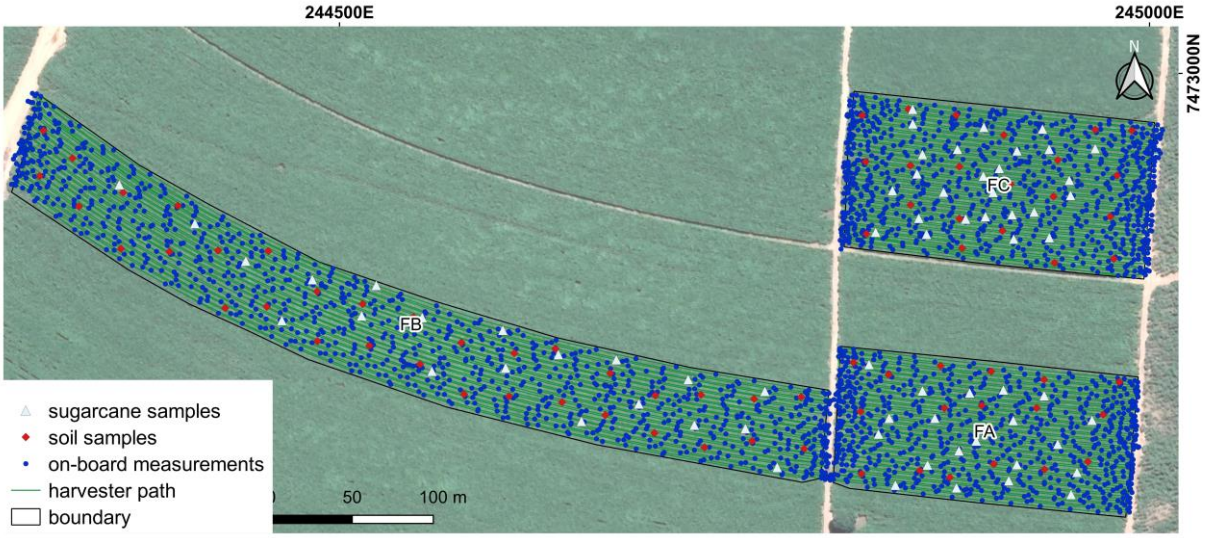

Figure 19. Location of experimental fields showing on-board measurement points measured during harvest, and sampling points for sugarcane and soil.

Table 9. Description of the database collected during harvesting in commercial sugarcane fields.

Field	Date	Area (ha)	# on-board measurements	# sugarcane samples	# soil samples
FA	09/26/2020	1.72	802	20	18
FB	09/27/2020	3.32	1109	20	34
FC	09/28/2020	2.02	977	26	21

4.2.4. Multivariate data analysis

The initial analysis was based on descriptive statistics of sugarcane quality values determined by conventional methods. Then, from the position and time at which each sample was collected, it was possible to identify the spectra corresponding to the samples. Thus, these spectra were isolated for modeling, corresponding to a first spectral data matrix with 66 samples. The spectral measurements of sugarcane samples, after being defibrinated in the laboratory procedure for reference analysis by conventional methods, constituted a second matrix corresponding to the same 66 samples described above. All measurements were made as absorbance values. The defibrated samples were analyzed at the same NIR sensor used for on-board measurements in bench condition, under controlled temperature environment (20 ± 5 ºC). Initially, the defibrated samples were arranged on a plate so that the surface was as even as possible. The spectral measurements were made in triplicate at three random points on the surface, so that each analyzed test sample was represented by its averaged spectrum. This procedure was performed for all test samples collected in the field on the same day as the laboratory analyses, providing two matrices corresponding to the same samples as defibrated cane and on-board samples. Also, the defibrated sugarcane samples were used to segment into calibration and external validation sets using the Kennard-Stone (KENNARD; STONE, 1969) algorithm applied to the spectral data and the reference data. The Kennard-Stone algorithm was

applied in each field samples data set individually, following the proportion of 70% of the samples for calibration data set, and 30% for external validation data set. This algorithm allows to obtain a representative calibration data set of sample population obtained (PASQUINI, 2018). Thus, the calibration data set was composed by 47 samples (14 of field A, 14 of field B, and 19 of field C), and 19 samples were destinated to the external validation data set (6 of field A, 6 of field B and 7 of field C).

The spectrum data were pre-processed by standard normal variate (BARNES; DHANOA; LISTER, 1989), which centers each spectrum on its mean and then scales it by its standard deviation, allowing eliminate the deviations caused by particle size and scattering. And it was applied the second derivative based on the Savitzky-Golay algorithm (SAVITZKY; GOLAY, 1964), with a window size of nine wavelengths and a second-order polynomial fitting, allowing to minimize hurdles such as baseline shifts drifts and to remove high-frequency noise from a spectrum and improve the signal-to-noise ratio (OLIVERI *et al.*, 2019). Also, the reference values were mean centered.

The piecewise direct standardization (PDS) (WANG, Yongdong.; VELTKAMP; KOWALSKI, 1991) was developed to apply calibration transfer method to measurements on the field conditions. PDS is a multivariate method that relates each wavelength of a master spectrum (e.g., defibrated sugarcane) and those of a secondary spectrum (e.g., field spectra). This method has been used with satisfactory results for calibration of soil spectra obtained by on-board sensing, based on spectra obtained for dry samples under bench conditions (off-line), allowing to eliminate noise and humidity effects, for example (JI; VISCARRA ROSSEL; SHI, 2015b; NAWAR; MUNNAF; MOUAZEN, 2020). The mainly advantages are for using a small number of samples in the transfer set, and for the noise-filtering effect due its multivariate characteristic (NAWAR; MUNNAF; MOUAZEN, 2020). It is a method proposed several years ago, but still often considered as a reference in studies evaluating new methods of model transfer because of its good performance (PASQUINI, 2018).

Based on the master and secondary spectra matrix, the algorithm performs the structured transformation matrix. Then, the transformation matrix is used to transform any spectra data set measured in another conditions, allowing calibration model transfer (WORKMAN, 2018). The method was implemented only in calibration data set. The absorbance of transfer samples of defibrated cane was related to the absorbance of field spectra that corresponded to the same wavelength and its neighbors within a predefined window (JI; VISCARRA ROSSEL; SHI, 2015b). After some tests, a window of 9 wavelengths was predefined for the present study. The PDS was performed based on the following three equations (Equations 3.1-3.3), as described by Workman (2018).

$$\mathbf{A}_{\mathrm{Pj}} = \mathbf{A}_{\mathrm{Cj}}\mathbf{b}_{\mathrm{j}} \tag{3.1}$$

where A_{Pj} corresponded to the absorbance response of standardization samples measured of defibrated cane at wavelength j; A_{Cj} corresponded to the absorbance response of standardization samples measured by on-board sensing at wavelength j; and b_j was the transformation (correction) vector of coefficients for the jth wavelength.

Then, the transformation matrix was calculated as following.

$$\widehat{\mathbf{T}}_{\mathrm{PDS}} = \mathrm{diag}(\mathbf{b}_1^{\mathrm{T}}, \mathbf{b}_2^{\mathrm{T}}, \dots, \mathbf{b}_{\mathrm{j}}^{\mathrm{T}}, \dots, \mathbf{b}_{\mathrm{k}}^{\mathrm{T}}) \tag{3.2}$$

where k corresponded to the total number of wavelengths in the standardization samples measured on both forms, as master and secondary spectra.

Thus, the transformed response for external validation samples and on-board measurements was estimated.

$$\hat{\mathbf{a}}^{\mathrm{T}} = \mathbf{a}^{\mathrm{T}}\hat{\mathbf{T}}_{\mathrm{PDS}} \tag{3.3}$$

where superscript **T** corresponded to the matrix transpose and **â** corresponded the estimated value for the measured sample **a**. The resulted matrix was used to transform the external validation data set and validate the models, and then, used to transform the on-board data set for the three experimental fields. Also, a principal component analysis (PCA) was performed as an exploratory analysis of the effect before and after the calibration transfer application.

The calibration models were development based on partial least square (PLS) regression (WOLD, S.; MARTENS; WOLD, 1983), in the same way that the development by Corrêdo *et al.* (2021a). Due to the small number of calibration samples, a leave-one-out cross-validation was applied here. The optimal PLS models were determined based on the lowest number of latent variables (LV), in which the root mean square error (RMSE) of cross-validation was not significantly higher than the minimum RMSE observed (Equation 3.4).

$$RMSE = \sqrt{\frac{\sum_i^N (y_i - \widehat{y_i})^2}{N}} \tag{3.4}$$

where N is the number of samples, y_i is the reference measurement of sample i, and $\hat{y}_i$ is the estimated result for sample i.

The outliers were evaluated on the calibration step based on high leverage and unmodeled residuals by Hotelling's T^2 and Q statistics. Samples with higher values for both statistics tests, at 95% of significance level, were removed from the spectral data set. Also, the reference values were evaluated by the RMSE of calibration (RMSEC). Samples that presented errors in prediction greater than $\pm 3 \times$ RMSEC were considered outliers and removed from the data set (VALDERRAMA; BRAGA; POPPI, 2007). The external validation samples were considered unknown samples. Therefore, outliers were evaluated only for the spectral data set. This procedure was carried out as recommended by international rules from ASTM- E1655-17 (ASTM E1655-17, 2017). All the modeling was performed in MATLAB framework (MATLAB R2015a, The MathWorks Inc., Natick, USA) and PLS- Toolbox 8.8 (Eigenvector Research Inc., Wenatchee, WA, USA).

The models were evaluated in terms of coefficient of determination (R^2, Equation 3.5), relative root mean square error (RRMSE, Equation 3.6) in calibration, cross-validation, and independent validation (external validation), and by the ratio performance of interquartile range (RPIQ, Equation 3.7).

$$R^2 = 1 - \frac{\sum_i^N (y_i - \widehat{y_i})^2}{\sum_i^N (y_i - \overline{y_i})^2} \tag{3.5}$$

$$RRMSE = \frac{RMSE}{\overline{y_i}} \times 100 \tag{3.6}$$

$$RPIQ = \frac{(Q_3 - Q_1)}{RMSE} \tag{3.7}$$

where y_i and $\hat{y}_i$ are the measured and predicted values for each ith sample, respectively. The $\bar{y}_i$ is the mean y values, and N is the number of samples. Q_3 and Q_1 are the upper and lower quartiles, respectively.

There is no well-designed method for filtering online measurements from spectral prediction models. For identifying outliers in the on-board data predicted by PDS-PLS, it was calculating the average distance to the k-nearest neighbors (KNN) in score space for prediction data. The high distance indicates samples which appeared in

regions of score space which were not well sampled by the calibration data. Thus, on-board samples with KNN value higher than 3 were considered outliers and removed from the dataset.

4.2.5. Spatial variability and site-specific assessment of the relations among sugarcane quality and soil attributes

The spatial analysis was based on geostatistics parameters and visual assessment over the spatial variability maps of sugarcane quality properties to identify patterns in the spatial variability, based on the values predict by the PDS-PLS models. The analysis was performed by ordinary kriging, which requires the adjustment of a variogram model. The experimental variograms for each quality properties of each commercial field could be obtained through (Equation 3.8):

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})}\sum_{\alpha=1}^{N(\mathbf{h})}\{Z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha)\}^2 \quad (3.8)$$

where, $\gamma(\mathbf{h})$ was the semivariogram for a distance vector (lag) $\mathbf{h}$ among the observations $z(\mathbf{x}\alpha)$ and $z(\mathbf{x}\alpha + \mathbf{h})$. $N(\mathbf{h})$ was the number of pairs separated by $\mathbf{h}$. The adjustment of semivariograms were performed individually testing spherical, exponential and gaussian models, selecting the one that shown the lower RMSE of cross-validation (MALDANER; MOLIN, 2020). Then, the data were interpolated into a 1 m grid by global point kriging. Geostatistics analysis, as well mapping editing were performed in QGIS 3.10.8 (QGIS Development Team 2018) software.

Based on the interpolated sugarcane quality maps, the values of each sugarcane quality parameters were extract in the same position which soil samples were collect. The soil parameters and sugarcane quality analysis were submitted to Pearson's correlation analysis, with 95% of confidence level, and to principal component analysis (PCA). Through PCA analysis it was sought to simplify the description of the set of variables, reducing their dimension to principal components, with the aim of simplifying the interpretation of cause-and-effect relationships. Thus, the method transformed the original variables into new uncorrelated variables in which each principal component is composed of the linear combination of the original variables (SANCHES; GRAZIANO MAGALHÃES; JUNQUEIRA FRANCO, 2019). The scores and loadings originated by the analysis were plotted according to the variance explained by the first two principal components (PASQUINI, 2018). Therefore, PCA allowed a robust qualitative interpretation of these relationships considering the specificities of each experimental area. These analyses were performed on the Rstudio software (R Core Team 2018).

4.3. Results and Discussion

4.3.1. Sugarcane quality properties characterization

Descriptive statistics of sugarcane quality properties show that there was a small variation in absolute values (Figure 20). Field C (FC) showed lower absolute mean values for all sugarcane quality properties than fields A and B (FA and FB, respectively). On the other hand, these values were higher for field B (FB) than the other fields.

The smallest interquartile distance was observed for field B, especially for Pol of cane and TRS. Also, when considering the complete data set used for the study (with data from all three fields), we observed variations, between the minimum and maximum values, ranging from about 4% for Brix, 4.9% for Pol, 3.2% for Fibre, 4.3% for Pol of cane, and 40 kg M^{-1} for TRS.

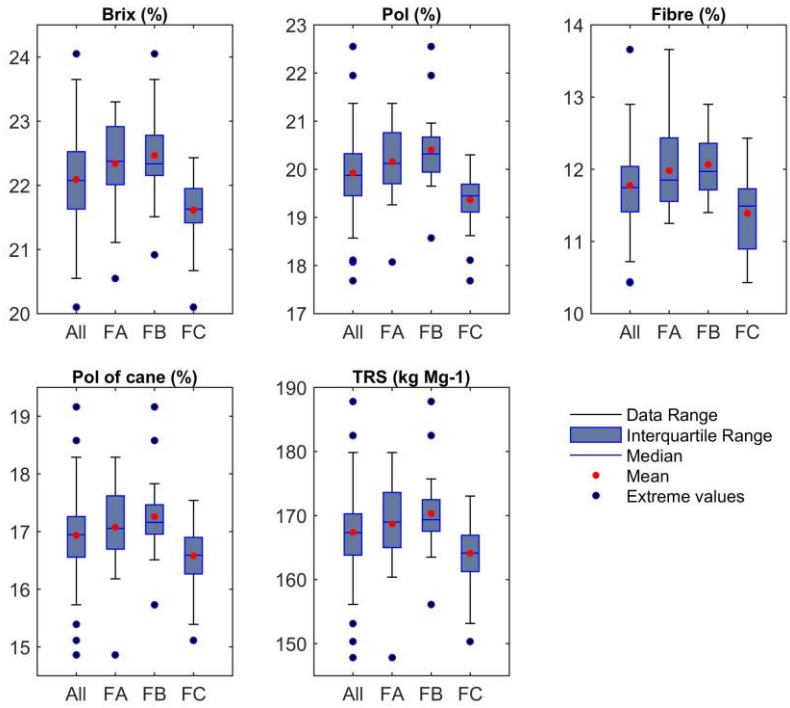

Figure 20. Boxplots of laboratory results for each quality property for all 66 samples collected from the three fields (All), and for the individual fields A, B and C (FA, FB and FC).

The Kennard-Stone algorithm for data segmentation allowed to obtain a calibration set whose maximum and minimum value range was equivalent to that of the complete data set (Table 10). This effect is desirable so that the calibration data set characterizes the existing variation in the entire data set that characterizes the study fields.

For all sugarcane quality properties, the CV value ranged from 3 to 5%, for calibration by cross-validation and prediction data sets. Also, the interquartile range, which 50% of the data occurs, showed differences ranging below 1% for Brix, Pol, Fibre and Pol of cane. Furthermore, this value for TRS ranged between 5.62 to 9.76 kg Mg^{-1} for cross-validation and prediction data sets, respectively. This variation was expected. Udompetaikul, Phetpan, and Sirisomboon (2021) achieved a C.V. around 3% with a calibration set consisting for 205 samples.

The effect of kurtosis values can be seen in the distribution histograms of the quality values determined by conventional methods in the laboratory, which did not show flattening (Figure 21). Moreover, Pol and Fibre showed a positive skewness, and Brix, Pol of cane and TRS presented a skewness closest to zero. Most sugarcane quality properties presented significant ($p < 0.05$) Pearson's correlation above 0.90, except for Fibre, which showed a non-significant ($p > 0.05$) correlation with all other properties evaluated. The high correlation between Brix and Pol values, for juice and sugarcane, is related to the constitution of soluble solids content in the sugarcane, estimated as Brix value, by mainly sucrose, estimated as Pol. Other soluble solids, such as reducing sugars and non-sugars, account for less than 2% of the total physicochemical constitution of sugarcane (COSTA *et al.*, 2021). Also, TRS is

estimated based on the sugarcane Pol, presenting high weight in its determination, besides the other sugars (fructose and glucose), with lower weight (CORRÊDO *et al.*, 2021a).

Table 10. Descriptive statistics of the sugarcane quality parameters for all collected samples (All data), for the cross-validation (47 samples) and prediction data set (19 samples).

Property	Data set	Min.	Max.	Mean	S.D.	C.V. (%)	Q_1	Q_3	Kurt.	Skew.
Brix (%)	All data	20.10	24.05	22.09	0.73	3.30	21.63	22.55	3.43	-0.06
	Cross-val.	20.10	24.05	22.14	0.79	3.57	21.63	22.58	3.34	-0.13
	Pred.	20.92	23.03	21.98	0.57	2.59	21.51	22.37	2.24	-0.09
Pol (%)	All data	17.68	22.55	19.92	0.85	4.27	19.50	20.35	4.15	0.20
	Cross-val.	17.68	22.55	19.93	0.93	4.67	19.50	20.39	3.99	0.21
	Pred.	18.57	20.94	19.88	0.65	3.27	19.40	20.30	2.30	-0.07
Fibre (%)	All data	10.43	13.66	11.77	0.61	5.18	11.42	12.04	3.54	0.22
	Cross-val.	10.43	13.66	11.77	0.62	5.27	11.50	11.99	3.92	0.46
	Pred.	10.44	12.67	11.78	0.60	5.09	11.37	12.25	2.50	-0.45
Pol of cane (%)	All data	14.86	19.16	16.93	0.74	4.37	16.56	17.30	4.41	0.00
	Cross-val.	14.86	19.16	16.95	0.79	4.66	16.57	17.19	4.42	0.03
	Pred.	15.73	17.87	16.90	0.59	3.49	16.37	17.36	2.05	-0.29
TRS (kg Mg^{-1})	All data	147.84	187.82	167.37	6.82	4.07	163.98	170.77	4.41	-0.05
	Cross-val.	147.84	187.82	167.53	7.32	4.37	164.05	169.67	4.45	-0.03
	Pred.	156.13	176.15	166.98	5.55	3.32	161.25	171.01	2.02	-0.28

Abbreviations: Minimum (Min.); Maximum (Max.); Standard Deviation (S.D.); Coefficient of Variation (C.V.); lower quartile (Q_1); upper quartile (Q_3); Kurtosis (Kurt.); Skewness (Skew.); Cross-val. (cross-validation); Pred. (Prediction or external validation); TRS (total recoverable sugars).

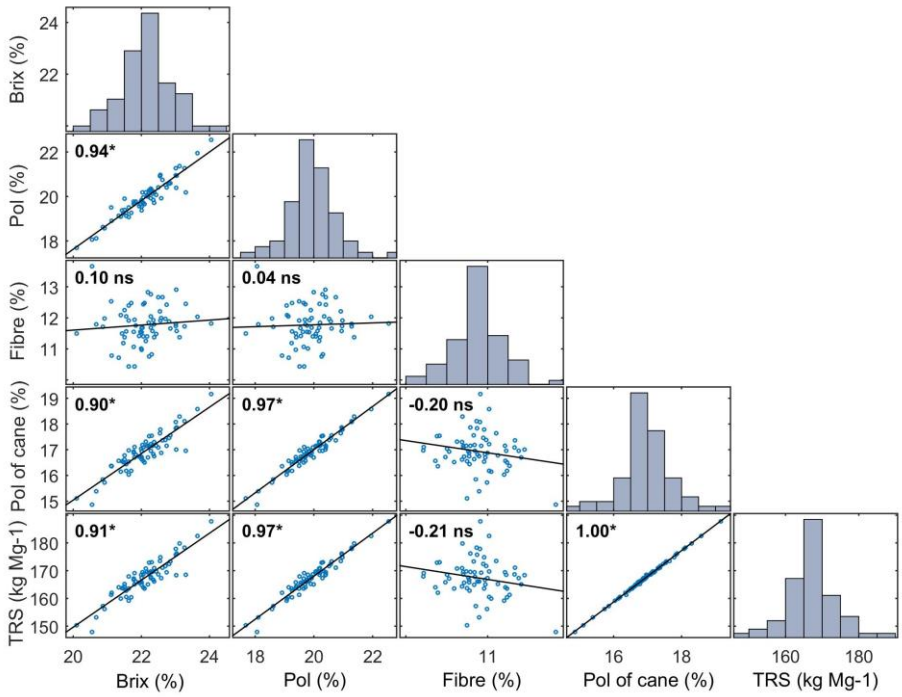

Figure 21. Pearson's correlation and distribution of frequency for sugarcane quality properties. * Significant correlation (*p*-value<0.05); ns non-significant correlation (*p*-value>0.05). Blue dots are the samples and black line is the correlation trend line. TRS (total recoverable sugars).

4.3.2. Sugarcane quality prediction models

The PLS regression method allowed to convert the dataset into a new multi-dimensional coordinate system (loadings) through the creation of a smaller number of orthogonal variables (latent variables - LV) (POMARES-VICIANA *et al.*, 2018; WOLD; SJÖSTRÖM; ERIKSSON, 2001). The differences among the first loadings for calibration models based on defibrated samples for sugarcane quality properties were visually very small, limiting itself to the percentage of variance explained by the first LV (Figure 22). The highest percentage of variance explained by the first latent variable occurred for Pol of cane, with 87.49% of the explained variance. On the other hand, the Fibre content obtained 99.30% of the spectral data variance explained by the first latent variable.

The first loading derived from the PLS calibration for the five sugarcane quality properties were very similar, where the smallest variation occurred between 900 and 1000 nm, largely attributed to saccharides and the third overtone of O-H, and normally associated with cellulosic Fibre (WORKMAN JR.; WEYER, 2012). Also, at around 960 nm there is a correspond effect of the third overtone of C-H stretching (GOLIC; WALSH; LAWSON, 2003), which may relate with polysaccharides such as sucrose. Then, between 1100 and 1200 nm an intermediate variation is evident. Between 1100 and 1230 nm occurs the second vibrational frequency overtones associated with C-H stretching and related to sugars (GOLIC; WALSH; LAWSON, 2003). Also, spectral bands at around 1170–1180 nm are related to the third overtone of C-H and unsaturated C=C double bonds, typically associated with Fibre, such as lignin (WORKMAN JR.; WEYER, 2012). Moreover, the highest variation identified for the wavelength range of the equipment used in this study was evidently between 1300 and 1450 nm, tending to stabilize close to zero of explained variance by the first latent variable after this interval. At 1360 nm there are effects possibly related to C-H combinations and the O-H first overtone (OSBORNE, 2006). Also, bands associated with sugars as

C-H and O-H related bands occurs at around 1450 nm, identified by the second overtone of O-H and polysaccharides lined to O-H (WORKMAN JR.; WEYER, 2012).

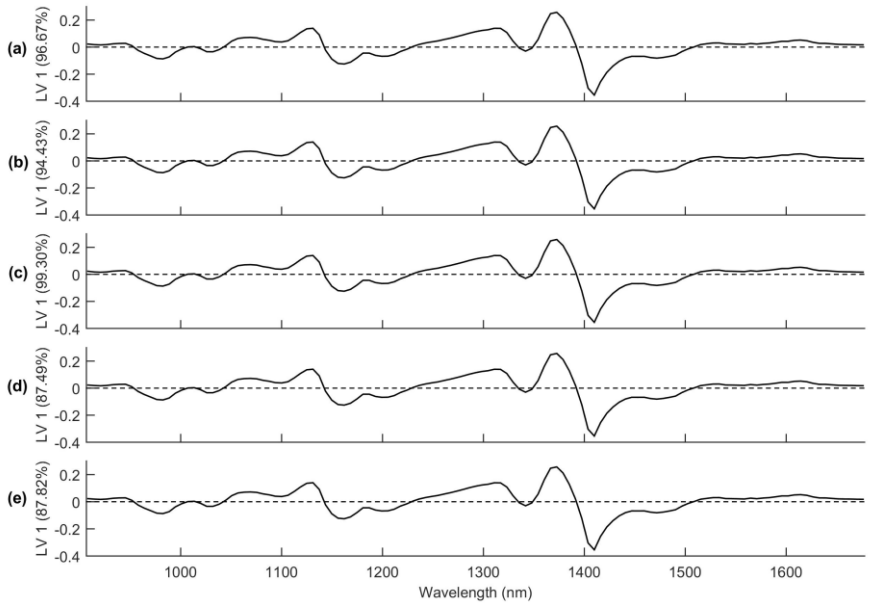

Figure 22. First Partial Least Square loadings for prediction of Brix (a), Pol (b), Fibre (c), Pol of cane (d), and total recoverable sugars - TRS (e) using near-infrared reflectance spectroscopy.

The sugarcane models based on defibrated samples allowed to obtain signals related to molecular vibrations of chemical bonds related to organic compounds such as saccharose. However, an in-field application of the NIR sensing must overcome intrinsic characteristics of the harvesting process, such as the presence of residues like straw and soil, the movement of the elevator conveyor belt, and other environmental factors thar could cause effects on the absorbance peaks related to the compounds of interest (CORRÊDO *et al.*, 2020; NAWI; CHEN; JENSEN, 2014).

The spectra obtained by on-board sensing showed substantial differences from those obtained under bench conditions, as expected (Figure 23a). The preprocessing of both data set, defibrated and on-board measurements, and the calibration transfer using the PDS method allowed a substantial reduction in these differences, reducing them to the order of 10^{-3} (Figure 23b). Part of this reduction could be attributable to the application of the second derivative to both spectral data. However, the PDS allowed the spectra obtained by on-board sensing to exhibit visually very similar characteristics to the spectra obtained on the bench, although it still maintains small differences in the bands corresponding to the absorbance peaks where there is most variance in the data (Figure 23c).

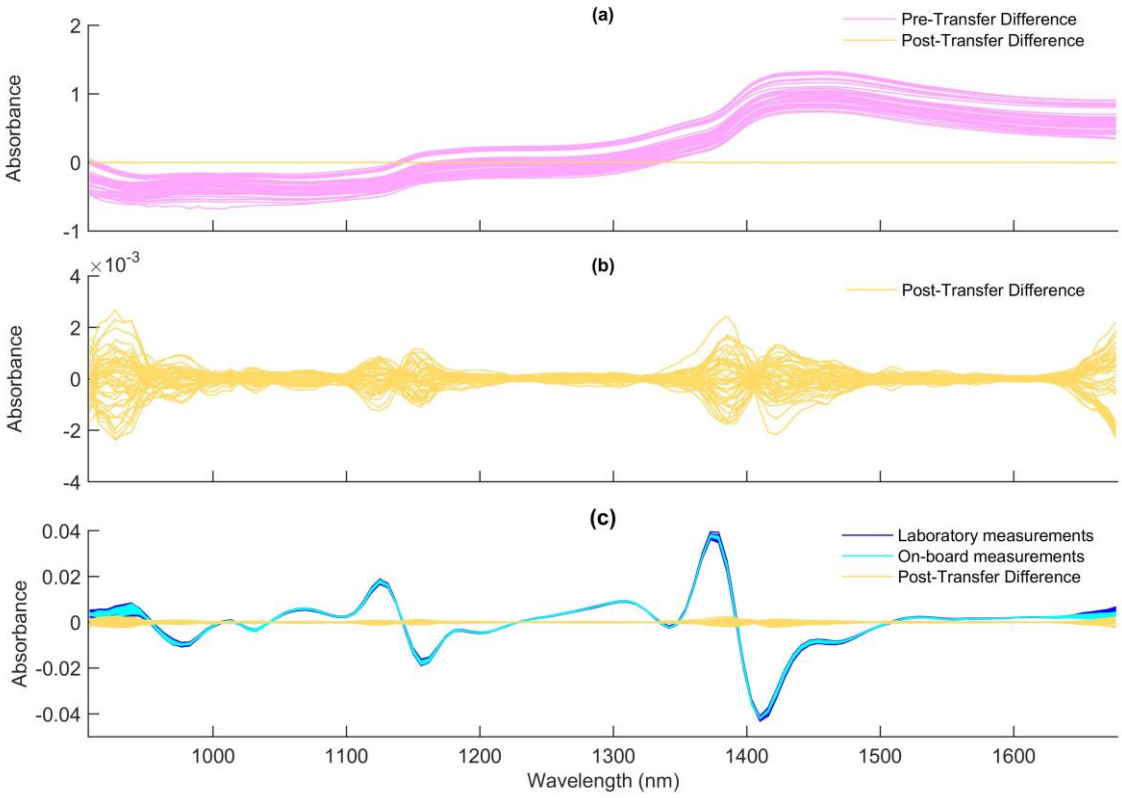

Figure 23. Differences among spectra of sugarcane billet samples with on-board readings by on-board NIR sensor during the harvest and after processing in the form of defibrated cane with bench readings before and after calibration transfer by picewise direct standardization (PDS).

The PCA analysis of the spectra from on-board measurements before and after the calibration transfer for the three fields, as well the on-board and defibrated spectral samples used to calibration of the models show both data sets in the same space in the score diagram after we apply the PDS method (Figure 24). As discussed previously, defibrated sugarcane samples measured in laboratory conditions were used as calibration transfer master data set. The scores of the laboratory spectra are plotted in a separate space to those of the spectra obtained by on-board sensing (Figure 24a), in which the clusters separation is evident with an offset along the principal component 2 (PC 2) due to the spectral differences between bench and online readings. The data variance obtained in on-board measurements may mostly be due to adverse environmental factors intrinsic of mechanized harvesting on the spectral measurements. On the other hand, after the PDS model application on the on-board measurements, both spectra data set are plotted within the space of the master data set (Figure 24b), i.e., defibrated cane spectra. Moreover, the first two components explained 90.73% of the data variance before the calibration transfer. Also, the data variance occurs mainly along the principal component 1 (PC 1) for both data sets. And after the calibration transfer application on the on-board measurements data set, based on the defibrated cane measurements, 99.60% of the data variance was explained by only the first component. This result indicates that the variance due to the difference between the spectral data corresponding to defibrated cane samples measured under controlled conditions and samples obtained with on-board measurements was substantially reduced.

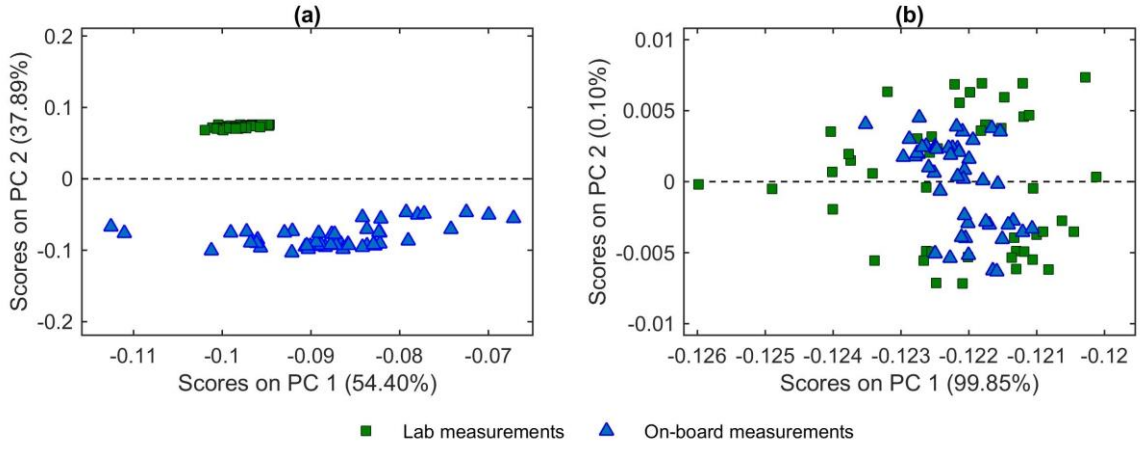

Figure 24. The principal component scores for on-board field prediction, and defibrated cane samples measured in the laboratory, before (a) and after (b) calibration transfer by Piecewise Direct Standardization (PDS). The percentage of variance explained by each component is shown on the axes in parentheses.

The calibration models, determined by leave-one-out cross-validation, employed 7 LV for all sugarcane quality properties, except for Fibre, with 4 LV (Table 11). The cross-validation by PLS allowed to obtain models with low bias values, with the highest values for TRS of 0.1712. Also, the RMSE of the cross-validation were less than 0.58% for Brix, Pol, Fibre and Pol of cane, and of 5.29 kg Mg^{-1} for TRS. Furthermore, the minimum R^2 values were of 0.33 for Fibre and the maximum of 0.65 for Pol.

The R^2 of calibration and cross-validation were used to determine the relation among reference and spectral values in the PLS models. Williams *et al.* (2001), apud Ruangratanakorn *et al.* (2020), described an interpretation for R^2 values for verification of the calibration models, which for values lower than 0.26, the NIR calibration is not usable. Values between 0.26 and 0.49 means a poor correlation of reference variables (quality parameter) accounted for spectral variables, and the reasons should be explored. Also, R^2 values of 0.50 until 0.65 could be interpreted by the possibility that the model discriminates high from low concentrations, as more than fifty percent of the variance in reference variables explained by spectral variables. Values between 0.66 and 0.81 mean a better potential prediction for an approximate quantification, and values between 0.82 and 0.90, and above 0.91 are interpreted as good and excellent prediction performance, respectively.

The calibration models for Brix and Pol showed good predictive performance (R^2 = 0.86), and the calibration models for Pol of cane and TRS showed predictive potential for an approximate quantification (R^2 = 0.80). The calibration models developed for Fibre may be interpreted as models that can enable the distinction of high and low concentrations (R^2 = 0.58). On the other hand, from the R^2 values for cross-validation for the quality parameters Brix, Pol, Pol of cane and TRS the models enable the distinction between low and high concentrations (R^2 = 0.50–0.65), while the cross-validation models for Fibre showed poor correlation between the reference values and the spectral variables (R^2 = 0.33).

Table 11. Results of partial least squares regression (PLSR) model performance in cross-validation, laboratory independent validation, and on-board validation for the prediction of solids soluble content (Brix), apparent sucrose in the juice (Pol), insoluble solids (Fibre), Pol of cane and total recoverable sugars (TRS).

		Brix	Pol	Fibre	Pol of cane	TRS
Calibration	LV	7	7	4	7	7
	RMSE*	0.29	0.34	0.36	0.35	3.25
	R^2	0.86	0.86	0.58	0.80	0.80
	Bias	0	0	0	0	0
Cross-validation	RMSE*	0.49	0.55	0.46	0.57	5.29
	R^2	0.63	0.65	0.33	0.51	0.50
	Bias	0.0057	0.0127	-0.0146	0.0186	0.1712
Laboratory validation (Defibrated samples)	RMSE*	0.35	0.50	0.49	0.46	4.27
	R^2	0.67	0.53	0.13	0.44	0.44
	Bias	0.0563	-0.0782	-0.1206	-0.0490	-0.3667
On-board validation (Transfer samples)	RMSE*	0.41	0.50	0.47	0.46	4.60
	R^2	0.33	0.28	0.18	0.32	0.19
	Bias	0.1402	0.0864	-0.0104	0.1505	0.7699

* Values in percentage for Brix, Pol, Fibre and Pol of cane, and in kg Mg^{-1} for TRS (total recoverable sugars). LV: Latent variables; R^2: coefficient of determination.

Results of this study have confirmed that the validation performed on defibrated samples with laboratory scanned spectra provided a higher predictive performance than that based on spectra obtained by on-board measurements. It is important to point out that the present study did not seek to achieve predictive performance equivalent to conventional measurement methods. An inconceivable reality if we keep in mind physical principles of NIR spectroscopy that optimize the measurement, such as uniformity of the sample surface and external interferences, such as oscillation by temperature, brightness, and distance from the sensor to the sample, that can be considered negligible (PASQUINI, 2018). However, we sought to present a calibration method that builds on these conditions to create a model for post-processing spectra collected on an on-board system with the closest possible performance in terms of accuracy.

The R^2 values for external validation data set were of minimum 0.13 for Fibre and of maximum 0.67 for Brix, for defibrated sugarcane samples analyzed on the bench. On the other hand, in the external validation of spectral data obtained in real time during harvest and transformed by PDS, the R^2 values were of minimum 0.18 for Fibre and of maximum 0.33 for Brix. The R^2 values of external validation for defibrated cane samples were lower than the values presented by Corrêdo *et al.* (2021a), with values of minimum 0.69 for Fibre prediction and maximum 0.83 for Pol prediction. Possibly, due to the greater variability of the calibration set in the mentioned study, from which sugarcane samples, included defibrated samples, were collected over a harvesting season. Low variability in the sugarcane quality values used for building the prediction models of this study possibly was caused by collecting only one sugarcane variety with a narrow range of days, reflected in the low R^2 values included for defibrated samples. Harvesting at different times during a harvest season, under different climatic conditions, allows for products with greater diversity (CORRÊDO *et al.*, 2021a).

Corrêdo *et al.* (2021a) evaluated different forms of sugarcane samples: billets, measured on the peel outer surface and cross-section, defibrated cane, and raw juice. Assuming that for on-board system development, absence

or minimal sample preparation is desired, the authors concluded that there was no statistically significant difference (p-value >0.05) among prediction accuracy (RMSEP) of the quality parameters evaluated from defibrated cane samples and whole billets. In addition, the predictive performance, evaluated by R^2 and RPIQ, for defibrated cane models was higher than for the other sample forms evaluated. The strategy of sugarcane collection samples at different periods throughout a harvest and from different varieties allowed Corrêdo *et al.* (2021a) to obtain greater variability in the contents of the quality parameters evaluated. The construction of robust PLS models requires high variability of the parameters to be predicted (PHETPAN; UDOMPETAIKUL; SIRISOMBOON, 2018). For future studies, we recommend that the trials be performed at different periods throughout a harvest and in different seasons, to obtain greater variability in the contents regarding product quality. However, both external validation models, i.e., laboratory and on-board measurements, has presented close or equal accuracy, determined by root mean square error of prediction (RMSEP) in the present study. It was possible to compare it with other previous studies based on the relative value of the accuracy, i.e., relative mean square error of prediction (RRMSEP). Also, the ratio of performance interquartile (RPIQ) values, which better represents the spread of the population (BELLON-MAUREL *et al.*, 2010), were used to compare the prediction performance among the measurements of defibrated cane made under controlled conditions (lab measurements) and the on-board measurements (onboard) transformed by the PDS model (Figure 25).

The RRMSE for Pol shown the same values for both measurements, i.e., in laboratory and on-board (2.51%). Furthermore, the values for Pol of the cane were very close, equal to 2.75% in lab and 2.73% for on-board measurements. For the relative accuracy of Brix and TRS close values were also observed. Brix and TRS measurements were equal to 1.86% and 2.75%, and equal to 1.59% and 2.56%, for on-board and laboratory measurements, respectively. The relative accuracy for Fibre prediction was slightly higher for on-board measurements (3.97%) than for laboratory measurements (4.16%). In addition, the RPIQ values allowed us to verify substantial similarities between the measurements reported under controlled conditions and online, with post processing by the calibration transfer method. The predictive performance for Pol and Pol of cane, were equal to 2.10 and 2.14 for on-board measurements, and equal to 1.81 and 2.13 for laboratory measurements, respectively. In addition, the RPIQ values for Brix and TRS were slightly higher for lab measurements (2.46 and 2.28, respectively), than for on-board system measurements (2.10 and 2.12, respectively). Fibre predictive performance was slightly higher for online measurements (1.88), than for laboratory measurements (1.79).

In a precursor study, Nawi *et al.* (2013) achieved RMSEP equal to 1.51% and 1.41%, i.e., relative accuracy of 8.47% and 7.91% for Brix prediction from NIR absorbance and reflectance measurements, respectively, for measurements performed on the peel outer surface under controlled conditions. In another study, based on cross-section measurements of billet samples under controlled conditions, Nawi, Chen and Jensen (2013) achieved RRMSEP values of up to 8.13% (RMSEP = 1.45%) for Brix prediction. Phuphaphud *et al.* (2019) obtained relative accuracy results of up to 5.49% (RMSEP = 0.63%) for Fibre prediction from measurements directly on the outer surface of the peel at the bottom section of sugarcane stalks. Maraphum *et al.* (2018), in a study on the effect of waxy material and measurement position in sugarcane stalks for Pol prediction, obtained relative accuracy results of up to 6.25% (RRMSEP = 1.20%). The results presented were inferior to those obtained in the present study, although performed under controlled conditions, possibly due to the calibration transfer procedure from defibrated samples. Furthermore, the preparation of samples through the removal of waxy material could make on-board sensing unfeasible, but according to the authors, it can optimize results for the implementation of NIR spectroscopy in sugarcane breeding programs.

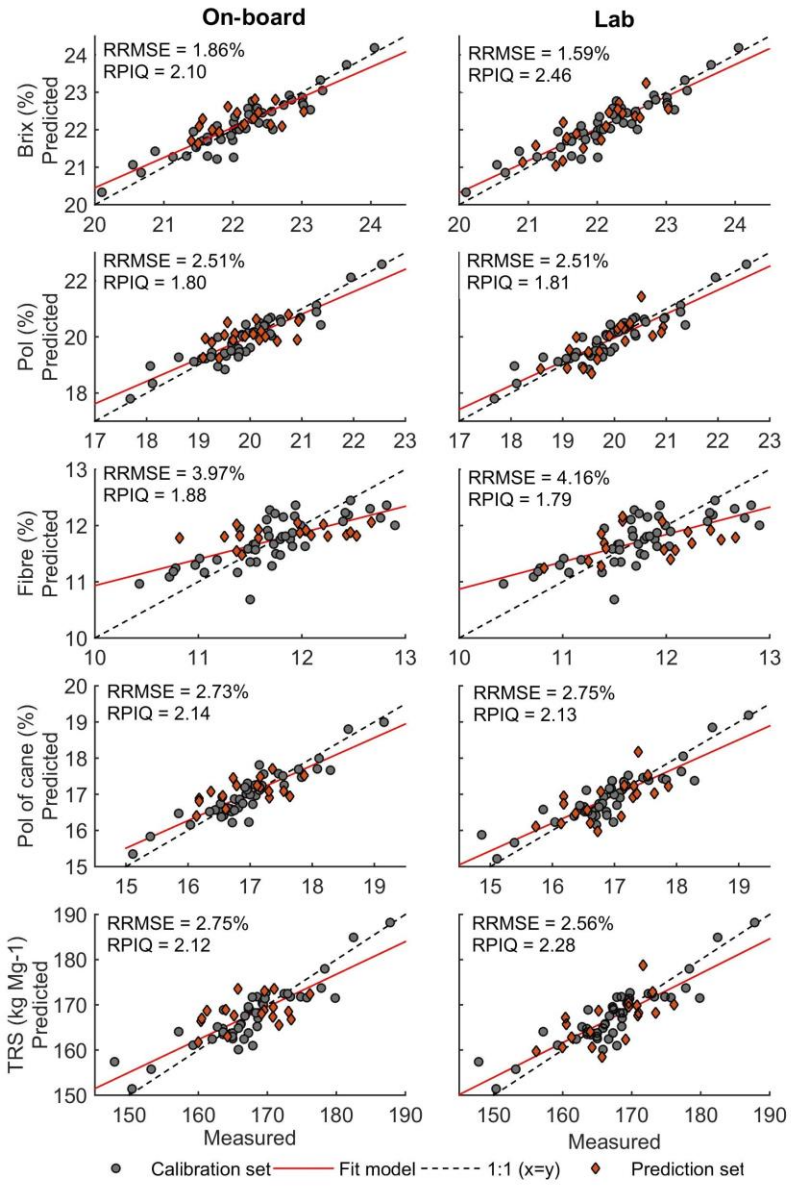

Figure 25. Predicted values by PLS-PDS models (on-board measurements) versus measured values by conventional laboratory methods (lab measurements) for sugarcane quality properties. TRS: total recoverable sugars.

Current studies on the use of a NIR spectroscopy equipment as an on-board sensor for soluble solids content (Brix) measurements were carried out by Phetpan, Udompetaikul and Sirisomboon (2018) and Udompetaikul, Phetpan and Sirisomboon (2021). To perform the study, the authors developed a static replica of a sugarcane harvester elevator, enabling the measurement of samples transported by the elevator belt under a measurement chamber. In the first one, they achieved metrics of R^2, RMSEP, and RPD equal 78.5, 0.30%, and 2.16, respectively. In the second study, the same authors have made an important contribution regarding the influence of different levels of sample being transported by the elevator. Three strategies were adopted. The first calibration model was built with full delivery data set, the second model with half delivery data set, from which each calibration set being validated with both full and half delivery data sets, and the third one was built with a global calibration model, composed of both data sets, and validated with mixed and unknown sample data sets. The best results were found for calibration models developed from the full delivery sample set and with validation from similar set samples

(R^2 of 0.81 and RMSEP of 0.30). Also, the authors concluded that different delivery levels affected the predictive accuracy. However, the third strategy yielded a satisfying result, with an R^2 of 0.69 for combined prediction data set, an R^2 of 0.56 for unknown data set, and an RMSEP of 0.42 for both. Moreover, the combined model tended to overcome effects of different cane levels.

In real field conditions, the control of cane levels in the elevator may make the implementation of an on-board system unfeasible. The present study allowed a further advance in that our system was subjected to the harsh field environment and rugged harvesting conditions. In addition, there was no control over the delivery levels, the intrinsic impurities of the harvest such as straw, dust, and other residues. In relative numbers, Udompetaikul, Phetpan and Sirisomboon, (2021) obtained relative accuracy (RRMSEP) equal to 1.86% and 1.83% for soluble solids (Brix) prediction from combined validation sets (full + half delivery sets) and unknown set, respectively. For the prediction of the same quality attribute, our study showed a relative accuracy of 1.86%, like that obtained by these authors under controlled conditions. Therefore, this study confirmed the potential of the PDS calibration transfer method adopted in this study to overcome the intrinsic adversities of mechanized sugarcane harvesting to achieve accuracy equivalent to that of previous studies conducted under controlled conditions.

4.3.3. On-board data analysis and spatial variability

Geostatistical analysis allowed to show that sugarcane quality presented spatial dependence, i.e., the sugarcane quality properties were not random across the fields, but also spatially correlated (Table 12). However, the nugget variance (C_0) represents a considerable portion of the sill variance ($C_0 + C_1$). This means that a portion of the variance is not spatially related. Furthermore, nugget-to-sill allowed to classify the spatial dependence according to classification proposed by Cambardella *et al.* (1994). Brix shown moderately spatially dependent in the three fields ($0.25 < C_0/(C_0 + C_1) < 0.75$). In the field A (FA), all sugarcane quality properties showed moderate spatial dependence, except Fibre, which shown weak spatial dependence ($C_0/(C_0 + C_1) > 0.75$). For the field B (FB), except Brix, all quality properties were weakly spatially dependents. For field C (FC), besides Brix, Pol was moderately spatially dependent, and the other properties were weakly spatial dependent.

Rodrigues, Magalhães and Franco (2013) fitted mathematical models to predict Brix, Pol and Fibre from physical-chemical soil parameters and leaf nitrogen as predictors variables. The models presented low to moderated values of R^2, ranging from 0.06 to 0.18 for Pol and Fibre prediction, respectively, in a first harvest evaluation season, and from 0.19 to 0.52 for Pol and Fibre prediction, in a second evaluation season. Also, the models allowed to describe the spatial structure of Brix, Pol and Fibre within the experimental field, with small errors. Moreover, the models allowed obtaining RMSE values lower than 1%. In relative metrics, the values were 2.36%, 3.06%, and 2.77%, for Brix, Pol and Fibre in the first harvest season, and 3.58%, 5.02%, and 5.32% for Brix, Pol and Fibre in the second harvest season, respectively. Also, these results allowed to recognize weakly to moderately spatial dependence of practically all the quality parameters evaluated, for observed and predicted values, except for Fibre, with strongly spatial dependence for predicted and observed values on the second harvest season.

Table 12. Semivariogram model parameters of sugarcane quality properties used for mapping the spatial variability for the three experimental fields.

Field	Property	Model fit	C_0	$C_0 + C_1$	A (m)	$C_0/(C_0 + C_1)$
FA	Brix (%)	Exponential	0.060	0.098	33.62	0.61
	Pol (%)	Spherical	0.091	0.223	71.53	0.41
	Fibre (%)	Gaussian	0.036	0.043	92.93	0.84
	Pol of cane (%)	Spherical	0.071	0.161	75.74	0.44
	TRS (kg Mg^{-1})	Spherical	6.250	13.019	73.78	0.48
FB	Brix (%)	Exponential	0.252	0.363	64.48	0.69
	Pol (%)	Spherical	0.349	0.444	62.98	0.79
	Fibre (%)	Spherical	0.108	0.117	264.23	0.92
	Pol of cane (%)	Spherical	0.293	0.360	66.12	0.81
	TRS (kg Mg^{-1})	Spherical	25.176	31.016	66.01	0.81
FC	Brix (%)	Exponential	0.381	0.517	38.76	0.74
	Pol (%)	Exponential	0.500	0.629	58.34	0.79
	Fibre (%)	Spherical	0.136	0.212	118.73	0.64
	Pol of cane (%)	Spherical	0.396	0.508	93.10	0.78
	TRS (kg Mg^{-1})	Spherical	34.424	44.395	97.67	0.78

* Header abbreviations: Nugget variance (C_0); spatial dependent variance (C_1); range (A); nugget-to-sill [$C_0/(C_0 + C_1)$]; total recoverable sugars (TRS).

A moderately spatial dependence for Brix and Pol predicted by NIR spectroscopy from defibrated sugarcane samples collecting manually across a commercial field was observed by Corrêdo *et al.* (2021b). The models developed in that study allowed ranges of 122 m for Brix and 124 m for Pol. The weak spatial dependence of some properties in this study must be due errors during the real time NIR measurement. Also, we observe linear patterns for the same quality parameter value in the crop direction, more evident for field A. This effect can be due to measurement error, such as the presence of dirt on the sensor window, or due to errors in crop management, such as nitrogen fertilizer application, for example, which is the main nutrient influencing sugarcane quality (DE CASTRO *et al.*, 2019; GOPALASUNDARAM; BHASKARAN; RAKKIYAPPAN, 2012; MUCHOW *et al.*, 1996). However, as discussed earlier, the models' accuracies were equivalent to defibrated sugarcane measurements in laboratory conditions due the calibration transfer process. In addition, despite possible measurement errors inherent to the limitations of the technique, such as variable sensor to target distance, presence of dirt, brightness, among others environmental challenges inherent of the mechanized harvesting, a spatial structure is perceptible in all variograms, even in small areas (1.7 to 3.3 ha), such as the experimental fields evaluated in the present study. Moreover, the range of sugarcane quality properties were approximately 33-264 m. If we consider only the parameters related to sugar concentration (Brix, Pol, Pol of cane and TRS), this value was at most 98 m. This indicates that there was variability at short distances, especially for Brix (range 34 to 65 m), but there were also large regions in the fields with distinct quality contents. A similar conclusion was found by Johnson and Richard (2005) regarding the presence of spatial correlation for yield data and sugarcane quality parameters in a study conducted in Louisiana (USA). Also, they obtained results ranging from 26 to 133 m for quality sugarcane properties, in two fields across three consecutive

years of study. On a recent study, Catelan *et al.* (2022) showed strong spatial dependence of the properties Brix, Pol, Fibre and TRS, as well as of the sugarcane yield from low sample density, collected manually, in an area of 445 ha.

Sugarcane quality maps showed evident effects of variability across the experimental fields (Figure 26). Also, the variability among the different fields is even more evident. Field C has a larger portion of area with low concentrations of all quality parameters than fields A and B. This trend can be seen in Figure 20, which the values measured in the laboratory for the samples collected for calibration and validation of the prediction models shown a trend of lower contents from quality parameters in the C field compared to the fields A and B. This indicates that the modeling allowed to discriminate high to low values of sugarcane quality properties, even within small areas, and especially over larger distances. For the parameters Brix, Pol, Fibre and Pol of cane, it was possible to verify relative variations of 7.14%, 8.91%, 7.63% and 9.27%, respectively, between minimum and maximum values in each field. As previously discussed, to advance in the application of the sensing technique it becomes essential the on-board sensing of extensive areas, in different moments throughout a harvest season, and between different harvests seasons. The acquisition of such a database will enable the development of more robust calibration models, with greater variability of quality properties. For the TRS it was possible to verify relative variations of 8.62% (14 kg Mg^{-1}). The use of this parameter combined with yield data, may allow the acquisition of maps of variability of the amount of sugar produced per area. It constitutes valuable information for the agronomic management of agricultural inputs aiming at the quality of the product supplied to the mills and for the financial management of the entire operation.

Moreover, it is possible to observe some patterns in the shape of the variability zones for each plot, for the different quality parameters, mainly for those regarding sugar contents (Brix, Pol, Pol of cane, TRS). Possibly, this effect occurred due to the existing correlation between these parameters (CORRÊDO *et al.*, 2021a). Regarding the experimental areas used in the study, it should be considered that they are small fields nearby, and the sensing technique captured the existing variability between high and low sugar contents. In future studies, the evaluation of this methodology in extensive commercial fields, across more than one harvest season, may enable clearer identification of spatial patterns of lower and higher values of sugarcane quality parameters in the fields, enabling to evaluate spatial and temporal variability for large commercial fields. In addition, it is recommended to evaluate the sensor acting at different distance settings from the target to minimize obstruction of the sapphire window by dirt, to evaluate different settings and power of the external illumination set, a higher frequency of data collection and to evaluate the influence of temperature variation on real-time measurements.

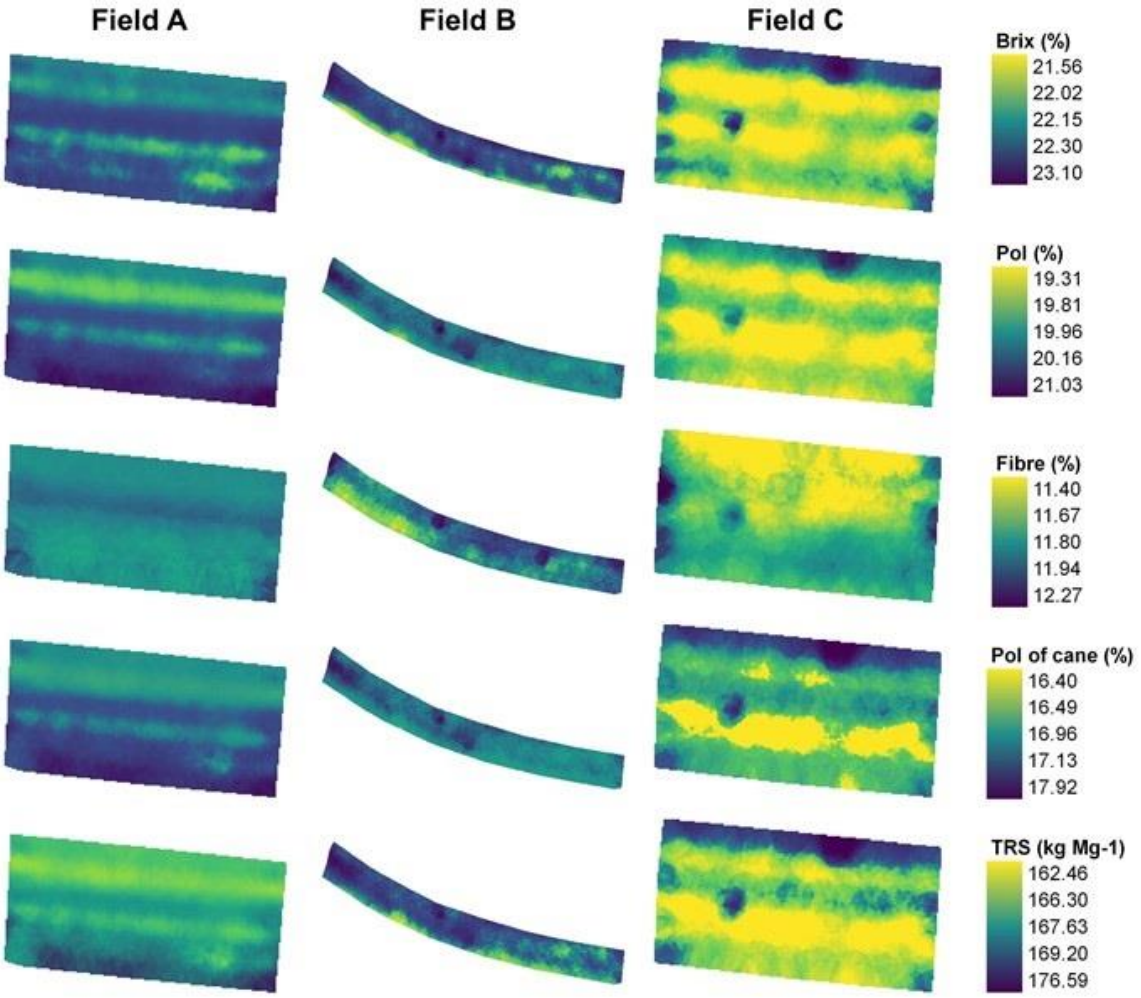

Figure 26. Sugarcane quality spatial variability maps for the three experimental fields measured by on-board NIR sensor on the harvester and predicted by PLS regression models combined with picewise direct standardization calibration transfer method.

4.3.4. Cause and effect relationships: sugarcane quality and soil parameters

Some attributes presented high variation of the contents within an experimental field, and on the other hand, low variation in another experimental field, as we could verify from interquartile range, such as P, CEC, and Cu, for example (Figure 27). Other attributes presented close average values for both experimental fields, such as K, Ca, S, B, Fe, Mn, and Zn. Also, there was no obvious trend between the variations of micronutrients (S, B, Cu, Fe, Mn and Zn) and sugarcane quality properties among the experimental fields. The most evident differences among both experimental fields could be observed from the interquartile range of the OM, Clay and Sand contents. Moreover, there was an apparent variation of the sugarcane quality properties predicted by the PDS-PLS models and kriging interpolation for estimated values in the same site-specific of soil samplings among the different fields, especially for fields FA and FC.

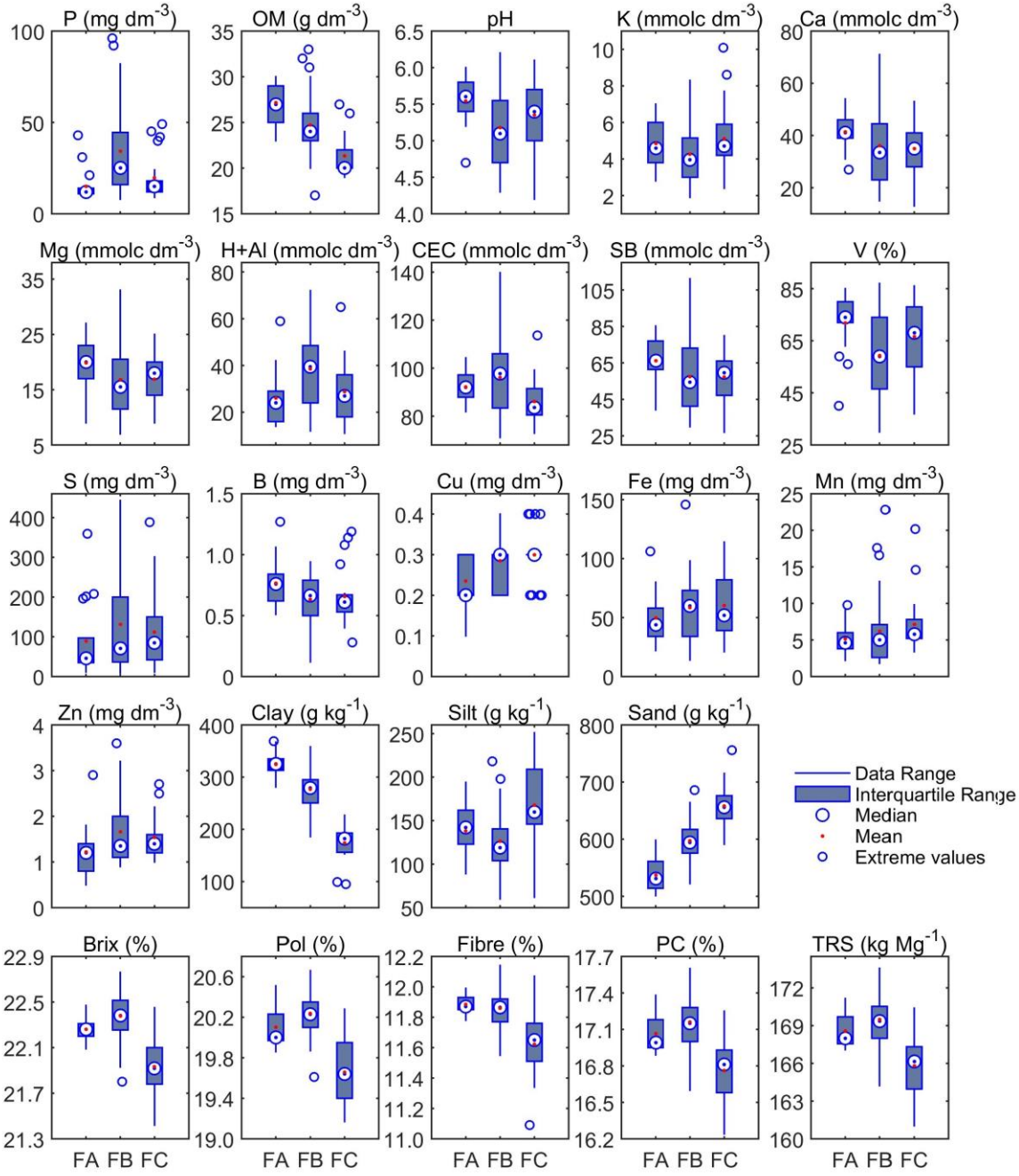

Figure 27. Boxplots of sugarcane quality properties and soil attributes used in this study for each experimental field (FA, FB, and FC).

There was a slightly variation of the quality parameters within FA, concentrated in higher values compared to the other experimental fields. Similar behavior was observed for OM, pH, V, and especially clay. On the other hand, there was more variation of quality parameters within FC, with higher concentration of low values compared to the other experimental fields. This trend was similar to that shown by OM and clay values. For sand values, this trend was inverse in both cases. Nonetheless, there was no significant correlation (p-value > 0.05) between texture variables and sugarcane quality properties for any experimental field (Figure 28).

The correlation of the quality parameters was maintained after prediction by the sensing technique and interpolation of the data by geostatistical methods. Brix, Pol, Pol of cane (PC) and TRS showed significant positive

correlation (*p*-value < 0.05) with each other, and Fibre showed non-significant correlation (*p*-value > 0.05) with the others sugarcane quality properties. Similar behavior to that presented by the descriptive statistical analysis of the results obtained by conventional laboratory analysis (Figure 21), and by other studies (CORRÊDO *et al.*, 2021a).

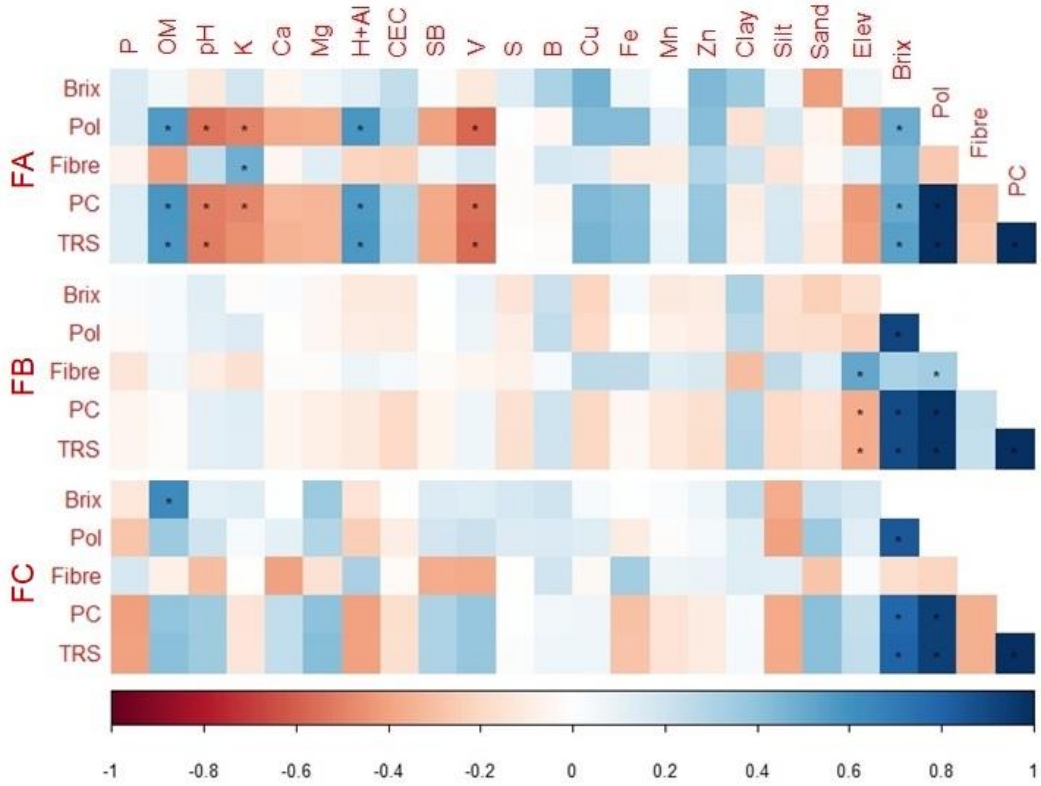

Figure 28. Pearson's correlation matrix between different soil attributes and sugarcane quality properties for each experimental field (FA, FB, and FC). * Significant correlation (p < 0.05).

Soil parameters OM, H+Al and sugarcane quality properties Pol, Pol of cane and TRS presented significant positive correlations (p-value < 0.05), and the same quality parameters presented negative correlations (*p*-value < 0.05) with pH and V in FA. In addition, K showed positive correlation (*p*-value < 0.05) with Fibre and negative correlation (*p*-value < 0.05) with Pol and Pol of the cane, in the same experimental field. The other parameters showed weak to moderate and non-significant (*p*-value > 0.05) correlations with the quality parameters evaluated. On the other hand, in field B, all soil attributes showed weak to moderate and non-significant correlations with sugarcane quality parameters, except elevation, with moderate and positive correlation with Fibre content (*p*-value < 0.05) and negative with Pol of cane and TRS (*p*-value < 0.05). Also, OM, pH, K, H+Al and V, which presented significant correlation with quality parameters, shown weak (near to zero) and non-significative (*p*-value > 0.05) correlation at FB. For FC, weak to moderate and non-significant (*p*-value > 0.05) correlations were observed between all sugarcane quality parameters and all edaphic attributes evaluated.

The correlations between quality properties and soil parameters showed no similarities for the different experimental fields, even for sugar-related parameters, except for Pol, PC and TRS on the FA. A similar conclusion was observed by Johnson and Richard (2005) for each one of the two fields evaluated by them. The authors attributed this effect to the differences of the soil type and crop age. In one of the evaluated fields, S, OM and Ca/Mg ratio presented the best inverse correlations with yield and quality parameters. On the other hand, in a

second experimental field, K, pH and OM presented the best results for negative correlation with the same parameters. The authors attribute the negative correlations to two factors. The increases in vegetative development in some portions of the field could imply a decrease in sugar accumulation, or in portions which sugar accumulation was greater, a greater depletion of soil nutrients could have occurred. Also, increases of OM should be associated with increases of vegetative yield, and consequently, occurred the strong negative correlations with sugar-related parameters accumulation by the plants (JOHNSON; RICHARD, 2005).

A principal component analysis shown the loadings of the variables and scores of samples (Figure 29). This analysis was based on the spatial distribution of quality properties and soil attributes. The first two principal components (PC1 and PC2) explained 29.3% and 22.3% of the data variance, respectively. Also, the projection of variables onto the factor plane suggests that sugar-related parameters (Brix, Pol, PC and TRS) were directly related, positively related to PC1 and PC2, with the variance explained mainly by PC2. Furthermore, Fibre content shown the same effect, but with less intensity, represented by the lower vector in relation to the other sugarcane quality parameters. Also, the soil parameters Clay, OM and CEC presented vectors projected on the positive direction of both principal components, as well. Moreover, P and S showed some relation with these parameters, due the variance explained mainly by PC2 in positive direction. On the other hand, the soil parameters Sand and Silt were inversely related with sugarcane quality properties, projected on negatively direction of both principal components. Also, elevation (Elev) as inversely related mainly with Fibre content, with most of the variance explained by PC2. Others soil properties showed variance explained mainly by PC1, and intermediately related with the quality parameters.

The soil attributes with the related variance to sugarcane quality properties, Brix, Pol and Fibre described by Rodrigues, Magalhães and Franco (2013), in two crop cycles, were OM, P and K. In the present study, K was not related to any quality parameters. Other studies point out OM as the most important attribute to impact sugarcane yield spatial variability (SANCHES; GRAZIANO MAGALHÃES; JUNQUEIRA FRANCO, 2019). Potassium (K) allow the synthesis and translocation of proteins and carbohydrates and the storage of sucrose (GOPALASUNDARAM; BHASKARAN; RAKKIYAPPAN, 2012). It is the most abundant element in the plant tissue (CHERUBIN *et al.*, 2019) and, consequently in the cane juice, being responsible to improve sucrose recovery through an increase in the Pol and reduction in Fibre content (MEYER; WOOD, 2001). The low correlation of K to the quality properties, shown by the low representation of this nutrient by the first two PCs (showed by an imperceptible loading vector), and low concentrations identified by laboratory analysis (Figure 27) may be attributed to the extraction of the nutrient from the soil by the plants, due to the samples being taken soon after harvesting.

Phosphorus (P) contributed to a small portion of variance explained by the first two PCs, mainly by PC2 as for both quality parameters. This nutrient is important to photosynthesis, rooting and to increase tillering, for example, and it may allow significantly increase both cane yield and quality (GOPALASUNDARAM; BHASKARAN; RAKKIYAPPAN, 2012). Mainly by the enhanced of sugar content and purity of juice, measured by the percentage of sucrose in the solids soluble content (ELAMIN *et al.*, 2007). Other attributes as SB, H+Al, pH, V, Ca, Mg, and micronutrients (Fe, Cu, Mn, Zn, B) showed intermediate to high loadings, with variance explained mainly by PC1. Also, presented low correlation with quality parameters, represents by the closest to right angles between these soil attributes and the quality properties meaning lack of correlation. These attributes are necessary for full crop development, as Ca it is necessary to cellular structure of cells, Mg is a component of chlorophyll, S compose amino acids, for example (CHERUBIN *et al.*, 2019). However, minor nutrients produce no substantial increases or decreases in sucrose content (MEYER; WOOD, 2001).

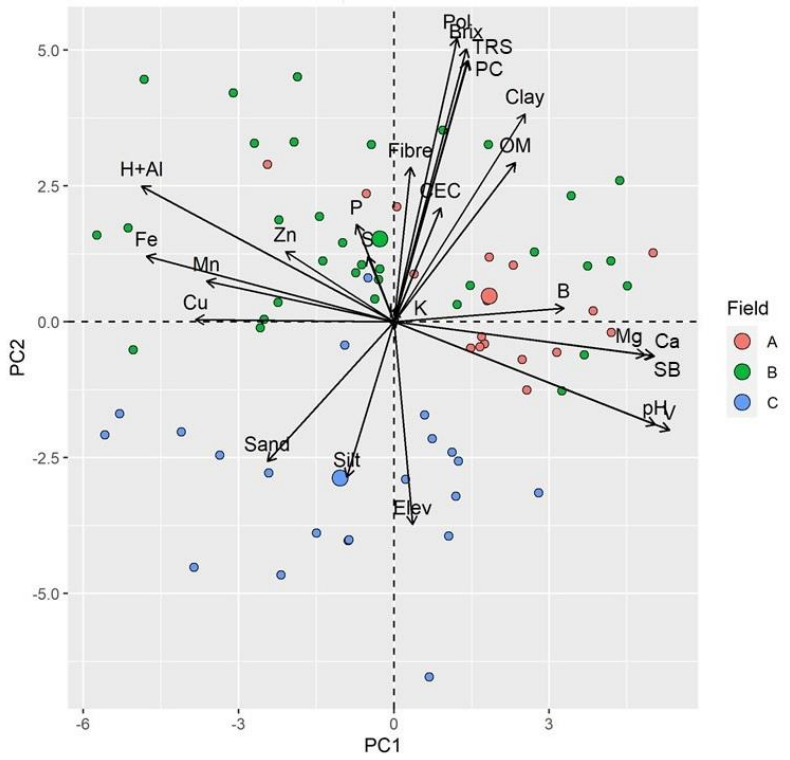

Figure 29. Principal component analysis (PCA) of soil attributes and sugarcane quality properties shown the loadings of the variables (vectors) and scores of samples (dots) for the first two principal components.

The scores of samples showed a clearly evidenced clusters of the fields, mainly for FC. FA scores presents itself somewhat merged with FB scores. Also, the soil parameters most related with FC were Sand, Silt and Elev. On the other hand, OM, Clay and CEC were more related with fields FA and FB. These results agree with the descriptive statistics that showed sandier soils in the field FC and clayey soils in the fields FA and FB (Figure 27). The positioning in nearly opposite directions between the attributes on the first quadrant (Clay, OM, and CEC), and those on the third quadrant (Sand and Silt) indicate a high negative correlation between these attributes. In this way, this interpretation is most evident between Clay and Sand. These physical soil attributes, as well OM, may be related to apparent electric conductivity (ECa), from which is related to water content on the soil (KITCHEN *et al.*, 2005). Marques *et al.* (2014) report a positive correlation between ECa and Clay, and between ECa and sugarcane quality properties Brix, Pol, Fibre and TRS. The authors point out, based in their results, that ECa may be used to identify areas with different potential to sugarcane quality. These results corroborating with those obtained by Catelan *et al.* (2022). These authors evaluated ECa to identify areas in the field with different potential for sugarcane yield and quality from soil texture, aiming to use it as a tool to direct the planting of varieties according to the potential of the area. The results showed a trend of higher values of sugarcane quality properties in site of higher ECa values, and a trend of lower results of sugarcane quality when ECa decrease. Therefore, these conclusions corroborate with the results obtained in our study, from which higher correlation between sugarcane quality properties were observed with Clay and OM, and an almost inverse correlation with Sand and Silt. Soil texture had an important effect on the field classification as could observed by the PCA scores. Moreover, FC scores were plotted in the opposite direction to the quality properties vectors. As demonstrated by the on-board sensing, FC showed lower concentrations of the sugarcane quality parameters than fields FA and FB (Figure 26), also evidenced by the values quantified in the

laboratory for the samples collected in the field (Figure 20), once again suggesting a relationship between the quality parameters and textural variations in the soil.

Through the cause-and-effect relationships observed between sugarcane quality properties and soil attributes, it was evident that the on-board NIR sensing method allowed the identification of regions with different sucrose yields. This way of acquisition of sugarcane crop data represents an additional advance for yield monitors, as it makes it possible to obtain spatialized sugar data per area. This information would enable trading and payment between producers and mills closer to the reality of production, assist in issues of harvest logistics and delivery to the mill, industrial planning, and above all, site-specific agronomic management of sugarcane crops based on product quality.

4.4. Conclusions

This study has shown that it is possible to overcome adverse effects of on-board sensing on sugarcane harvesters and provide spatial data of crop quality. Piecewise direct standardization allowed transferring NIR calibration models developed by partial least square regression on the bench from defibrated cane samples to perform post-processing of data obtained by the on-board sensor in the harvester. Furthermore, the on-board sensor and the proposed method for prediction allowed the spatial variability of sugarcane quality attributes to be characterized. The results were consistent with previous studies. Moreover, the analysis of relationships with soil attributes allowed the identification of trends with textural attributes, already reported in other studies, which strengthens the conclusion of the effectiveness of the proposed method for onboard proximal sensing of sugarcane quality. Moreover, further studies should be conducted to evaluate the method with collections in different months throughout a season and in different harvests, seeking to increase the robustness of the prediction models.

References

ABASI, S.; MINAEI, S.; JAMSHIDI, B.; FATHI, D. Dedicated non-destructive devices for food quality measurement: A review. **Trends in Food Science and Technology**, vol. 78, p. 197–205, 1 Aug. 2018. https://doi.org/10.1016/j.tifs.2018.05.009.

ASTM E1655-17. **Standard Practices for Infrared Multivariate Quantitative Analysis**. [*S. l.*: *s. n.*], 2017. https://doi.org/10.1520/E1655-05R12.2.

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. **Applied Spectroscopy**, vol. 43, no. 5, p. 772–777, 1989. https://doi.org/10.1366/0003702894202201.

BELLON-MAUREL, V.; FERNANDEZ-AHUMADA, E.; PALAGOS, B.; ROGER, J. M.; MCBRATNEY, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. **TrAC - Trends in Analytical Chemistry**, vol. 29, no. 9, p. 1073–1081, Oct. 2010. https://doi.org/10.1016/j.trac.2010.05.006.

BRAMLEY, R. G. V. Lessons from nearly 20 years of Precision Agriculture research, development, and adoption as a guide to its appropriate application. **Crop and Pasture Science**, vol. 60, no. 3, p. 197–217, 2009. https://doi.org/10.1071/CP08304.

CAMBARDELLA, C. A.; MOORMAN, T. B.; NOVAK, J. M.; PARKIN, T. B.; KARLEN, D. L.; TURCO, R. F.; KONOPKA, A. E. Field-Scale Variability of Soil Properties in Central Iowa Soils. **Soil Science Society of America Journal**, vol. 58, no. 5, p. 1501–1511, 1994. https://doi.org/10.2136/sssaj1994.03615995005800050033x.

CATELAN, M. G.; MARQUES JÚNIOR, J.; SIQUEIRA, D. S.; GOMES, R. P.; BAHIA, A. S. R. de S. Sugarcane yield and quality using soil magnetic susceptibility. **Scientia Agricola**, vol. 79, no. 4, p. 2022, 2022. DOI 10.1590/1678-992x-2020-0329. Available at: http://doi.org/10.1590/1678-992X-2020-0329. Accessed on: 20 Aug. 2021.

CHERUBIN, M. R.; LISBOA, I. P.; SILVA, A. G. B.; VARANDA, L. L.; BORDONAL, R. O.; CARVALHO, J. L. N.; OTTO, R.; PAVINATO, P. S.; SOLTANGHEISI, A.; CERRI, C. E. P. Sugarcane Straw Removal: Implications to Soil Fertility and Fertilizer Demand in Brazil. **Bioenergy Research**, vol. 12, no. 4, p. 888–900, 1 Dec. 2019. DOI 10.1007/s12155-019-10021-w. Available at: https://doi.org/10.1007/s12155-019-10021-w. Accessed on: 2 Jul. 2021.

CONSECANA. National Concill of Sugarcane Producers of São Paulo State. vol. 6, p. 81, 2015. .

CORRÊDO, L. de P.; CANATA, T. F.; MALDANER, L. F.; DE LIMA, J. de J. A.; MOLIN, J. P. Sugarcane Harvester for In-field Data Collection: State of the Art, Its Applicability and Future Perspectives. **Sugar Tech**, vol. 22, no. 4, p. 1–14, 10 Aug. 2020. DOI 10.1007/s12355-020-00874-3. Available at: http://link.springer.com/10.1007/s12355-020-00874-3. Accessed on: 10 Aug. 2020.

CORRÊDO, L. de P.; MALDANER, L. F.; BAZAME, H. C.; MOLIN, J. P. Evaluation of Minimum Preparation Sampling Strategies for Sugarcane Quality Prediction by vis-NIR Spectroscopy. **Sensors**, vol. 21, no. 6, p. 2195, 21 Mar. 2021. DOI 10.3390/s21062195. Available at: https://doi.org/10.3390/s21062195. Accessed on: 21 Mar. 2021.

CORRÊDO, L. P.; WEI, M. C. F.; FERRAZ, M. N.; MOLIN, J. P. Near-infrared spectroscopy as a tool for monitoring the spatial variability of sugarcane quality in the fields. **Biosystems Engineering**, vol. 206, p. 150–161, 1 Jun. 2021. DOI 10.1016/j.biosystemseng.2021.04.001. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1537511021000854. Accessed on: 28 Apr. 2021.

CORTÉS, V.; BLASCO, J.; ALEIXOS, N.; CUBERO, S.; TALENS, P. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. **Trends in Food Science and Technology**, vol. 85, p. 138–148, 1 Mar. 2019. https://doi.org/10.1016/j.tifs.2019.01.015.

COSTA, M. V. A. da; FONTES, C. H.; CARVALHO, G.; JÚNIOR, E. C. de M. UltraBrix: A Device for Measuring the Soluble Solids Content in Sugarcane. **Sustainability**, vol. 13, no. 3, p. 1227, 25 Jan. 2021. DOI 10.3390/su13031227. Available at: https://www.mdpi.com/2071-1050/13/3/1227. Accessed on: 27 Jan. 2021.

DE CASTRO, S. G. Q.; NETO, J. R.; KÖLLN, O. T.; BORGES, B. M. M. N.; FRANCO, H. C. J. Decision-making on the optimum timing for nitrogen fertilization on sugarcane ratoon. **Scientia Agricola**, vol. 76, no. 3, p. 237–242, 2019. DOI 10.1590/1678-992x-2017-0365. Available at: http://dx.doi.org/10.1590/1678-992X-2017-0365. Accessed on: 15 Aug. 2019.

DELEFORTRIE, S.; SAEY, T.; DE PUE, J.; VAN DE VIJVER, E.; DE SMEDT, P.; VAN MEIRVENNE, M. Evaluating corrections for a horizontal offset between sensor and position data for surveys on land. **Precision Agriculture**, vol. 17, no. 3, p. 349–364, 2016. https://doi.org/10.1007/s11119-015-9423-8.

ELAMIN, E. A.; EL-TILIB, M. A.; ELNASIKH, M. H.; IBRAHIM, S. H.; ELSHEIKH, M. A.; BABIKER, E. E. The influence of phosphorus and potassium fertilization on the quality of sugar of two sugarcane varieties grown on three soil series of Sudan. **Journal of Applied Sciences**, vol. 7, no. 16, p. 2345–2350, 15 Aug. 2007. https://doi.org/10.3923/jas.2007.2345.2350.

FEUDALE, R. N.; WOODY, N. A.; TAN, H.; MYLES, A. J.; BROWN, S. D.; FERRÉ, J. Transfer of multivariate calibration models: A review. **Chemometrics and Intelligent Laboratory Systems**, vol. 64, no. 2, p. 181–192, 28 Nov. 2002. https://doi.org/10.1016/S0169-7439(02)00085-0.

FRANCESCHINI, M. H. D.; DEMATTÊ, J. A. M.; KOOISTRA, L.; BARTHOLOMEUS, H.; RIZZO, R.; FONGARO, C. T.; MOLIN, J. P. Effects of external factors on soil reflectance measured on-the-go and assessment of potential spectral correction through orthogonalisation and standardisation procedures. **Soil and Tillage Research**, vol. 177, p. 19–36, 1 Apr. 2018. https://doi.org/10.1016/j.still.2017.10.004.

GOLIC, M.; WALSH, K.; LAWSON, P. Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature. **Applied Spectroscopy**, vol. 57, no. 2, p. 139–145, 2003. DOI 10.1366/000370203321535033. Available at: https://journals.sagepub.com/doi/pdf/10.1366/000370203321535033. Accessed on: 9 Jan. 2019.

GOPALASUNDARAM, P.; BHASKARAN, A.; RAKKIYAPPAN, P. Integrated Nutrient Management in Sugarcane. **Sugar Tech**, vol. 14, no. 1, p. 3–20, 1 Dec. 2012. DOI 10.1007/s12355-011-0097-x. Available at: https://link.springer.com/article/10.1007/s12355-011-0097-x. Accessed on: 2 Jul. 2021.

JI, W.; VISCARRA ROSSEL, R. A.; SHI, Z. Accounting for the effects of water and the environment on proximally sensed vis-NIR soil spectra and their calibrations. **European Journal of Soil Science**, vol. 66, no. 3, p. 555–565, 1 May 2015a. DOI 10.1111/ejss.12239. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/ejss.12239. Accessed on: 1 Jun. 2021.

JI, W.; VISCARRA ROSSEL, R. A.; SHI, Z. Improved estimates of organic carbon using proximally sensed vis-NIR spectra corrected by piecewise direct standardization. **European Journal of Soil Science**, vol. 66, no. 4, p. 670–678, 1 Jul. 2015b. DOI 10.1111/ejss.12271. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/ejss.12271. Accessed on: 2 Jul. 2021.

JOHNSON, R. M.; RICHARD, E. P. Sugarcane yield, sugarcane quality, and soil variability in Louisiana. **Agronomy Journal**, vol. 97, no. 3, p. 760–771, 2005. https://doi.org/10.2134/agronj2004.0184.

KENNARD, R. W.; STONE, L. A. Computer Aided Design of Experiments. **Technometrics**, vol. 11, no. 1, p. 137–148, 1969. https://doi.org/10.1080/00401706.1969.10490666.

KITCHEN, N. R.; SUDDUTH, K. A.; MYERS, D. B.; DRUMMOND, S. T.; HONG, S. Y. Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. **Computers and Electronics in Agriculture**, vol. 46, no. 1-3 SPEC. ISS., p. 285–308, 1 Mar. 2005. https://doi.org/10.1016/j.compag.2004.11.012.

KUANG, B.; TEKIN, Y.; MOUAZEN, A. M. Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content. **Soil and Tillage Research**, vol. 146, no. PB, p. 243–252, 1 Mar. 2015. DOI 10.1016/j.still.2014.11.002. Available at: http://dx.doi.org/10.1016/j.still.2014.11.002. Accessed on: 13 Feb. 2019.

LONG, D. S.; MCCALLUM, J. D. Adapting a relatively low-cost reflectance spectrometer for on-combine sensing of grain protein concentration. **Computers and Electronics in Agriculture**, vol. 174, p. 105467, 1 Jul. 2020. https://doi.org/10.1016/j.compag.2020.105467.

MAGALHÃES, P. S. G. G.; CERRI, D. G. P. P. Yield Monitoring of Sugar Cane. **Biosystems Engineering**, vol. 96, no. 1, p. 1–6, Jan. 2007. DOI 10.1016/j.biosystemseng.2006.10.002.

MALDANER, L. F.; CORRÊDO, L. de P.; CANATA, T. F.; MOLIN, J. P. Predicting the sugarcane yield in real-time by harvester engine parameters and machine learning approaches. **Computers and Electronics in Agriculture**, vol. 181, p. 105945, 1 Feb. 2021. https://doi.org/10.1016/J.COMPAG.2020.105945.

MALDANER, L. F.; MOLIN, J. P. Data processing within rows for sugarcane yield mapping. **Scientia Agricola**, vol. 77, no. 5, p. 2020, 2020. DOI 10.1590/1678-992x-2018-0391. Available at: http://dx.doi.org/10.1590/1678-992X-2018-0391. Accessed on: 24 Dec. 2019.

MARAPHUM, K.; CHUAN-UDOM, S.; SAENGPRACHATANARUG, K.; WONGPICHET, S.; POSOM, J.; PHUPHAPHUD, A.; TAIRA, E. 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**, vol. 26, no. 5, p. 287–296, 2018. DOI 10.1177/0967033518795810. Available at: https://journals.sagepub.com/doi/pdf/10.1177/0967033518795810. Accessed on: 11 Mar. 2019.

MARQUES, J.; SIQUEIRA, D. S.; CAMARGO, L. A.; TEIXEIRA, D. D. B.; BARRÓN, V.; TORRENT, J. Magnetic susceptibility and diffuse reflectance spectroscopy to characterize the spatial variability of soil properties in a brazilian haplustalf. **Geoderma**, vol. 219–220, p. 63–71, 1 May 2014. DOI 10.1016/j.geoderma.2013.12.007. Available at: http://dx.doi.org/10.1016/j.geoderma.2013.12.007. Accessed on: 27 Dec. 2018.

MAYRINK, G. O.; VALENTE, D. S. M.; QUEIROZ, D. M.; PINTO, F. A. C.; TEOFILO, R. F. Determination of chemical soil properties using diffuse reflectance and ion-exchange resins. **Precision Agriculture**, vol. 20, p. 541–561, 2019. DOI 10.1007/s11119-018-9597-y. Available at: https://doi.org/10.1007/s11119-018-9597-y. Accessed on: 22 May 2019.

MEYER, J. H.; WOOD, R. A. The effects of soil fertility and nutrition on sugarcane quality: a review. **Proc S Afr Sug Technol Ass**, vol. 75, no. March, p. 242–247, 2001. .

MOLIN, J. P.; BAZAME, H. C.; MALDANER, L.; CORREDO, L. D. P.; MARTELLO, M.; CANATA, T. F. Precision agriculture and the digital contributions for site-specific management of the fields. **Revista Ciência Agronômica**, vol. 51, no. Special Agriculture 4.0, p. 1–10, 2020. DOI 10.5935/1806-6690.20200088. Available at: http://ccarevista.ufc.br/seer/index.php/ccarevista/article/view/7720. Accessed on: 5 Aug. 2021.

MORELLOS, A.; PANTAZI, X. E.; MOSHOU, D.; ALEXANDRIDIS, T.; WHETTON, R.; TZIOTZIOS, G.; WIEBENSOHN, J.; BILL, R.; MOUAZEN, A. M. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. **Biosystems Engineering**, vol. 152, p. 104–116, 2016. DOI 10.1016/j.biosystemseng.2016.04.018. Available at: https://ac.els-cdn.com/S1537511015304165/1-s2.0-S1537511015304165-main.pdf?_tid=9bd2a49e-74fd-4d85-879e-0462d1e3e1e5&acdnat=1546893219_bfe1381c3b989b935bfd4e7d9069837f. Accessed on: 7 Jan. 2019.

MUCHOW, R. C.; ROBERTSON, M. J.; WOOD, A. W.; KEATING, B. A. Effect of nitrogen on the time-course of sucrose accumulation in sugarcane. **Field Crops Research**, vol. 47, no. 2–3, p. 143–153, Aug. 1996. DOI 10.1016/0378-4290(96)00022-6. Available at: https://linkinghub.elsevier.com/retrieve/pii/0378429096000226. Accessed on: 15 Aug. 2019.

MUNNAF, M. A.; MOUAZEN, A. M. Development of a soil fertility index using on-line Vis-NIR spectroscopy. **Computers and Electronics in Agriculture**, vol. 188, p. 106341, 1 Sep. 2021. DOI 10.1016/j.compag.2021.106341. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0168169921003586. Accessed on: 26 Jul. 2021.

NAWAR, S.; MUNNAF, M. A.; MOUAZEN, A. M. Machine learning based on-line prediction of soil organic carbon after removal of soil moisture effect. **Remote Sensing**, vol. 12, no. 8, p. 1308, 1 Apr. 2020. DOI 10.3390/RS12081308. Available at: www.mdpi.com/journal/remotesensing. Accessed on: 1 Jul. 2020.

NAWI, N. M.; CHEN, G.; JENSEN, T. In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: a review. **Precision Agriculture**, vol. 15, no. 6, p. 684–703, 2014. https://doi.org/10.1007/s11119-014-9362-9.

NAWI, N. M.; CHEN, G.; JENSEN, T. Visible and shortwave near infrared spectroscopy for predicting sugar content of sugarcane based on a cross-sectional scanning method. **Journal of Near Infrared Spectroscopy**, vol. 21, no. 4, p. 289–297, 2013. https://doi.org/10.1255/jnirs.1060.

NAWI, N. M.; CHEN, G.; JENSEN, T.; MEHDIZADEH, S. A. Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared. **Biosystems Engineering**, vol. 115, p. 154–161, 2013. DOI 10.1016/j.biosystemseng.2013.03.005. Available at: http://dx.doi.org/10.1016/j.biosystemseng.2013.03.005.

OLIVERI, P.; MALEGORI, C.; SIMONETTI, R.; CASALE, M. The impact of signal pre-processing on the final interpretation of analytical outcomes – A tutorial. **Analytica Chimica Acta**, vol. 1058, p. 9–17, 13 Jun. 2019. https://doi.org/10.1016/j.aca.2018.10.055.

OSBORNE, B. G. Near-Infrared Spectroscopy in Food Analysis. **Encyclopedia of Analytical Chemistry**, Chichester, UK, , p. 14, 30 Oct. 2006. DOI 10.1002/9780470027318.a1018. Available at: http://doi.wiley.com/10.1002/9780470027318.a1018. Accessed on: 11 Dec. 2018.

PANCHUK, V.; KIRSANOV, D.; OLENEVA, E.; SEMENOV, V.; LEGIN, A. Calibration transfer between different analytical methods. **Talanta**, vol. 170, p. 457–463, 1 Aug. 2017. https://doi.org/10.1016/j.talanta.2017.04.039.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, vol. 1026, p. 8–36, 2018. https://doi.org/10.1016/j.aca.2018.04.004.

PHETPAN, K.; UDOMPETAIKUL, V.; SIRISOMBOON, P. An online visible and near-infrared spectroscopic technique for the real-time evaluation of the soluble solids content of sugarcane billets on an elevator conveyor. **Computers and Electronics in Agriculture**, vol. 154, p. 460–466, Nov. 2018. DOI 10.1016/j.compag.2018.09.033. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0168169918302096. Accessed on: 7 Jan. 2019.

PHUPHAPHUD, A.; SAENGPRACHATANARUG, K.; POSOM, J.; MARAPHUM, K.; TAIRA, E. Prediction of the fibre content of sugarcane stalk by direct scanning using visible-shortwave near infrared spectroscopy. **Vibrational Spectroscopy**, vol. 101, p. 71–80, Mar. 2019. DOI 10.1016/j.vibspec.2019.02.005. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0924203118302613. Accessed on: 3 Jun. 2019.

PISSARD, A.; MARQUES, E. J. N.; DARDENNE, P.; LATEUR, M.; PASQUINI, C.; PIMENTEL, M. F.; FERNÁNDEZ PIERNA, J. A.; BAETEN, V. Evaluation of a handheld ultra-compact NIR spectrometer for rapid and non-destructive determination of apple fruit quality. **Postharvest Biology and Technology**, vol. 172, p. 111375, 1 Feb. 2021. https://doi.org/10.1016/j.postharvbio.2020.111375.

POMARES-VICIANA, T.; MARTÍNEZ-VALDIVIESO, D.; FONT, R.; GÓMEZ, P.; DEL RÍO-CELESTINO, M. Characterisation and prediction of carbohydrate content in zucchini fruit using near infrared spectroscopy. **Journal of the Science of Food and Agriculture**, vol. 98, no. 5, p. 1703–1711, 30 Mar. 2018. https://doi.org/10.1002/jsfa.8642.

POREP, J. U.; KAMMERER, D. R.; CARLE, R. On-line application of near infrared (NIR) spectroscopy in food production. **Trends in Food Science and Technology**, vol. 46, no. 2, p. 211–230, 1 Dec. 2015. https://doi.org/10.1016/j.tifs.2015.10.002.

RISIUS, H.; HAHN, J.; HUTH, M.; TÖLLE, R.; KORTE, H. In-line estimation of falling number using near-infrared diffuse reflectance spectroscopy on a combine harvester. **Precision Agriculture**, vol. 16, no. 3, p. 261–274, 1 Jun. 2015. DOI 10.1007/s11119-014-9374-5. Available at: https://link.springer.com/article/10.1007/s11119-014-9374-5. Accessed on: 26 Mar. 2021.

RODRIGUES, F. A.; MAGALHÃES, P. S. G.; FRANCO, H. C. J. Soil attributes and leaf nitrogen estimating sugar cane quality parameters: Brix, pol and fibre. **Precision Agriculture**, vol. 14, no. 3, p. 270–289, 2013. https://doi.org/10.1007/s11119-012-9294-1.

RUANGRATANAKORN, J.; SUWONSICHON, T.; KASEMSUMRAN, S.; THANAPASE, W. Installation design of on-line near infrared spectroscopy for the production of compound fertilizer. **Vibrational Spectroscopy**, vol. 106, p. 103008, 1 Jan. 2020. https://doi.org/10.1016/j.vibspec.2019.103008.

SALGUERO-CHAPARRO, L.; PEÑA-RODRÍGUEZ, F. On-line versus off-line NIRS analysis of intact olives. **LWT - Food Science and Technology**, vol. 56, no. 2, p. 363–369, 1 May 2014. https://doi.org/10.1016/j.lwt.2013.11.032.

SANCHES, G. M.; GRAZIANO MAGALHÃES, P. S.; JUNQUEIRA FRANCO, H. C. Site-specific assessment of spatial and temporal variability of sugarcane yield related to soil attributes. **Geoderma**, vol. 334, p. 90–98, 15 Jan. 2019. DOI 10.1016/j.geoderma.2018.07.051. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0016706118303513. Accessed on: 27 Dec. 2018.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, vol. 36, no. 8, p. 1627–1639, 1964. DOI 10.1021/ac60214a047. Available at: https://pubs.acs.org/sharingguidelines. Accessed on: 28 Mar. 2020.

SEXTON, J.; EVERINGHAM, Y.; DONALD, D.; STAUNTON, S.; WHITE, R. Investigating the identification of atypical sugarcane using NIR analysis of online mill data. **Computers and Electronics in Agriculture**, vol. 168, 1 Jan. 2020. DOI 10.1016/j.compag.2019.105111. Available at: https://doi.org/10.1016/j.compag.2019.105111. Accessed on: 3 Jan. 2021.

SILVA, C. B.; DE MORAES, M. A. F. D.; MOLIN, J. P. Adoption and use of precision agriculture technologies in the sugarcane industry of São Paulo state, Brazil. **Precision Agriculture**, vol. 12, no. 1, p. 67–81, 6 Jan. 2011. DOI 10.1007/s11119-009-9155-8. Available at: https://link.springer.com/article/10.1007/s11119-009-9155-8. Accessed on: 11 Aug. 2021.

TÜMSAVAŞ, Z.; TEKIN, Y.; ULUSOY, Y.; MOUAZEN, A. M. Prediction and mapping of soil clay and sand contents using visible and near-infrared spectroscopy. **Biosystems Engineering**, vol. 177, p. 90–100, 1 Jan. 2019. https://doi.org/10.1016/j.biosystemseng.2018.06.008.

UDOMPETAIKUL, V.; PHETPAN, K.; SIRISOMBOON, P. Development of the partial least-squares model to determine the soluble solids content of sugarcane billets on an elevator conveyor. **Measurement**, vol. 167, p. 107898, 1 Jan. 2021. DOI 10.1016/j.measurement.2020.107898. Available at: https://linkinghub.elsevier.com/retrieve/pii/S026322412030436X. Accessed on: 17 Feb. 2021.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy. **Journal of Agricultural and Food Chemistry**, vol. 55, no. 21, p. 8331–8338, 2007. DOI 10.1021/jf071538s. Available at: https://pubs.acs.org/doi/pdf/10.1021/jf071538s. Accessed on: 3 Jul. 2018.

WALSH, K. B.; BLASCO, J.; ZUDE-SASSE, M.; SUN, X. Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. **Postharvest Biology and Technology**, vol. 168, p. 111246, 1 Oct. 2020. https://doi.org/10.1016/j.postharvbio.2020.111246.

WANG, H.; PENG, J.; XIE, C.; BAO, Y.; HE, Y. Fruit quality evaluation using spectroscopy technology: A review. **Sensors**, vol. 15, no. 5, p. 11889–11927, 21 May 2015. DOI 10.3390/s150511889. Available at: http://www.mdpi.com/1424-8220/15/5/11889. Accessed on: 12 Aug. 2019.

WANG, Y.; VELTKAMP, D. J.; KOWALSKI, B. R. Multivariate Instrument Standardization. **Analytical Chemistry**, vol. 63, no. 23, p. 2750–2756, 1 Dec. 1991. DOI 10.1021/ac00023a016. Available at: https://pubs.acs.org/doi/abs/10.1021/ac00023a016. Accessed on: 2 Jul. 2021.

WOLD, S.; MARTENS, H.; WOLD, H. The multivariate calibration problem in chemistry solved by the PLS method. [*S. l.*]: Springer, Berlin, Heidelberg, 1983. p. 286–293. DOI 10.1007/bfb0062108. Available at: https://link.springer.com/chapter/10.1007/BFb0062108. Accessed on: 30 Dec. 2020.

WOLD, Svante; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: A basic tool of chemometrics. 58., 2001. **Chemometrics and Intelligent Laboratory Systems** [...]. [*S. l.*: *s. n.*], 2001. vol. 58, p. 109–130. DOI 10.1016/S0169-7439(01)00155-1. Available at: www.elsevier.comrlocaterchemometrics. Accessed on: 4 Jan. 2019.

WORKMAN, J. J. A Review of Calibration Transfer Practices and Instrument Differences in Spectroscopy. **Applied Spectroscopy**, vol. 72, no. 3, p. 340–365, 1 Mar. 2018. DOI 10.1177/0003702817736064. Available at: http://journals.sagepub.com/doi/10.1177/0003702817736064. Accessed on: 24 Jun. 2020.

WORKMAN JR., J.; WEYER, L. **Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy**. [*S. l.*: *s. n.*], 2012.

XU, X.; MO, J.; XIE, L.; YING, Y. Influences of Detection Position and Double Detection Regions on Determining Soluble Solids Content (SSC) for Apples Using On-line Visible/Near-Infrared (Vis/NIR) Spectroscopy. **Food Analytical Methods**, vol. 12, no. 9, p. 2078–2085, 15 Sep. 2019. DOI 10.1007/s12161-019-01530-7. Available at: https://doi.org/10.1007/s12161-019-01530-7. Accessed on: 15 Feb. 2021.

YANG, S.; WU, Q.; CHANG, H.; LI, Q.; XU, H. Effect of grain density to near infrared spectra and design of a laboratory evaluation system for combine harvester. **2017 ASABE Annual International Meeting**, , p. 1–8, 2017. https://doi.org/10.13031/aim.201700493.

5. FINAL REMARKS

The knowledge of the spatial variability of agricultural products quality may provide valuable information and enhance yield data in crop management, given the importance of these attributes for the market and due they are the result of interventions made throughout the crop cycle. For sugarcane, it will be the watershed for effective implementation of PA techniques in the management of sugarcane plantations, by providing quality data, such as total recoverable sugar, by area. On-board proximal sensing techniques in harvesters are the most promising solution to this aim. Sensors based on near-infrared spectroscopy have been widely used for quality assessment of agricultural products in industry and evaluated on board harvesters and combines. The present work evaluated this technology for the identification of spatial variability of sugarcane quality, different levels of processing of samples collected throughout a harvest, and an on-board sensor in a sugarcane harvester for real-time quality measurements. The system was able to provide sugarcane quality data with minimal sample preparation required in the calibration procedure. Moreover, it allowed representing the spatial variability of these properties in thematic maps for useful for decision making regarding targeted diagnostics, prescriptions, and site-specific interventions.

The spatial variability of quality parameters was characterized by means of prediction models developed from NIR spectroscopy and the results were confronted with those obtained by conventional analyses. In addition, the evaluation of different sample forms allowed the obtainment of models based on minimum preparation. However, further progress can be made in alternative methods so that no preparation is necessary. Despite this, the minimal preparation in the calibration process does not present itself as a limiting factor, because from some samples it was possible to develop models to fit the data collected in real time. Moreover, the evaluation of cause-and-effect relationships proved that the spatialized quality data were consistent with previous studies relating edaphic attributes and sugarcane quality.

The research showed promising and unprecedented results regarding the evaluation of an on-board NIR spectroscopy to estimate sugarcane quality data during the harvest. This may be the first on-board sensor for this purpose. Also, it was proposed ways to improve the sensing technique to large-scale application, mainly along different harvest seasons.

APPENDIX

APPENDIX A. Sugarcane varieties used to perform the study; Principal Component Analysis (PCA) for the varieties, from the spectra of the different sample forms; Standardized residuals generated from partial least squares (PLS) regression analysis; and loadings for each quality attribute predicted from different types of sugarcane samples.

Table 13. Number of samples of each Brazilian sugarcane variety used in the study.

Variety	# Samples	Variety	# Samples	Variety	# Samples
CT96-1007	12	CTC 9001	1	RB966928	7
CT96-3346	7	CTC 9005	2	RB975201	1
CTC 11	19	CV 6654	3	RB975952	2
CTC 14	10	IACSP95-5000	1	RB985476	1
CTC 15	12	RB855002	3	SP80-3280	9
CTC 17	7	RB855156	33	SP83-2847	31
CTC 2	19	RB855536	4	SP83-5073	2
CTC 20	26	RB867515	9	RB965621	1
CTC 22	2	RB935621	4	SP91-1049	1
CTC 4	34	RB965621	1	Various	26
CTC 7	1	RB965902	12		

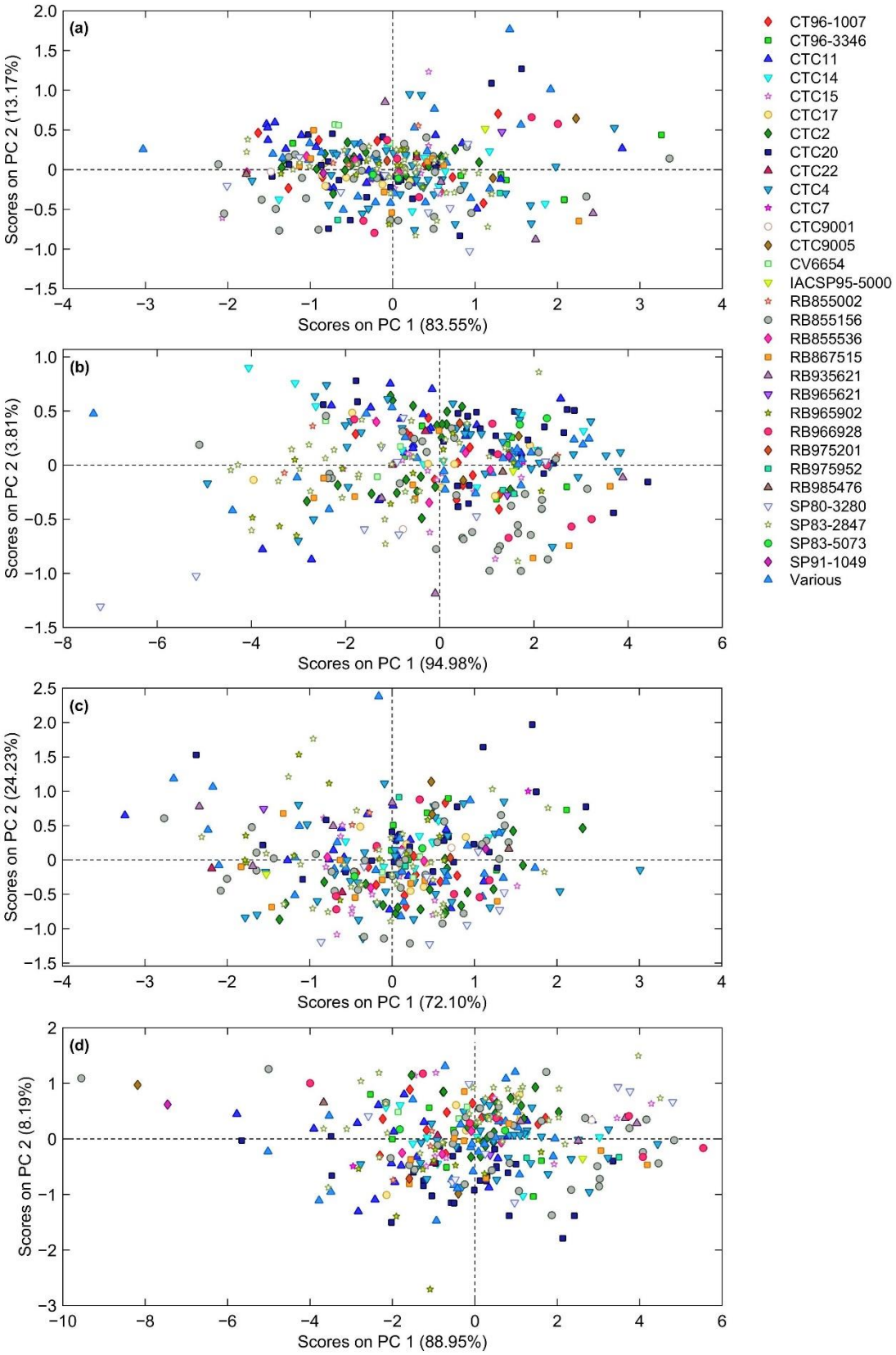

Figure 30. Principal component scores of different sugarcane varieties for spectral data of skin scanning of sugarcane billets (a), cross-sectional scanning of sugarcane billets (b), defibrated cane samples (c), and raw juice samples (d).

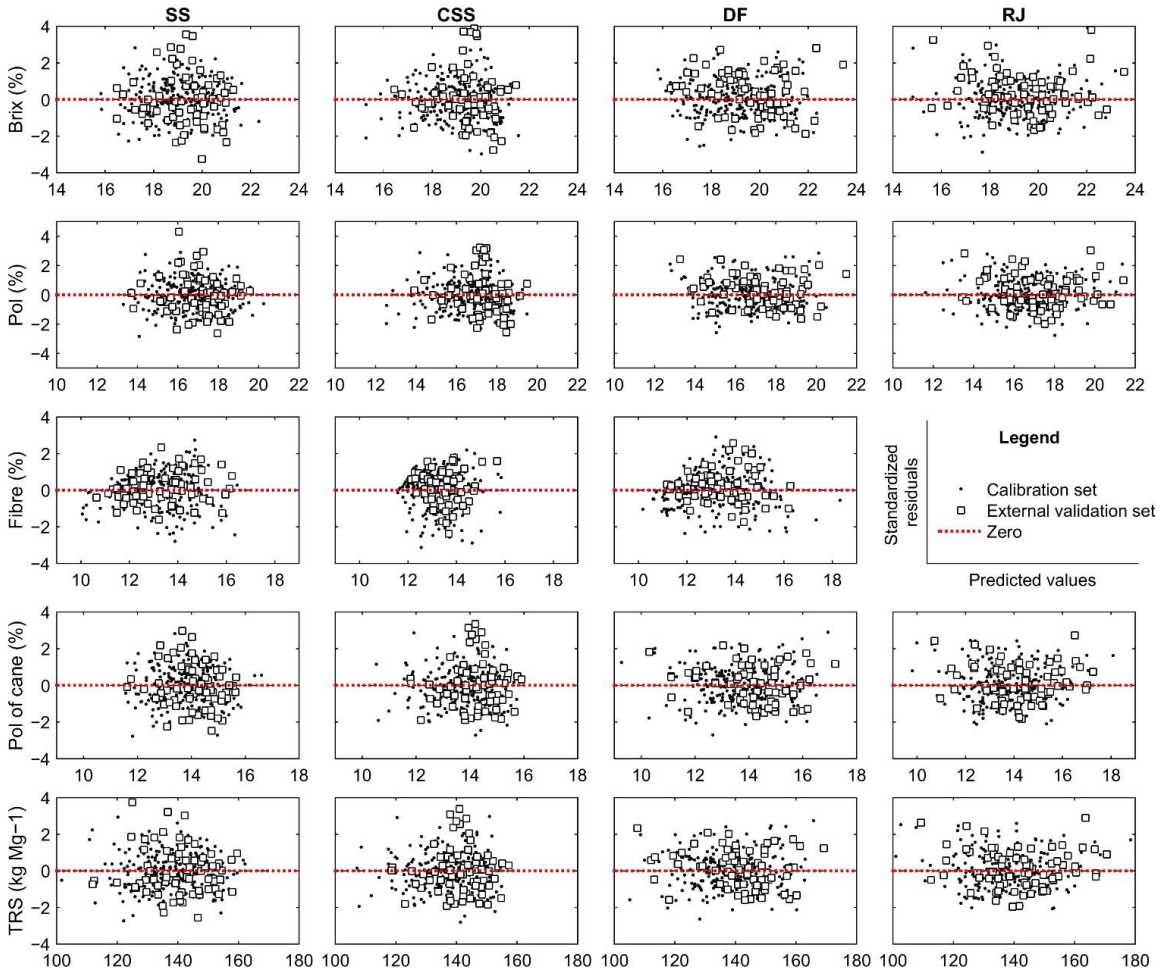

Figure 31. The plot of standardized residuals versus predicted sugarcane quality values from vis-NIR by skin (SS) and cross-sectional (CSS) scanning of billets, defibrated (DF), and raw juice (RJ) samples. Black dots are calibration samples, and red dots are external validation samples.

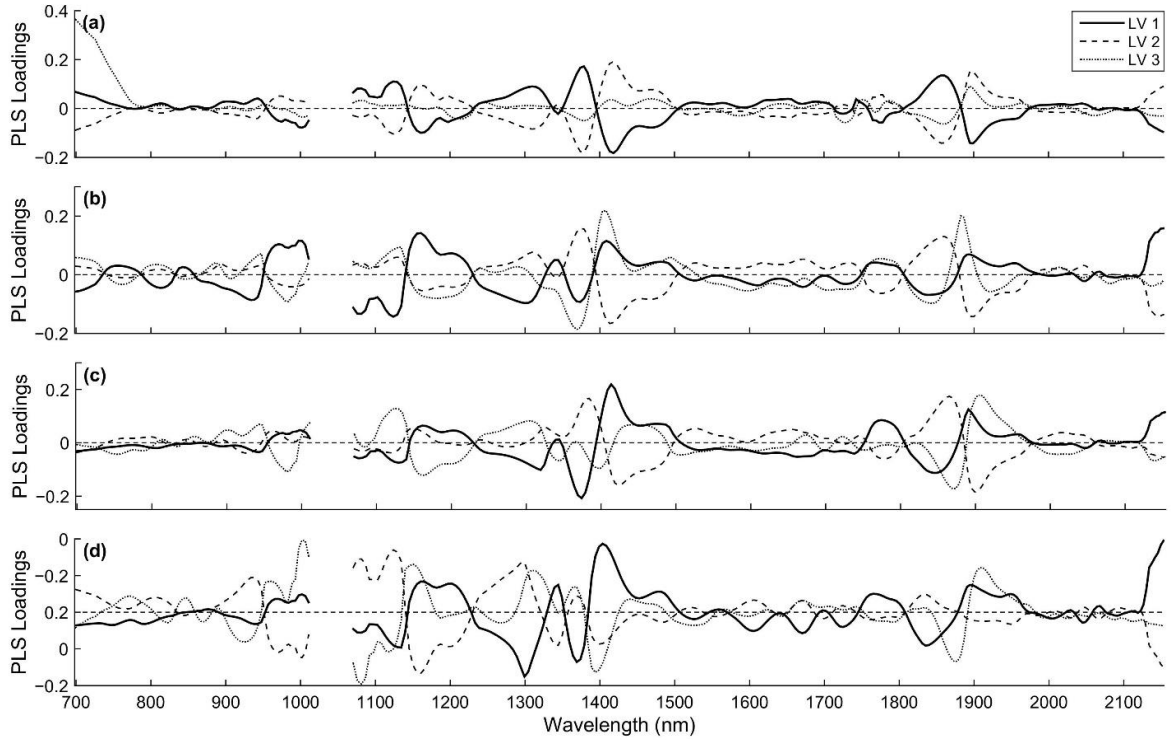

Figure 32. The three first Partial least squares loadings for Brix prediction using near-infrared reflectance spectroscopy from the skin (a) and cross-sectional (b) scanning of billets, defibrated cane (c), and extracted raw juice (d).

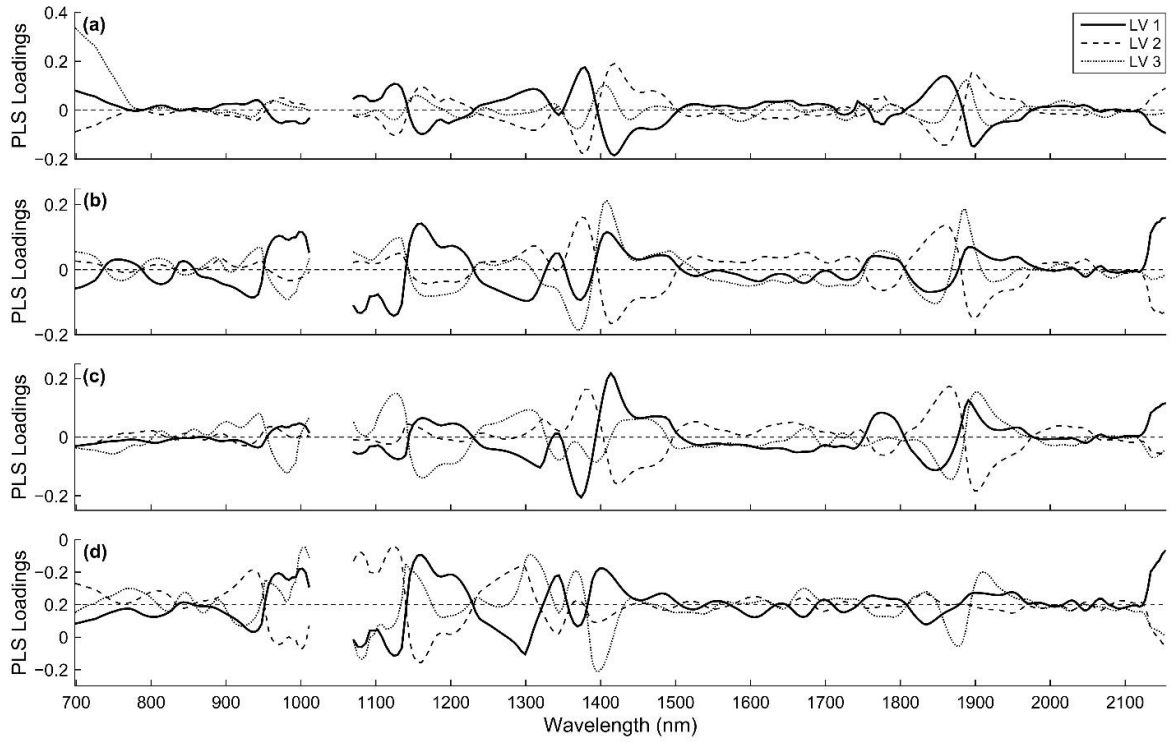

Figure 33. The three first Partial least squares loadings for Pol prediction using near-infrared reflectance spectroscopy from the skin (a) and cross-sectional (b) scanning of billets, defibrated cane (c), and extracted raw juice (d).

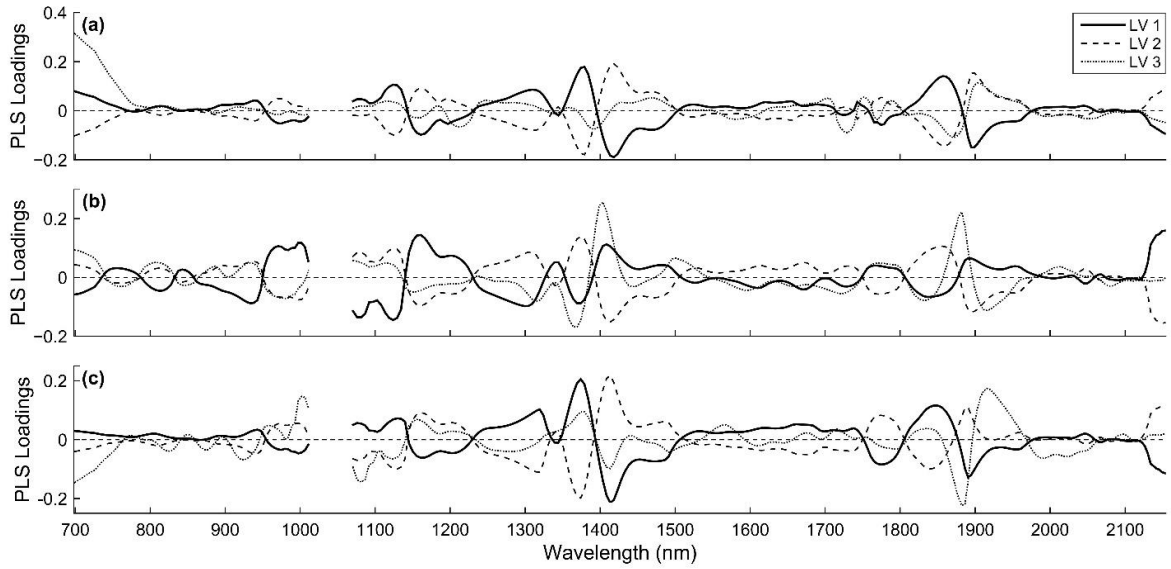

Figure 34. The three first Partial least squares loadings for Pol of cane prediction using near-infrared reflectance spectroscopy from the skin (a) and cross-sectional (b) scanning of billets, defibrated cane (c), and extracted raw juice (d).

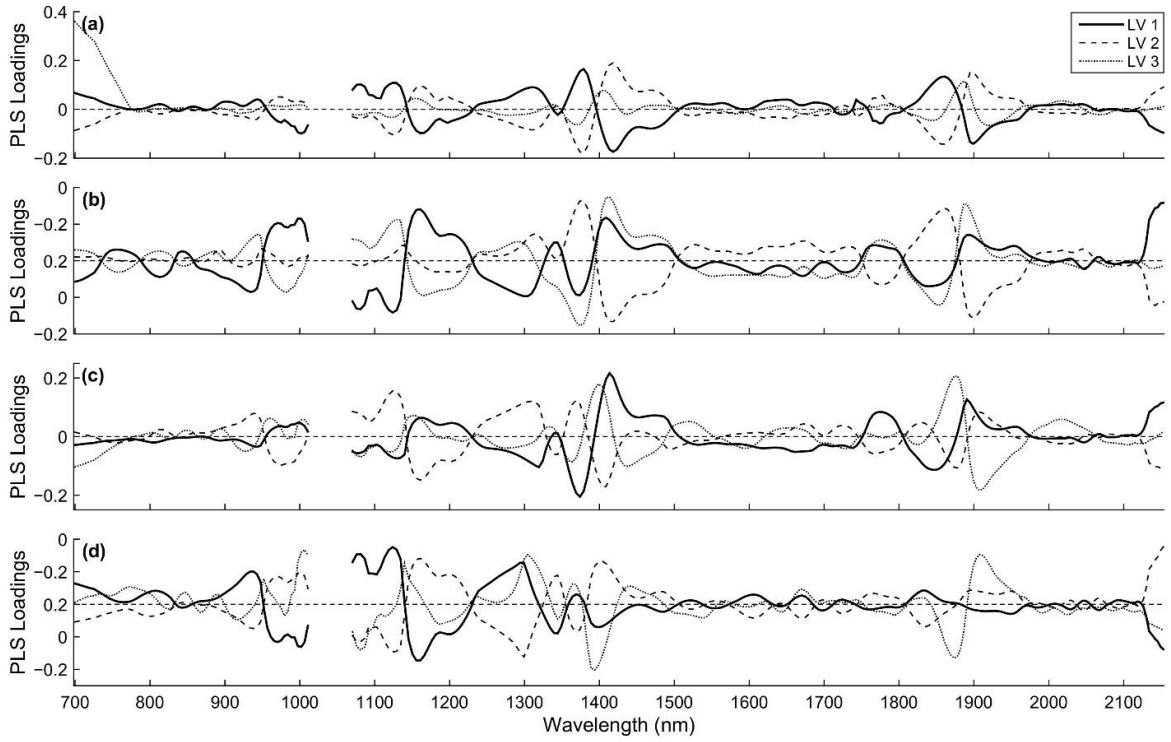

Figure 35. The three first Partial least squares loadings for Pol of cane prediction using near-infrared reflectance spectroscopy from the skin (a) and cross-sectional (b) scanning of billets, defibrated cane (c), and extracted raw juice (d).

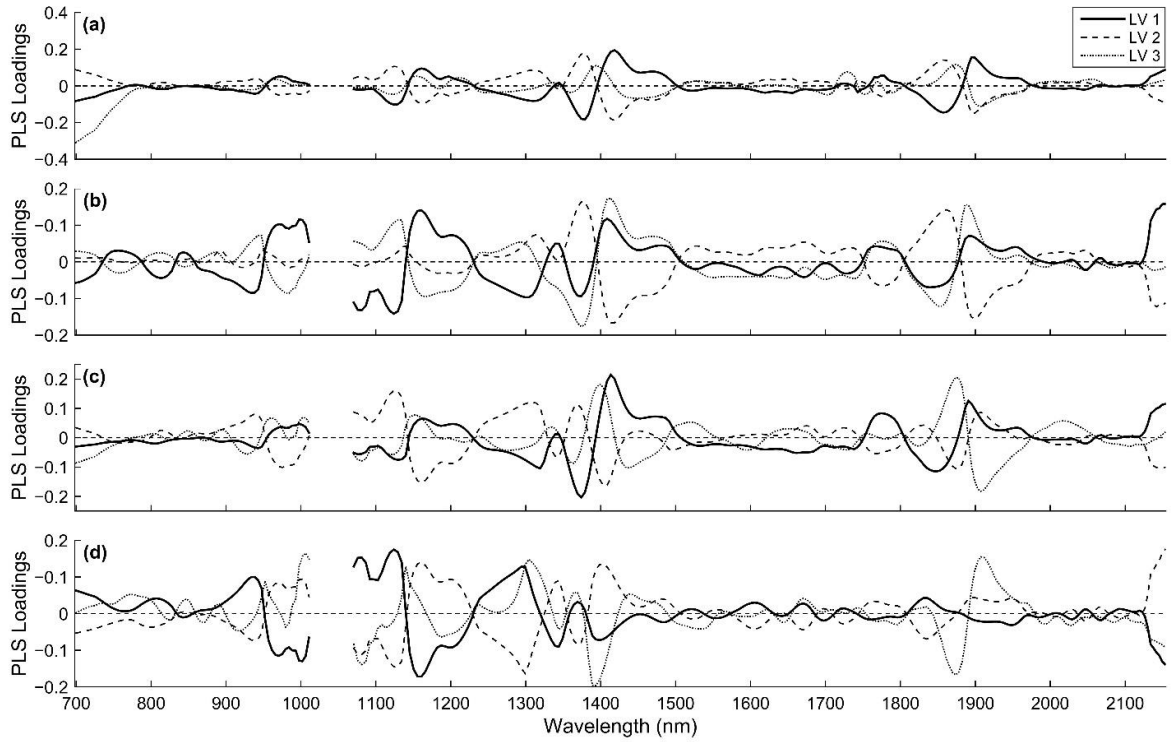

Figure 36. The three first Partial least squares loadings for total recoverable sugars (TRS) prediction using near-infrared reflectance spectroscopy from the skin (a) and cross-sectional (b) scanning of billets, defibrated cane (c), and extracted raw juice (d).