



Contents lists available at ScienceDirect

## Food Research International

journal homepage: [www.elsevier.com/locate/foodres](http://www.elsevier.com/locate/foodres)

# NIR spectroscopy and chemometrics-based traceability of specialty Brazilian green canephora coffee<sup>☆</sup>

Venancio Ferreira de Moraes-Neto<sup>a</sup>, Michel Rocha Baqueta<sup>a</sup>, Alexsandro Lara Teixeira<sup>b</sup>,  
Elem Tamirys dos Santos Caramês<sup>a,c</sup>, Juliana Azevedo Lima Pallone<sup>a,\*</sup>

<sup>a</sup> University of Campinas (UNICAMP), School of Food Engineering, Department of Food Science and Nutrition, Campinas, São Paulo, Brazil

<sup>b</sup> Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA Rondônia, Porto Velho, Rondônia, Brazil

<sup>c</sup> University of São Paulo – USP, Institute of Biomedical Sciences, Department of Microbiology, São Paulo, Brazil

## ARTICLE INFO

## Keywords:

Vibrational spectroscopy  
Discriminant analysis  
Coffee varieties  
Geographical origin  
Green coffee beans

## ABSTRACT

There is a need to develop efficient analytical methods to recognize the origins of coffee beans, especially from large producers such as Brazil, which offers high value-added Geographical Indication (GI) coffees. However, the challenge is not only the need for fast and clean techniques but also understanding how sample preparation and data treatments directly affect the performance of the applied technique. In this study, Near Infrared (NIR) Spectroscopy was combined with Partial Least Squares Discriminant Analysis (PLS-DA) to assess the ability to discriminate green coffee samples with recognized GI (Robusta Amazônico from Rondônia and Conilon from Espírito Santo), examining the influence of sample presentation (ground or whole bean) and spectral pre-processing. The results demonstrated that NIR performed with high efficiency for both whole beans and ground green coffee, achieving 100 % correct prediction. The most effective pre-processing was the combination of the 1st derivative of Savitzky-Golay and Multiplicative Scatter Correction (MSC). This suggests that the technique can be used for rapid discrimination in green coffee trading, with the direct analysis of natural whole beans being much more advantageous, as it avoids milling, which requires liquid nitrogen and a specific mill. Thus, NIR coupled with PLS-DA is a non-invasive, easy-to-operate, low-cost, and sensitive technique that can be applied directly to intact canephora coffee samples.

## 1. Introduction

Coffee is among the most popular drinks consumed worldwide, and its consumption is increasing due to recent improvements in product quality, better agronomic practices, health benefits when consumed in moderation, and the availability of new products (Ludwig et al., 2014; Vegro & Almeida, 2019). Among the various species of coffee, the arabica coffee (*Coffea arabica* L.) stands out economically for its aroma and flavor. However, recently, canephora coffee (*Coffea canephora* Pierre ex A. Froehner) has been managed by some producers aiming to obtain specialty coffees (Corrêa et al., 2016; Farah, 2018).

In the 2022–2023 harvest, coffee production in Brazil will exceed 67 million bags, of which 70 % corresponds to arabica coffee and 30 % to canephora coffee. The production of canephora coffee has been growing significantly and the beverage obtained from the species has been increasingly included in world consumption (CONAB, 2024). In Brazil, the canephora coffee is divided into two large groups, the “Conilon” and

the “Robusta”, being mainly cultivated in the states of Espírito Santo and Rondônia respectively, as specialty coffees. In these regions, the plantations are conducted predominantly in small-scale family modules, with family labor (Baqueta et al., 2024). In 2021, the National Institute of Industrial Property (INPI) granted indication of origin “Espírito Santo” for Conilon coffee and then the denomination of geographical origin “Matas de Rondônia” for Robusta Amazônico (de Moraes-Neto et al., 2024; INPI, 2021). Rondônia has unique characteristics in coffee production. In addition to benefiting from the Amazon biome, which influences the sensory attributes of the beans, the state stands out for cultivating Amazonian robusta by both traditional coffee growers and indigenous communities. This diverse production not only enhances the territorial identity of the coffee but also promotes sustainable practices and the preservation of traditional knowledge in coffee farming (Baqueta et al., 2024).

The origin of coffee beans, whether geographical or botanical, has a direct relation with the quality of the beverage, especially with regard to

<sup>☆</sup> This article is part of a Special issue entitled: ‘Miniaturized Spectroscopy’ published in Food Research International.

\* Corresponding author.

<https://doi.org/10.1016/j.foodres.2025.117093>

Received 3 December 2024; Received in revised form 25 June 2025; Accepted 26 July 2025

Available online 5 August 2025

0963-9969/© 2025 Elsevier Ltd. All rights reserved, including those for text and data mining, AI training, and similar technologies.

sensory properties, as well as the added economic value as in the case of specialty canephora coffees (Hoyos Ossa et al., 2018). This factor can cause fraud in labeling, that is, incorrect geographical or botanical origin attribution of coffee beans, marketing a cheaper product as higher-priced varieties. Consequently, both coffee producers and industries need to seek ways to protect market reputation from fraudulent issues. In this context, there is encouragement for the development of efficient analytical methods capable of assessing coffee authenticity, enabling fair trade throughout the coffee chain (Malta et al., 2020; Spink, 2019).

In the last decades, researches have tested analytical methods to correctly discriminate the geographical origin of green coffee through information about the chemical composition, biochemical aspects or fingerprinting of the samples by chromatographic techniques (De Luca et al., 2016; Mehari et al., 2019); spectroanalytical analysis, with emphasis on infrared spectroscopy and absorption and atomic emission spectrometry (Botelho et al., 2017; Endaye et al., 2020; Link et al., 2014; Marquetti et al., 2016); and data fusions of different analytical techniques including conventional techniques used for obtaining the centesimal composition, chromatography and mass spectrometry (Babova et al., 2016; Worku et al., 2019; Zhu et al., 2021). Although many techniques are considered promising, most of the analytical strategies adopted have made use of methods that require the use of expensive, laborious, and time-consuming techniques and reagents. Some require the use of potentially toxic reagents and are not feasible for large numbers of samples (Abernethy & Higgs, 2013).

In this scenario, the use of near infrared spectroscopy (NIR) raises as a promising alternative analytical tool, once it presents as a result a vibrational spectrum that depends on the chemical composition of the samples (Manley & Baeten, 2018). NIR has characteristics that fit the principles of green analytical chemistry, as it allows the development of analytical methods that do not require extensive sample pretreatment, are non-destructive, economical, and their applications go beyond the laboratory scale using data science applied to chemical data, as chemometric tools (Gatuszka et al., 2013). Nowadays, this has become a modern and important tool in process engineering through the possibility of real-time and on-line monitoring, quality control and process automation in all processing scales (Pallone et al., 2018). Chemometrics is the use of statistical, mathematical and graphical tools to extract information from data from chemical systems (Otto, 2023). Its application can be directed to several purposes, such as: exploratory data analysis by Principal Component Analysis (PCA) and supervised discrimination by Partial Least Squares with Discriminant Analysis (PLS-DA) (Ferreira, 2015; Oliveri, 2017). Chemometric methods, whether supervised or unsupervised, are widely reported in the literature as suitable and promising tools for discriminating between authentic and inauthentic samples (Caballero et al., 2021; Efenberger-Szmechtyk et al., 2018).

For coffees, NIR has already been applied to discriminate samples of different geographical origins (Bona et al., 2017; Giraudo et al., 2019; Marquetti et al., 2016), arabica and canephora blends (Mees et al., 2018; Myles et al., 2006) and coffee defects (Santos et al., 2012), obtaining satisfactory results. However, there are still few studies using it to discriminate the geographic/botanical origin of raw coffee beans. Giraudo et al. (2019), Marquetti et al. (2016), and Bona et al. (2017) verified the efficiency of NIR for discrimination of Arabica coffees. For Brazilian green canephora coffees, only one study using UV-Vis spectrophotometry was found (de Moraes-Neto et al., 2024), but infrared spectroscopy presents itself as a possible application. It is worth noting that Brazilian coffee is predominantly exported and sold as unripe beans, which highlights not only the need to develop effective methods to guarantee the traceability of this product, but also to understand how the conditions under which the methods are carried out can affect performance efficiency.

In this study, NIR was associated PLS-DA to (i) verify the discrimination capacity of green coffee canephora samples with recognized geographical and botanical origins (Robusta Amazônico from Rondônia

and Conilon from Espírito Santo), (ii) and the influence of the impact of intact beans and ground green coffees on NIR performance, as an unprecedented application to control of coffees with GI.

## 2. Materials and method

### 2.1. Samples

In this study 222 samples of green (raw) beans of canephora coffee supplied by the Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA-Rondônia) were analyzed, of which 114 were from the Robusta Amazônico, coded as RA and 108 samples from the Conilon from the state Espírito Santo (Conilon from Espírito Santo), coded as CES. It is important to note that each sample is from a different producer, with a total of 222 distinct producers involved. The beans were not dried, and their natural moisture content was preserved.

The NIR spectra data were collected for the intact beans (without pulp and parchment) and for the ground beans for comparison purposes. The ground samples were prepared by freezing them using liquid nitrogen and then grinding them with a home coffee grinder (Cadence, model MR35). In this study, unroasted coffee beans were used, in which the silver skin naturally remains adhered to the seeds, regardless of the postharvest processing method. Consequently, in all the analyzed samples, whether whole or ground, the NIR spectra were acquired with the silver skin present. The granulometry of the ground samples was standardized for all experiments using a 20 mesh/tyler - 0.85 mm sieve. All the samples were stored in laminated stand-up pouch packaging at  $-18^{\circ}\text{C}$  during the experiments.

### 2.2. NIR data acquisition

NIR spectra were obtained directly from coffee intact beans (without sample preparation) and ground coffees, in order to evaluate the influence of sample preparation on the performance of the analytical technique. The samples were directly analyzed by diffuse reflectance in a FT-NIR system (Spectrum 100 - Perkin Elmer, USA). The intact coffee beans were analyzed using a NIRA accessory (Perkin Elmer-Waltman, USA, NIRA model) equipped with a spinning sample module (Sample spinner, Perkin Elmer-Waltman, USA, NIRA model). The ground coffees were analyzed in appropriate vials (Wheaton, USA, shell vial). The scanning spectra were obtained in triplicate, for both intact coffee beans and ground coffees, in the spectral range between  $10,000\text{ cm}^{-1}$  and  $4000\text{ cm}^{-1}$ , with a resolution of  $4\text{ cm}^{-1}$  and 16 scans (Ferreira et al., 2015; Silva et al., 2021).

### 2.3. NIR spectra data preprocessing and exploratory analysis

NIR spectra data were imported into Matlab R2019a software (The Mathworks, Natick, MA) with the PLS\_Toolbox 8.6 computer package. The 222 original spectroscopic profiles of samples with 6001 variables per spectrum were transformed into pseudo-absorbance ( $\log 1/R$ ) and the wavenumbers, in  $\text{cm}^{-1}$  were converted into wavelengths in nanometers (nm).

Different preprocessing techniques were evaluated prior to exploratory analysis, including first-order (1st) and second-order (2nd) derivatives of Savitzky-Golay (10-point window), multiplicative scatter correction (MSC), and standard normal variate (SNV). The impact of each approach was assessed in terms of enhancing the differences between sample classes.

Subsequently, the preprocessed spectra were subjected to exploratory analysis. Principal Component Analysis (PCA), an unsupervised method, was applied to examine sample distribution and identify trends in the data, evaluating the influence of each preprocessing technique on the distinction between intact and ground coffee beans. Prior to PCA, the data were mean-centered.

#### 2.4. Model for discriminating canephora coffee beans with GI by PLS-DA

For the development of the PLS-DA discrimination models, the samples were separated by the Kennard-Stone algorithm (Kennard & Stone, 1969), in which the training data sets comprised 75 % of the samples of each geographical origin and the prediction (external validation) sets, formed by the remaining 25 % of the samples of each class. The presence of outliers was assessed by analyzing Q residuals and Hotelling's  $T^2$  values, but no outliers were detected thus no samples were excluded.

The internal model validation was performed using the Venetian Blinds cross-validation method with five data splits. In this procedure, the dataset was divided into five intercalated subsets, ensuring that each split contained a balanced representation of the data. In each iteration, one subset was left out for validation, while the remaining four were used for training the model. The number of samples per blind was set to 33 samples per subset, considering the total of 167 samples. The maximum number of latent variables (LVs) was 10 according to the Bayes' Theorem (based on classification errors), explained variance of the Y-block and the ROC curves in order to obtain the best model based on sensitivity and specificity (Eqs. (1) and (2)).

The Soft Independent Modeling by Class Analogy (SIMCA) model is one of the most widely applied algorithms for class modeling problems in food science and was also determined to be used as a classification method for comparison with the PLS-DA discriminant method.

Performance metrics, such as sensitivity and specificity, were evaluated during the training, cross-validation, and prediction phases, providing a comprehensive assessment of the classification model's performance. Where, sensitivity is the ability to correctly identify validation samples from a specific class, while specificity measures the correct rejection of samples from other classes. These values can range from 0 % to 100 % and are calculated based on true positives, false positives, and false negatives in the prediction dataset. Models with sensitivity and specificity greater than 80 % in their predictions are considered effective (Ferreira, 2015).

$$\text{Sensitivity (\%)} = \frac{TP}{TP + FN} \times 100 \quad (1)$$

$$\text{Specificity (\%)} = \frac{TN}{TN + FP} \times 100 \quad (2)$$

Where TP refers to true positives, FN to false negatives, TN to true negatives, and FP to false positives.

The importance of the variable in the projection, known as VIP scores was used to access the variables responsible for discrimination (Baqueta et al., 2019a; Ribeiro et al., 2011).

### 3. Results and discussion

#### 3.1. NIR spectra data preprocessing and exploratory analysis

The raw spectra of intact beans (Fig. 1.A) and ground beans (Fig. 1.B) show overlap and indistinguishable spectral profiles to the naked eye, making it difficult to visually identify the geographical origins of the coffees. When comparing the average NIR spectra of coffees from different geographical origins, for both intact and ground beans, the spectral profiles were very similar, displaying primarily absorptions corresponding to the characteristic vibrational modes of coffee. However, distinct spectral variations were observed in the following regions: 1700–1750 nm, mainly associated with absorptions from the stretching vibration of the C=O bond (carbonyls, aldehydes, ketones, esters, amides), and 1950–2500 nm, which includes C–H stretching vibrations, particularly for methyl and methylene groups, as well as H-O-H absorptions related to water. The pre-processing using Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) showed similar effects, both between the techniques and across the sample types,

highlighting differences primarily between 1500 and 1600 nm, and from 2000 nm to 2500 nm. These differences may be related to variations in the composition of the green coffee beans, influenced by factors such as the processing method and geographical origin. The application of the first derivative, to both sample types (intact Fig. 1.G and ground Fig. 1.H), emphasized differences primarily in the region between 2000 and 2500 nm and around 1600 nm, regions sensitive to variations in the bean composition, such as the presence of organic compounds and water, which can differ depending on the origin and processing of the beans. Finally, the second derivative contributed only slight differences in the average intensity of the spectra, keeping the class profiles very similar. This technique provided subtle improvements in distinguishing between the samples, but the variations remained small, indicating that geographical origin separation might be more challenging to identify using this pre-processing alone.

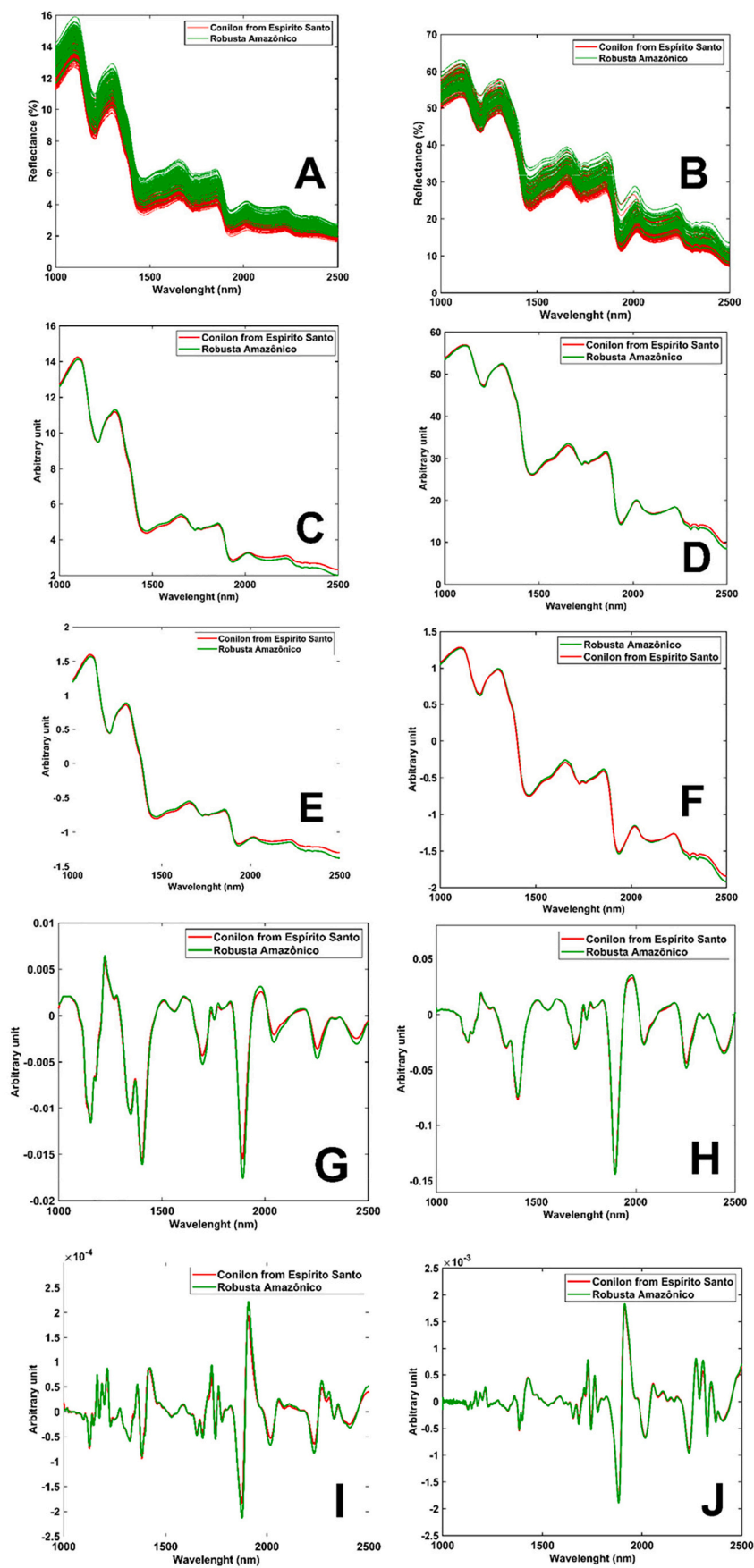
Despite these differences, the lack of selectivity of the technique and the high similarity between the signals of different sample classes make spectral interpretation challenging. This highlights the need to employ discriminant chemometric tools to explore the spectral data, ensuring greater reliability in the evaluation of the results. The differences between the average spectra of each sample type can be explained by the physical characteristics of the samples.

Another relevant factor is the chemical transformations that occur in green coffee during and after the grinding process. In general, grinding food exposes it to light and air, causing alterations mainly related to oxidative processes in lipids and changes in phenolic compounds (Batista et al., 2021; Rodriguez-Amaya & Shahidi, 2021). According to de Pires et al. (2021), grinding roasted coffee beans can cause significant physical and chemical changes, such as the breaking of cellular structures and walls, exposing previously inaccessible compounds. This process can trigger the oxidation of lipids and phenolic compounds present in the beans, as well as the release of organic acids and sugars, which in turn influence the spectral characteristics of the coffee. The release of volatile compounds and modification of antioxidant compounds can also be observed, affecting spectral profiles and making them distinct between intact and ground samples. A similar behavior can be expected for green coffee, which, although a different matrix from roasted coffee, also contains lipids, phenolic compounds, volatiles, antioxidants, acids, and sugars (Farah, 2018). Studies on green tea (Zhao et al., 2023) have demonstrated that grinding can disrupt the intramolecular hydrogen bonds of cellulose and lignin, leading to modifications in the spectral profile. Given that green coffee beans also contain significant amounts of cellulose and lignin (Farah, 2018), it is plausible that a similar effect occurred in this study. These observations further support the idea that the differences in NIR fingerprints are primarily driven by differences in chemical composition.

These chemical transformations are responsible for the variations observed in the spectra, particularly in the regions of 1500–1600 nm, associated with the C–H stretch vibration (methylene and methyl), and in the 1700–1750 nm region, associated with the C=O stretch vibration (carbonyls), in addition to the regions between 1950 and 2500 nm, related to C–H stretching vibrations (mainly for methyl and methylene groups) and H-O-H, associated with changes in water levels and volatile compounds present in the coffee (de Pires et al., 2021; Kucharska-Ambrożej et al., 2021; Yüksel et al., 2020).

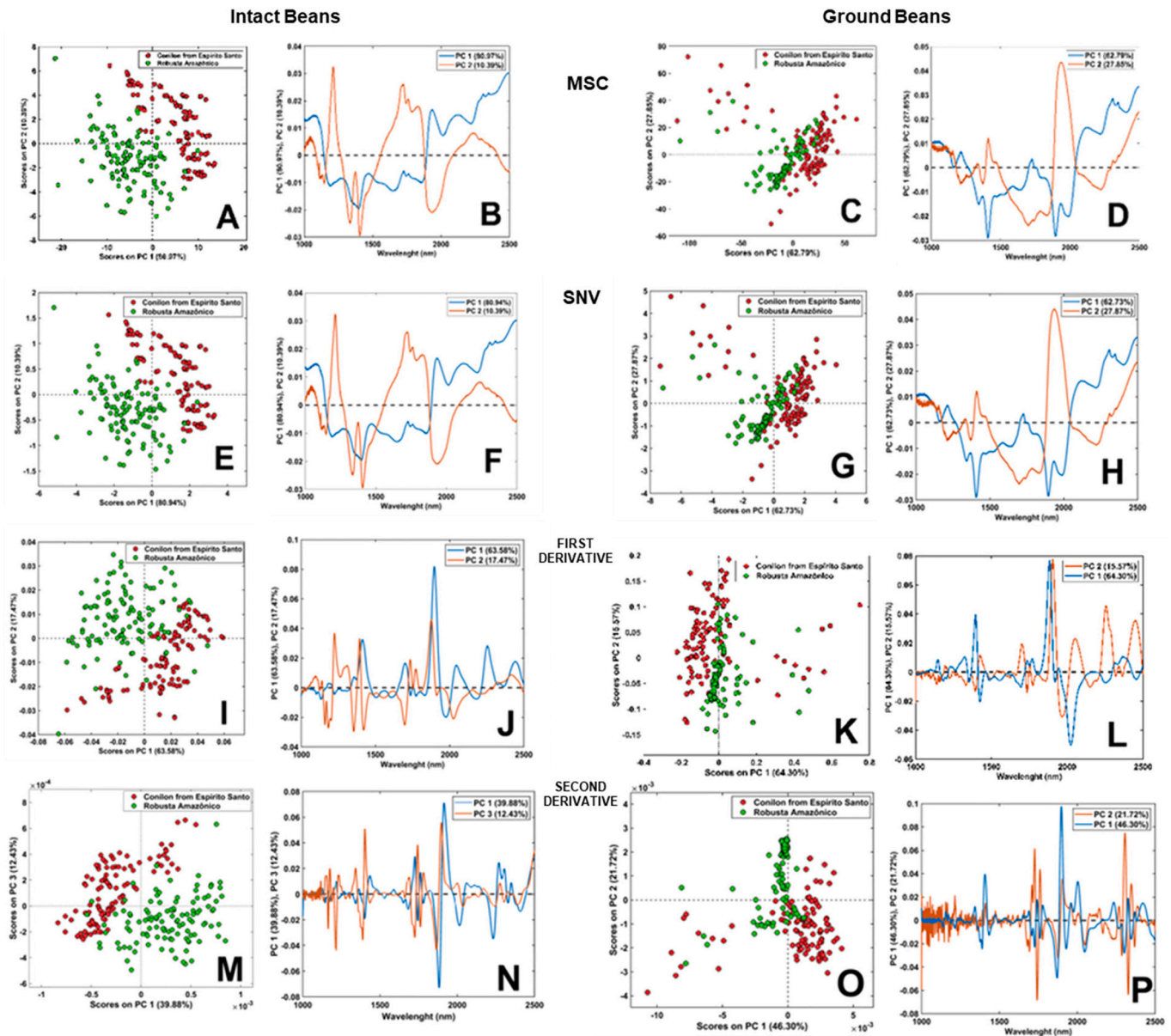
The PCAs (Fig. 2) revealed distinct patterns between Robusta Amazônico and Conilon from Espírito Santo, both in intact and ground samples. The application of pre-processing techniques (MSC, SNV, and 1st and 2nd derivatives by Savitzky-Golay enhanced the visualization of the grouping trends, leading to a clear separation between the coffee varieties analyzed. Ground coffee samples exhibited greater spectral complexity compared to intact samples, as evidenced by the requirement for additional principal components (PCs) to capture the variability in the data (Table 1).

PCA indicated that, when considering all pre-processing methods, the spectral regions of 1000–1256 nm and 1895–2500 nm were the most



(caption on next page)

**Fig. 1.** Raw (A), averages of pre-processed with MSC (C), SNV (E), first derivative (G) and second derivative (I) spectra for the samples of canephora coffee beans. And raw (B), averages pre-processed with MSC (D), SNV (F), first derivative (H) and second derivative (J) spectra for the ground samples. The red lines represent Conilon samples from Espírito Santo and the green lines represent Robusta Amazônico samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 2.** PCA with scores and respective loadings of the beans spectra pre-processed with MSC (A-B), SNV (E-F), first derivative (I-M), second derivative (M-N). And for ground samples pre-processed with MSC (C-D), SNV (G-H), first derivative (K-L) and second derivative (O-P). The samples represented in the scores in red represent Conilon from Espírito Santo coffees and in green Robusta Amazônico coffees. In the loadings, the blue line represents the loadings in the first PC used and the orange line represents the second PC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

significant for differentiating the Robusta and Conilon coffees. These bands correspond to absorptions from C–H bonds (found in lipids and carbohydrates), O–H bonds (associated with moisture and polyphenols), and N–H bonds (linked to proteins and alkaloids such as caffeine and trigonelline) (Giraud et al., 2019). The PCA score plots, using combinations of two PCs, revealed clear graphical separation based on geographic origin, particularly for the intact samples. The variance explained by the first two components ranged from 52.31 % (second derivative – PC1 versus PC3 – Fig. 2M) to 91.30 % (MSC – PC1 versus PC2 – Fig. 2A) for intact grains. For the ground grains, this variation ranged from 67.0 % (second derivative – PC1 versus PC2 –

Fig. 2O) to 90.6 % (MSC and SNV – PC1 versus PC2 – Fig. 2C and G), the main features with detailed results of each PCA model are described in Table 1.

The SNV and MSC pre-processing methods exhibited similar effects, revealing a separation trend based on geographic origin. However, for the ground samples, greater intra- and interclass clustering was observed, indicating reduced spectral variability within the samples. This effect can be attributed to the homogenization of compounds in the samples after grinding. Nevertheless, some Robusta samples overlapped with those of Conilon, particularly in the ground samples, suggesting that grinding led to greater homogenization among these varieties,

**Table 1**  
Main features of each PCA model generated.

Pre-processing	Intact Samples			Ground samples		
	PCs	EV (%)	Most important ranges in PCA (nm)	PCs	EV (%)	Most important ranges in PCA (nm)
MSC	8	99.3	PC1: (+) 1000–1142 (+) 1895–2500 (-) 1143–1894	9	99.7	PC1: (+) 1000–1256 (+) 2039–2500 (-) 1267–1701 (-) 1757–2038
SNV	6	99.1	PC1: (+) 1000–1142 (+) 1895–2500 (-) 1143–1894	9	99.7	PC1: (+) 1000–1256 (+) 2039–2500 (-) 1267–1701 (-) 1757–2038
1st D	4	91.7	PC1: (+) 1212–1333 (+) 1375–1404 (+) 1720–1792 (+) 1820–1903 (+) 2299–2479	5	96.6	PC1: (+) 1367–1411 (+) 1826–1912 (-) 1955–2147
2nd D	5	79.3	PC1*: (+) 1701–1735 (+) 1897–1977 (+) 2254–2320 (+) 2444–2500 (-) 1837–1897 (-) 1975–2042 (-) 2188–2253	3	78.5	PC1: (+) 1407 (+) 1896 (+) 2002 (-) 1876 (-) 2046 (-) 2301
			PC2: (+) 1164–1264 (+) 1564–1876 (-) 1269–1554 (-) 1877–2079			PC2: (+) 1000–1728 (+) 1392–1483 (+) 1880–2043 (+) 2297–2500 (-) 1881–1317 (-) 1349–1391 (-) 1484–1879 (-) 2044–2280
			PC2: (+) 1164–1264 (+) 1564–1876 (-) 1269–1554 (-) 1877–2079			PC2: (+) 1000–1728 (+) 1392–1483 (+) 1880–2043 (+) 2297–2500 (-) 1881–1317 (-) 1349–1391 (-) 1484–1879 (-) 2044–2280
			PC3*: (+) 1174 (+) 1400 (+) 1744 (+) 1881–1912 (+) 2481–2500 (-) 1164 (-) 1214 (-) 1372–1391 (-) 1706–1730 (-) 1760 (-) 1846–1877			PC2: (+) 1672–1723 (+) 1873–1935 (+) 2014–2107 (+) 2217–2314 (+) 2370–2500 (-) 1935–2013
			PC2: (+) 1164–1264 (+) 1564–1876 (-) 1269–1554 (-) 1877–2079			PC2: (+) 1705–1733 (+) 1766 (+) 1897–1916 (+) 2295–2316 (+) 2350 (-) 1693 (-) 1742 (-) 1879 (-) 2284 (-) 2527

PCs: Principal Components; EV: Explained Variance; D: Derivative; (+) and (–) indicated the positive and negative portion of loading scores.

making their distinction more challenging. Analysis of other principal components beyond PC1, which captures the majority of the variance, revealed that more subtle differences among the varieties could still be identified.

The observed variability also reflects intrinsic differences among producers within the same geographical indication (GI), suggesting that the separation of geographical origins can be influenced by both natural variability and the specific agricultural practices of each producer (Farah, 2018; Pham et al., 2019). Analysis of the loadings (Fig. 2 and Table 1) revealed that, for MSC and SNV preprocessing, Conilon is more associated with the spectral region of 1000–1100 nm, for both intact and ground grains. This suggests that specific Conilon compounds exert a greater influence on this spectral range. In the case of Robusta, the association occurs in different spectral bands depending on the grain condition. For intact grains, the region of 1100–1900 nm shows the greatest contribution, highlighting the importance of this range for distinguishing Robusta. This trend was observed for both MSC and SNV, demonstrating that these techniques preserve Robusta's distinct spectral characteristics. For intact grains, this influence extends up to 2000 nm, suggesting that the intact grain's structure retains additional spectral signatures that may be lost after milling.

Analysis of the PCA models set after NIR data was derivate by 1st and/or 2nd derivative, revealed more prominent peaks in the 1700–2500 nm region, crucial for geographic differentiation (Fig. 2.J, 2. L, 2.N and 2.P). This observation indicating that the use of derivatives improved spectral resolution, emphasizing small variations that could be masked and/or not captures when the data was preprocessed by MSC and SNV. These differences are associated with specific chemical

compounds, such as C–H, O–H, and N–H bonds, which are essential in coffee beans' composition and reflect variations in lipid, protein, and water concentrations between varieties (Munyendo et al., 2021). The analysis of ground samples indicated that grinding resulted in a more compact grouping, with less spectral variability within coffee classes. This may relate to the exposure of intracellular compounds, such as lipids, whose absorptions are detected in the 1701–1735 nm and 1897–1977 nm regions, and proteins and alkaloids like caffeine, in bands around 2100–2200 nm, associated with NH bond vibrations (Giraud et al., 2019; Magalhaes et al., 2016). Applying second-derivative preprocessing (Fig. 2.N and 1.P), regarding the PC1 loadings, it was possible to observe similar bands previously observed when 1st derivative preprocessing was applied, while regarding the PC2 it was possible to notice higher presence of noise across the spectral region, indicating introduction of non-informative data on PCA model developed on the data preprocesses by 2nd derivative.

Overall, PCAs revealed a clear trend toward separation between classes, with PC1 explaining most of the variance and reflecting the major chemical differences between the two groups. Ground coffee samples exhibited a more compact distribution in PCA space, indicating that grinding significantly affects the spectral profile of the beans. This highlights the need to account for this factor in chemometric modeling for future geographic origin discrimination. The exploratory analysis showed that geographical origin directly influences NIR spectra, a crucial factor for distinguishing samples based on their chemical composition. The spectral differences between Robusta and Conilon varieties, as well as between geographic origins, can be attributed to variations in constituent concentrations, involving numerous complex

factors such as plant variety, soil type, geographic location, climatic conditions, and agricultural practices. The differentiation trend observed in PCA reflects these variations and suggests that a supervised discrimination model could be particularly effective in recognizing these two geographic origins, especially when using combinations of preprocessing techniques.

### 3.2. Models for discrimination of samples of green coffee beans with GI by PLS-DA

To achieve this, PLS-DA models were developed using different preprocessing techniques, either individually or in combination. In both cases, the most effective preprocessing was the combination of MSC with the 1st derivative. The best model for discriminating intact coffee beans was built with 5 LVs, while 8 LVs were required for ground coffee samples. As previously discussed, ground coffee samples exhibit greater chemical complexity and different reflectance effects, possibly leading to the loss of relevant chemical information for discrimination, particularly those associated with the surface or husk of the coffee bean. The results are described in Table 2.

The analysis of the discriminant models revealed consistent and robust performance. For intact samples, models based on SNV and MSC alone already exhibited sensitivity and specificity values of at least 97.7 % in the training and test phases, indicating high discriminative ability without requiring more complex approaches. However, adding the first derivative further improved overall performance, achieving values of at least 98.8 % in both training and test sets with only 5 LVs. Notably,

applying only the first derivative yielded similar performance but required 7 LVs. Conversely, a performance reduction was observed when preprocessing methods included the second derivative, either alone or in combination with other approaches. This suggests that, the second derivative amplified noise, as discussed in the PCA results in Section 3.1, thereby reducing model effectiveness.

The best model developed for intact coffee beans demonstrated exceptional performance in distinguishing between the Robusta Amazônico and Conilon from Espírito Santo classes. For the Robusta Amazônico class, sensitivity reached 100 % in both the training and test sets, with 96.5 % in cross-validation. Specificity values were 98.8 %, 96.5 %, and 100 % for the training, cross-validation, and test sets, respectively. For the Conilon from Espírito Santo class, sensitivity ranged from 98.8 % in training, 96.5 % in cross-validation, to 100 % in the test set, while specificity was 100 % in both training and test sets and 96.3 % in cross-validation. The slight reduction in cross-validation suggests that the model may be somewhat sensitive to intra-class variations, but the consistently high predictive performance confirms its robustness.

When the same preprocessing strategy (MSC and first derivative) was applied to ground green coffee, performance was even higher, achieving 100 % sensitivity and specificity across all datasets (training, cross-validation and test). Grinding contributed to reducing spectral variability within each class, which may have enhanced the performance of the supervised models. However, the need for a higher number of latent variables for ground samples suggests that grinding increases spectral complexity due to the exposure of new chemical structures. This result

**Table 2**

Figures of merit of the PLS-DA models for Conilon from Espírito Santo (CES) and Robusta Amazônico (RA) samples.

Intact beans									
Pré-processing	LVs	Class	Training		Cross Validation		Prediction		
			Sens. (%)	Spec. (%)	Sens. (%)	Spec. (%)	Sens. (%)	Spec. (%)	
SNV	5	CES	100	97.7	97.5	96.5	100	100	
		RA	97.7	100	96.5	97.5	100	100	
MSC	5	CES	100	97.7	97.5	96.5	100	100	
		RA	97.7	100	96.5	97.5	100	100	
1st D	7	CES	98.8	100	96.3	97.7	100	100	
		RA	100	98.8	97.7	96.3	100	100	
2nd D	5	CES	98.8	100	91.4	94.2	96.3	100	
		RA	100	98.8	94.2	91.4	100	96.3	
SNV + 1st D	5	CES	98.8	97.7	91.4	95.3	96.3	100	
		RA	97.7	98.8	95.3	91.4	100	96.3	
SNV + 2nd D	5	CES	100	100	93.8	96.5	88.9	100	
		RA	100	100	96.5	93.8	100	88.9	
<b>MSC + 1st D</b>	<b>5</b>	<b>CES</b>	<b>100</b>	<b>98.8</b>	<b>96.3</b>	<b>96.5</b>	<b>100</b>	<b>100</b>	
		<b>RA</b>	<b>98.8</b>	<b>100</b>	<b>96.5</b>	<b>96.3</b>	<b>100</b>	<b>100</b>	
MSC + 2nd D	5	CES	100	100	91.4	96.5	88.9	100	
		RA	100	100	96.5	91.4	100	88.9	
Ground beans									
Pre-processing	LVs	Class	Training		Cross validation		Prediction		
			Sens. (%)	Spec. (%)	Sens. (%)	Spec. (%)	Sens. (%)	Spec. (%)	
SNV	8	CES	98.8	98.8	95.1	94.2	100	100	
		RA	98.8	98.8	94.2	95.1	100	100	
MSC	8	CES	98.8	98.8	95.1	94.2	100	100	
		RA	98.8	98.8	94.2	95.1	100	100	
1st D	7	CES	98.8	100	93.8	84.2	100	100	
		RA	100	98.8	94.2	93.8	100	100	
2nd D	8	CES	100	100	95.1	97.7	100	100	
		RA	100	100	97.7	95.1	100	100	
SNV + 1st D	8	CES	100	100	100	98.8	100	100	
		RA	100	100	98.8	100	100	100	
SNV + 2nd D	7	CES	100	100	96.3	97.7	100	100	
		RA	100	100	97.7	96.3	100	100	
<b>MSC + 1st D</b>	<b>8</b>	<b>CES</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	
		<b>RA</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	
MSC + 2nd D	6	CES	100	100	95.1	98.8	100	100	
		RA	100	100	98.8	95.1	100	100	

LVs – Latent Variables; Sens.: Sensitivity; Spec.: Specificity; D: Derivative.

highlights the effectiveness of the combined strategy, ensuring complete class separation regardless of the modeling stage. Nevertheless, the similar performance values across different preprocessing approaches indicate that, for ground samples, the model's high performance may be more stable against preprocessing variations (Engel et al., 2013; Rinnan et al., 2009). These findings emphasize the critical role of careful preprocessing selection in maximizing PLS-DA performance.

VIP score analysis was performed to identify the most important spectral variables for discriminating geographical indication of coffee beans (Fig. 3). The highly influential variables are those with VIP scores greater than 1 (Galindo-Prieto et al., 2015).

The most influential regions in discriminating the geographical origin of the intact beans were 1219 to 1227 nm, 1378 to 1415 nm, 1689 to 1711 nm, 1728 to 1740 nm, 1853 to 1933 nm, 1945 to 1973 nm, 2002 to 2091 nm, 2224 to 2300 nm, and 2416 to 2500 nm. For the ground coffee, the significant VIP score regions were 1135 to 1139 nm, 1149 to 1158 nm, 1219 to 1227 nm, 1329 to 1354 nm, 1378 to 1432 nm, 1686 to 1716 nm, 1726 to 1742 nm, 1748 to 1757 nm, 1865 to 1928 nm, 1954 to 2047 nm, 2056 to 2097 nm, 2230 to 2307 nm, and 2409 to 2500 nm.

There are no significant VIP scores among 1135–1139 nm, 1149–1158 nm, and 1148–1757 nm for ground beans relative to intact beans, indicating that these bands can be considered specific NIR-VIP variables for discriminating the geographical origins of ground coffee. The VIP scores for intact beans reached higher values, suggesting a stronger spectral contribution of certain compounds in their composition. Spectral regions around 1135–1139 nm and 1149–1158 nm are typically associated with second overtones of C–H stretching vibrations, which can be linked to lipids and certain carbohydrates. Meanwhile, the broader region 1148–1757 nm encompasses bands related to first overtones and combination bands of O–H, C–H, and N–H bonds, often attributed to moisture, proteins, and complex carbohydrates such as cellulose and hemicellulose. The higher VIP scores in intact beans suggest that these compounds contribute more significantly to their spectral profile, potentially due to differences in structure which contributes to differences in the results obtained by reflectance and influences radiation scattering differently.

The other similarities or distinctions of the VIP scores are associated with the intensities of the absorption peaks of the components present in the beans. Among these components, some have higher absorption, from 1400 to 2500 nm. The peak near 1400 nm is associated with carbohydrates, chlorogenic acids, and lipids. The scores between 1600 and 1800 nm showed that in both types of samples, the wavelengths provide important information to differentiate the samples and are related to

caffeine, but carbohydrates also presents bands in this region. In the region between 2000 and 2500 nm, the scores also have importance in the discrimination of ground and intact beans of different GI and relevant predictors and associated with the bands of combined NH, OH, and CO bonds that are characteristic of carbohydrates, lipids, proteins, chlorogenic acids, and caffeine (Baqueta et al., 2019; Barbin et al., 2014; Ribeiro et al., 2011).

Considering previous studies, classification methods are also widely applied for coffee recognition involving spectroscopic techniques. Thus, SIMCA classification models were developed using different preprocessing techniques, either individually or in combination, and the results are expressed in the supplementary material in Table S1. The most effective model for evaluating coffee beans was obtained using only SNV as pre-processing, with 5 PCs. For ground coffee samples, despite the high prediction merit values for the various pre-processing methods, all of them had unsatisfactory sensitivity and specificity values.

Robusta Amazônico coffee beans displayed distinguishable differences from Conilon beans from Espírito Santo when analyzed using NIR, and PLS-DA supervised classification demonstrated its effectiveness in differentiating the samples. Therefore, it was possible to discriminate the botanical/geographical geographical origin of raw beans of Brazilian specialty canephora coffees with GIs, granted through NIR spectroscopy of the intact or ground samples. It is possible to state that besides the geographical origin, the physical characteristics of the samples exert changes in the spectral profiles, but both models presented excellent performance for use as GIs samples control. The analysis of ground green beans samples was more efficient, with values of 100 % for all the parameters evaluated, despite the greater number of variables. However, it requires higher costs and is more laborious due to the sample preparation technique, while the analysis of intact beans is faster, non-destructive and cheaper, thus being more viable for the developed application, and can be used in quality control of producers, industries, certification agencies and associations involved in the evaluation of certified coffee beans (Giraudó et al., 2019). Once intact grain analysis proved satisfactory and more feasible considering sample preparation, based on the VIP scores of the developed model.

The results of the models obtained in this study are similar and even superior to those of previous research with various applications of green coffee geographical origin discriminant analysis, such as in the discrimination of beans from different cities in the state of Paraná, Brazil, with 66.67 % correct classification using MIR spectroscopy and radial-basis function network modeling (Link et al., 2014). Marquetti

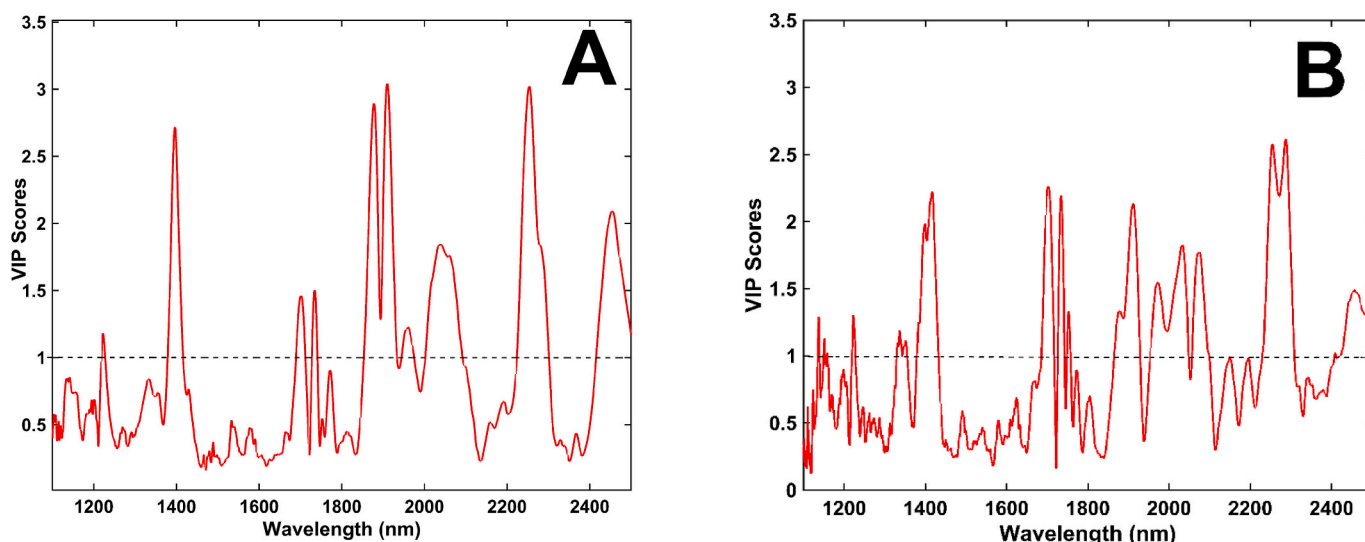


Fig. 3. VIP scores from intact (A) and ground beans (B). The dashed line indicates the threshold.

et al. (2016) using NIR spectroscopy with PLS-DA, obtained specificity and sensitivity values of 35.7 % in the discrimination of grains also from different cities in the state of Paraná, Brazil. A study of the recognition of the geographical origin of grains from different countries in Asia and Africa using NIR spectroscopy and PLS-DA modeling, obtained efficiency results from 93 % (Giraudou et al., 2019).

In previous studies conducted with Brazilian canephora coffees, similar samples to those used in this research, sensitivity and specificity rates ranging from 89.3 % to 100 % were obtained for the discrimination between green Conilon beans from Espírito Santo and Robusta from the Amazon, using UV-Vis spectroscopy and PLS-DA analysis (de Moraes-Neto et al., 2024). These results highlight an excellent discriminatory capacity between the two coffee varieties, demonstrating the effectiveness of this technique in identifying differences between the green canephora coffee beans.

It is important to highlight that, although NIR spectroscopy combined with chemometrics is widely used for coffee authentication, the present study brings significant advances by focusing on a scarcely explored matrix: green beans of Brazilian canephora coffees with recognized GIs. An important differentiator is the direct comparison between intact and ground samples, showing that the analysis of intact beans can provide comparable or even superior results, eliminating the need for grinding. Additionally, we rigorously evaluated different spectral pre-processing strategies, and this approach allowed us to develop models with high sensitivity and specificity, reinforcing the applicability of the method. However, the methodology cannot be universally applied to all coffee origins and types. Each case must be evaluated individually, as the characteristics of coffee can vary significantly between different regions, varieties, production processes, and sampling conditions, directly influencing the spectral response and, consequently, the performance of chemometric models (Zhang et al., 2018).

#### 4. Conclusion

In this study, NIR spectroscopy associated with chemometrics showed promise and discrimination of botanical variety and geographical origin in green beans of Brazilian specialty canephora coffees, with granted GIs. The exploratory analysis provided a comprehensive understanding, pointing out the spectrochemical differences between Amazon Robusta and Conilon Espírito Santo coffees, as well as between intact and ground beans.

NIR spectroscopy coupled with PLS-DA modeling was performed with high efficiency when the 1st Savitzky-Golay derivative combined with MSC was used as pre-processing, and can be applied for rapid discrimination regardless of the type of sample preparation. The NIR spectra provided information about the geographical origin and variety of the samples and indicated that the main differences among the classes were related to carbohydrates, chlorogenic acids, lipids, proteins, and caffeine. Therefore, the developed method is feasible for applications in control and geographical origin certification procedures, since it is a non-invasive and easy-to-operate technique, presented low-cost, sensitivity and can be applied directly to intact canephora coffees green beans samples.

#### CRedit authorship contribution statement

**Venancio Ferreira de Moraes-Neto:** Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Michel Rocha Baqueta:** Writing – original draft, Methodology, Conceptualization. **Alexsandro Lara Teixeira:** Resources. **Elem Tamirys dos Santos Caramês:** Writing – original draft, Methodology. **Juliana Azevedo Lima Pallone:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

#### Declaration of competing interest

None.

#### Acknowledgements

The authors would like to thank the Coordination for the Improvement of Higher Education Personnel (CAPES), Finance Code 001, and the National Council for Scientific and Technological Development (CNPq), grant #162505/2022-1 and #310933/2023-4, for their support in conducting this study. The authors also acknowledge the São Paulo Research Foundation (FAPESP) for the financial support, grants #2021/14060-1 and #2022/03268-3.

#### Data availability

Data will be made available on request.

#### References

- Abernethy, G., & Higgs, K. (2013). Rapid detection of economic adulterants in fresh milk by liquid chromatography-tandem mass spectrometry. *Journal of Chromatography A*, 1288, 10–20. <https://doi.org/10.1016/j.chroma.2013.02.022>
- Babova, O., Occhipinti, A., & Maffei, M. E. (2016). Chemical partitioning and antioxidant capacity of green coffee (*Coffea arabica* and *Coffea canephora*) of different geographical origin. *Phytochemistry*, 123, 33–39. <https://doi.org/10.1016/j.phytochem.2016.01.016>
- Baqueta, M. R., Coqueiro, A., & Valderrama, P. (2019). Brazilian coffee blends: A simple and fast method by near-infrared spectroscopy for the determination of the sensory attributes elicited in professional coffee cupping. *Journal of Food Science*, 84(6), 1247–1255. <https://doi.org/10.1111/1750-3841.14617>
- Baqueta, M. R., Marini, F., Teixeira, A. L., Goulart, B. H. F., Pilau, E. J., Valderrama, P., & Pallone, J. A. L. (2024). Spectroscopic and sensory characterization of Brazilian *Coffea canephora* terroir and botanical varieties produced in the Amazon and Espírito Santo implementing multi-block approaches. *Journal of Food Composition and Analysis*, 133, Article 106442. <https://doi.org/10.1016/j.jfca.2024.106442>
- Barbin, D. F., de Felicio, A. L. S. M., Sun, D. W., Nixdorf, S. L., & Hirooka, E. Y. (2014). Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview. *Food Research International*, 61, 23–32. <https://doi.org/10.1016/j.foodres.2014.01.005>
- Batista, A. G., da Silva-Maia, J. K., & Maróstica, M. R. (2021). Generation and alterations of bioactive organosulfur and phenolic compounds. In D. B. Rodriguez-Amaya, & J. Amaya-Farfan (Eds.), *Chemical changes during processing and storage of foods* (pp. 537–577). Academic Press. <https://doi.org/10.1016/B978-0-12-817380-0.00011-7>.
- Bona, E., Marquetti, I., Link, J. V., Makimori, G. Y. F., da Costa Arca, V., Guimarães Lemes, A. L., ... Poppi, R. J. (2017). Support vector machines in tandem with infrared spectroscopy for geographical classification of green arabica coffee. *LWT - Food Science and Technology*, 76, 330–336. <https://doi.org/10.1016/j.lwt.2016.04.048>
- Botelho, B. G., Oliveira, L. S., & Franca, A. S. (2017). Fluorescence spectroscopy as tool for the geographical discrimination of coffees produced in different regions of Minas Gerais state in Brazil. *Food Control*, 77, 25–31. <https://doi.org/10.1016/j.foodcont.2017.01.020>
- Caballero, D., Ríos-Reina, R., & Amigo, J. M. (2021). Chemometrics and food traceability. In *Comprehensive Foodomics* (pp. 387–406). Elsevier. <https://doi.org/10.1016/B978-0-08-100596-5.22859-x>.
- CONAB. (2024). Acompanhamento da Safra Brasileira de café - Safra, Primeiro levantamento. *Obs. Agrícola*, 10, 10–11. <https://www.conab.gov.br/info-agro/safra/s/cafe>.
- Corrêa, P. C., de Oliveira, G. H. H., de Oliveira, A. P. L. R., Vargas-Elfas, G. A., & Baptestini, F. M. (2016). Particle size and roasting on water sorption in conilon coffee during storage. *Coffee Science*, 11(2), 221–233. <https://doi.org/10.25186/cs.v11i2.1061>
- De Luca, S., De Filippis, M., Bucci, R., Magri, A. D., Magri, A. L., & Marini, F. (2016). Characterization of the effects of different roasting conditions on coffee samples of different geographical origins by HPLC-DAD, NIR and chemometrics. *Microchemical Journal*, 129, 348–361. <https://doi.org/10.1016/j.microc.2016.07.021>
- Efenberger-Szmechtyk, M., Nowak, A., & Kregiel, D. (2018). Implementation of chemometrics in quality evaluation of food and beverages. *Critical Reviews in Food Science and Nutrition*, 58(10), 1747–1766. <https://doi.org/10.1080/10408398.2016.1276883>
- Endaye, M., Atlabachew, M., Mehari, B., Alemayehu, M., Mengistu, D. A., & Kerisew, B. (2020). Combining multi-element analysis with statistical modeling for tracing the origin of green coffee beans from Amhara region, Ethiopia. *Biological Trace Element Research*, 195(2), 669–678. <https://doi.org/10.1007/s12011-019-01866-5>
- Engel, J., Gerretzen, J., Szymańska, E., Jansen, J. J., Downey, G., Blanchet, L., & Buydens, L. M. (2013). Breaking with trends in pre-processing? *TrAC Trends in Analytical Chemistry*, 50, 96–106. <https://doi.org/10.1016/j.trac.2013.04.015>

- Farah, A. (2018). Coffee: production, quality and chemistry. In *Coffee: Production, quality and chemistry*. <https://books.google.es/books/about/Coffee.html?id=LMk0jEACAAJ&pgis=1>.
- Ferreira, D. S., Pallone, J. A. L., & Poppi, R. J. (2015). Direct analysis of the main chemical constituents in *Chenopodium quinoa* grain using Fourier transform near-infrared spectroscopy. *Food Control*, *48*, 91–95. <https://doi.org/10.1016/j.foodcont.2014.04.016>
- Ferreira, M. M. C. (2015). *Quimiometria: conceitos, métodos e aplicações*. Editora da Unicamp.
- Galindo-Prieto, B., Eriksson, L., & Trygg, J. (2015). Variable influence on projection (VIP) for OPLS models and its applicability in multivariate time series analysis. *Chemometrics and Intelligent Laboratory Systems*, *146*, 297–304. <https://doi.org/10.1016/j.CHEMOLAB.2015.05.001>
- Gatuszka, A., Migaszewski, Z., & Namieśnik, J. (2013). The 12 principles of green analytical chemistry and the SIGNIFICANCE mnemonic of green analytical practices. In , Vol. 50. *TrAC - trends in analytical chemistry* (pp. 78–84). Elsevier B.V. <https://doi.org/10.1016/j.trac.2013.04.010>.
- Giraud, A., Grassi, S., Savorani, F., Gavoci, G., Casiraghi, E., & Geobaldo, F. (2019). Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis. *Food Control*, *99*(December 2018), 137–145. <https://doi.org/10.1016/j.foodcont.2018.12.033>
- Hoyos Ossa, D. E., Gil-Solsona, R., Peñuela, G. A., Sancho, J. V., & Hernández, F. J. (2018). Assessment of protected designation of origin for Colombian coffees based on HRMS-based metabolomics. *Food Chemistry*, *250*(January), 89–97. <https://doi.org/10.1016/j.foodchem.2018.01.038>
- INPI. (2021). *Brazil, Instituto Nacional da Propriedade Industrial, Indicações Geográficas*.
- Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics*, *11*(1), 137–148. <https://doi.org/10.1080/00401706.1969.10490666>
- Kucharska-Ambrożej, K., Martyna, A., Karpińska, J., Kietyka-Dadasiewicz, A., & Kubat-Sikorska, A. (2021). Quality control of mint species based on UV-VIS and FTIR spectral data supported by chemometric tools. *Food Control*, *129*, Article 108228. <https://doi.org/10.1016/J.FOODCONT.2021.108228>
- Link, J. V., Lemes, A. L. G., Marquetti, I., dos Santos Scholz, M. B., & Bona, E. (2014). Geographical and genotypic classification of arabica coffee using Fourier transform infrared spectroscopy and radial-basis function networks. *Chemometrics and Intelligent Laboratory Systems*, *135*, 150–156. <https://doi.org/10.1016/j.chemolab.2014.04.008>
- Ludwig, I. A., Clifford, M. N., Lean, M. E. J., Ashihara, H., & Crozier, A. (2014). Coffee: Biochemistry and potential impact on health. *Food & Function*, *5*(8), 1695–1717. <https://doi.org/10.1039/c4fo00042k>
- Magalhaes, L. M., Machado, S., Segundo, M. A., Lopes, J. A., & Páscoa, R. N. (2016). Rapid assessment of bioactive phenolics and methylxanthines in spent coffee grounds by FT-NIR spectroscopy. *Talanta*, *147*, 460–467. <https://doi.org/10.1016/j.talanta.2015.10.022>
- Malta, M. R., de Fassi, L. O., Liska, G. R., Carvalho, G. R., Pereira, A. A., Botelho, C. E., ... Pereira, R. G. F. A. (2020). Discrimination of genotypes coffee by chemical composition of the beans: Potential markers in natural coffees. *Food Research International*, *134*(March). <https://doi.org/10.1016/j.foodres.2020.109219>
- Manley, M., & Baeten, V. (2018). Spectroscopic technique: Near infrared (NIR) spectroscopy. In *Modern techniques for food authentication* (pp. 51–102). Elsevier. <https://doi.org/10.1016/b978-0-12-814264-6.00003-7>.
- Marquetti, I., Link, J. V., Lemes, A. L. G., dos Scholz, M. B. S., Valderrama, P., & Bona, E. (2016). Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. *Computers and Electronics in Agriculture*, *121*, 313–319. <https://doi.org/10.1016/j.compag.2015.12.018>
- Mees, C., Souard, F., Delporte, C., Deconinck, E., Stoffelen, P., Stévigny, C., ... de Braekeleer, K. (2018). Identification of coffee leaves using FT-NIR spectroscopy and SIMCA. *Talanta*, *177*, 4–11. <https://doi.org/10.1016/J.TALANTA.2017.09.056>
- Mehari, B., Redi-Abshiro, M., Chandravanshi, B. S., Combrinck, S., McCrindle, R., & Atlabachew, M. (2019). GC-MS profiling of fatty acids in green coffee (*Coffea arabica* L.) beans and chemometric modeling for tracing geographical origins from Ethiopia. *Journal of the Science of Food and Agriculture*, *99*(8), 3811–3823. <https://doi.org/10.1002/jsfa.9603>
- de Moraes-Neto, V. F., Baqueta, M. R., dos Caramês, E. T. S., de Santana, F. B., Alves, E. A., & Pallone, J. A. L. (2024). Discrimination of Brazilian green canephora coffee beans by ultraviolet-visible spectroscopy as a non-target analysis: A tool for recognizing geographical indications. *Microchemical Journal*, *202*(February). <https://doi.org/10.1016/j.microc.2024.110737>
- Munyendo, L., Njoroge, D., & Hitzmann, B. (2021). The potential of spectroscopic techniques in coffee analysis—A review. *Processes*, *10*(1), 71. MDPI <https://doi.org/10.3390/pr10010071>.
- Myles, A. J., Brown, S. D., & Zimmerman, T. A. (2006). Transfer of multivariate classification models between laboratory and process near-infrared spectrometers for the discrimination of Green Arabica and Robusta Coffee Beans. *Applied Spectroscopy*, *60*(10), 1198–1203. <https://opg.optica.org/as/abstract.cfm?uri=as-60-10-1198>.
- Oliveri, P. (2017). Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues – A tutorial. In , Vol. 982. *Analytica Chimica Acta* (pp. 9–19). Elsevier B.V. <https://doi.org/10.1016/j.aca.2017.05.013>.
- Otto, M. (2023). *Chemometrics: Statistics and computer application in analytical chemistry*. John Wiley & Sons.
- Pallone, J. A. L., dos Caramês, E. T. S., & Alamar, P. D. (2018). Green analytical chemistry applied in food analysis: Alternative techniques. *Current Opinion in Food Science*, *22*(Figure 1), 115–121. <https://doi.org/10.1016/j.cofs.2018.01.009>
- Pham, Y., Reardon-Smith, K., Mushtaq, S., & Cockfield, G. (2019). The impact of climate change and variability on coffee production: A systematic review. *Climatic Change*, *156*, 609–630. <https://doi.org/10.1007/s10584-019-02538-y>
- de Pires, F. C., Pereira, R. G. F. A., Baqueta, M. R., Valderrama, P., & Alves da Rocha, R. (2021). Near-infrared spectroscopy and multivariate calibration as an alternative to the Agron to predict roasting degrees in coffee beans and ground coffees. *Food Chemistry*, *365*(June), Article 130471. <https://doi.org/10.1016/j.foodchem.2021.130471>
- Ribeiro, J. S., Ferreira, M. M. C., & Salva, T. J. G. (2011). Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. *Talanta*, *83*(5), 1352–1358. <https://doi.org/10.1016/J.TALANTA.2010.11.001>
- Rinnan, Å., Van Den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, *28*(10), 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- Rodriguez-Amaya, D. B., & Shahidi, F. (2021). Oxidation of lipids. In D. B. Rodriguez-Amaya, & J. Amaya-Farfan (Eds.), *Chemical Changes During Processing and Storage of Foods* (pp. 125–170). Academic Press. <https://doi.org/10.1016/B978-0-12-817380-0.00004-X>.
- Santos, J. R., Sarraguça, M. C., Rangel, A. O. S. S., & Lopes, J. A. (2012). Evaluation of green coffee beans quality using near infrared spectroscopy: A quantitative approach. *Food Chemistry*, *135*(3), 1828–1835. <https://doi.org/10.1016/J.FOODCHEM.2012.06.059>
- Silva, J. G. S., dos Caramês, E. T. S., & Pallone, J. A. L. (2021). Additives and soy detection in powder rice beverage by vibrational spectroscopy as an alternative method for quality and safety control. *LWT*, *137*, Article 110331. <https://doi.org/10.1016/j.lwt.2020.110331>
- Spink, J. W. (2019). *Food fraud prevention: Introduction, implementation, and management* (In Food Fraud Prevention).
- Vegro, C. L. R., & Almeida, L. F. (2019). Global coffee market: Socio-economic and cultural dynamics. In *Coffee consumption and industry strategies in Brazil: A volume in the consumer science and strategic marketing series* (pp. 3–19). Elsevier. <https://doi.org/10.1016/B978-0-12-814721-4.00001-9>.
- Worku, M., Upadhayay, H. R., Latruwe, K., Taylor, A., Blake, W., Vanhaecke, F., ... Boeckx, P. (2019). Differentiating the geographical origin of Ethiopian coffee using XRF- and ICP-based multi-element and stable isotope profiling. *Food Chemistry*, *290* (July 2018), 295–307. <https://doi.org/10.1016/j.foodchem.2019.03.135>
- Yüksel, A. N., Özkara Barut, K. T., & Bayram, M. (2020). The effects of roasting, milling, brewing and storage processes on the physicochemical properties of Turkish coffee. *LWT*, *131*, Article 109711. <https://doi.org/10.1016/J.LWT.2020.109711>
- Zhang, C., Shen, T., Liu, F., & He, Y. (2018). Identification of coffee varieties using laser-induced breakdown spectroscopy and Chemometrics. *Sensors*, *18*(1), 95. <https://doi.org/10.3390/s18010095>
- Zhao, Z., Dai, Z., Jiang, X., Yu, L., Hu, M., Peng, J., & Zhou, F. (2023). Influence and optimization of long-time superfine grinding on the physicochemical features of green tea powder. *Journal of Food Composition and Analysis*, *117*, Article 105124. <https://doi.org/10.1016/J.JFCA.2022.105124>
- Zhu, M., Long, Y., Ma, Y., Chen, Y., Yu, Q., Xie, J., Li, B., & Tian, J. (2021). Comparison of chemical and fatty acid composition of green coffee bean (*Coffea arabica* L.) from different geographical origins. *LWT*, *140*(December 2020), Article 110802. <https://doi.org/10.1016/j.lwt.2020.110802>