Journal of Food Composition and Analysis 116 (2023) 105065

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

journal homepage: www.elsevier.com/locate/jfca

Brazilian Canephora coffee evaluation using NIR spectroscopy and discriminant chemometric techniques

Michel Rocha Baqueta [a], Enrique Anastácio Alves [b], Patrícia Valderrama [c,*,1], Juliana Azevedo Lima Pallone [a,*,2]

[a] University of Campinas – UNICAMP, School of Food Engineering, Department of Food Science and Nutrition, Campinas, São Paulo, Brazil
[b] Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA Rondônia, Porto Velho, Rondônia, Brazil
[c] Universidade Tecnológica Federal do Paraná – UTFPR, Campo Mourão, Paraná, Brazil

ARTICLE INFO

Keywords:
Amazonian robusta
Conilon
Geographical origin
Multivariate classification
NIR spectroscopy
PLS-DA

ABSTRACT

High-quality Brazilian Canephora coffees are rising to the level of specialty coffees in the face of a new industry perception. In this framework, spectra from 527 coffees were analyzed in the near-infrared (NIR) region. Principal component analysis distinguished Brazilian Canephora producing states, botanical varieties, low and high-quality Canephora, Canephora and Arabica, and Canephora with geographical indication (GI) from those without GI. Also, Canephora coffee cultivars from Western Brazilian Amazon were distinguished. Three multi-class PLS-DA (traditional, hard, and soft versions) were compared to discriminate 5 classes: Robusta Amazônico from traditional (1) and indigenous (2) producers of Rondônia, Conilon from Espírito Santo (3), Conilon from Bahia (4), and specialty Arabica (5). Binary PLS-DA discriminated GI Canephora and non-GI Canephora with 100% sensitivity and specificity. Carbohydrates, chlorogenic acids, lipids, caffeine, and proteins were dominant absorption bands in coffee classifications. The proposed method is objective, simple, fast, and could be used in the routine analysis of coffee to verify claims of identity, variety, and origin.

1. Introduction

The rise in quality standards of Brazilian Canephora coffees to the level of specialty ones has been contributing to change the current industry's perception on Canephora quality. Many efforts have been made to understand more about it and an example is the application of analytical methods that are more objective, simpler, and faster to evaluate coffee identity claims. However, coffee matrix is challenging because the beans contain many different chemical compounds that influence the analysis, requiring a coupling of these methods with chemometrics to produce models with the chemical information that has been obtained.

Among coffee (*Coffea* spp.) species, *Coffea canephora* is the second most important, representing 40% of the global crop (ICO, 2021). Despite this, it has historically been described as a low-quality coffee compared to *Coffea arabica* and has led many coffee studies to differentiate Arabica from Canephora. Brazil is the second largest world producer of Canephora (ICO, 2021). Espírito Santo (Southeast region), Rondônia (North region), and Bahia (Northeast region) are the main producing states (Brazil, 2021a), see Fig. 1. They grow Robusta and Conilon (Souza et al., 2021, 2018). While Espírito Santo produces Conilon mainly in its north and Bahia mainly in its extreme south, region known as Atlântico Baiano, Rondônia produces Robusta. Despite their diversity, they has not yet been subjected to a comprehensive analysis together, in view that origin and variety have been important factors in coffee discrimination (Robert et al., 2022).

Brazilian Conilon and Robusta have shown an evolution in quality standards over the last five years, with several factors, including cultivation, processing, and post-harvest, contributing to improve their sensory quality. They have been called high quality Canephora, Fine Robusta or specialty Canephora, and differ greatly from a low-quality Canephora considered as commodity. They have also been qualified with equal and even higher sensorial quality when compared to specialty Arabica (Alves et al., 2020; Dalazen et al., 2020; Fioresi et al., 2021; Lemos et al., 2020; Machado et al., 2021; Oliveira et al., 2020a, 2020b; Pereira et al., 2019).

* Corresponding authors.
E-mail addresses: pativalderrama@gmail.com, patriciav@utfpr.edu.br (P. Valderrama), jpallone@unicamp.br (J.A.L. Pallone).
[1] ORCID: 0000-0002-6680-1260
[2] ORCID: 0000-0002-8517-2922

https://doi.org/10.1016/j.jfca.2022.105065
Received 9 September 2022; Received in revised form 19 November 2022; Accepted 24 November 2022
Available online 26 November 2022
0889-1575/

Canephora quality has been a hot topic for research and has received more attention in the last five years, especially after two geographical indication (GI) registrations. This register is a sign used in Brazil to identify and protect specialty foods. It is divided into Denomination of Origin (DO), which is linked to a specific geographical area, and Indication of Origin (IO), which is linked to the notoriety of where it is produced (Brazil, 2021b). In 2021, Robusta from Rondônia and Conilon from Espírito Santo were nationally recognized and protected with GI seals of DO and IO, respectively (Brazil, 2021c, 2021d). However, Conilon from Bahia is not currently registered with GI.

Robusta Amazônico is the name given to the Robusta coffee exclusively produced in Rondônia, in the DO Matas de Rondônia, which is located in the Western Brazilian Amazon region, resulting in a unique terroir (Brazil, 2021d). Such coffees may exhibit chocolate, woody, fruity, spicy, rooty, and/or herbaceous characteristics (Brazil, 2021e). Matas de Rondônia region covers fifteen cities, highlighted in green in Fig. 1. Robusta Amazônico has a sustainable and agroforestry production model and became the first Canephora to be registered with a GI worldwide. Its producers are also distinctive, since there are indigenous and non-indigenous people (Brazil, 2021f, 2019). The indigenous people comprise 127 families of different ethnicities living in two protected Indigenous Lands called "Sete de Setembro" and "Rio Branco", in purple and yellow, respectively in Fig. 1, which are in the Matas de Rondônia boundaries (Brazil, 2019). Since its coffee has a positive impact on the promotion of these populations and social inclusion, which is a public policy issue, a differentiation between Robusta Amazônico producers is required to increase its value and appeal.

Conilon from Espírito Santo is another outstanding coffee and is also known as Conilon Capixaba (Brazil, 2021c). The state is the largest national Conilon producer and cultivates it mainly in its northern region. (Brazil, 2021a). However, there are also producers based in the southern or central region, as shown in Fig. 1. Its sensory profile may present chocolate, almond, floral, or fruity characteristics (Brazil, 2021e).

Brazilian Canephora requires a protection tool to defend its identity, origin, and quality due to its increasing popularity among consumers and producers. In this sense, near-infrared (NIR) spectroscopy is a relevant technique for food geographical verification and traceability. It performs a non-targeted spectrochemical analysis, providing a spectral fingerprint that can be used to confirm the identity of a sample (Ríos-Reina et al., 2021). Also, it is easy-to-use, instantly scans the sample with just one click, allows direct solid sample analysis, is relatively inexpensive and industrially applicable, allowing large-scale fingerprinting (Baqueta et al., 2021). Despite its functionality, it suffers in identifying specific chemical compounds in a sample, because they are evaluated through the combined absorptions (Ríos-Reina et al., 2021).

Chemometric tools are required to analyze the chemical information contained in NIR spectra. There are several discriminant analysis available, but the Partial Least Squares with Discriminant Analysis (PLS-DA) is the most popular (Pomerantsev and Rodionova, 2018). It is widely used in food analytical chemistry (Foschi et al., 2021; Oliveri and Downey, 2013; Rodionova and Pomerantsev, 2020). PLS-DA can be applied to model two classes in binary discriminations or three or more classes in multi-class situations. It is typically applied based on PLS scores, but recently, new versions have been proposed. One of them called soft is based on Quadratic Discriminant Analysis (QDA), while a hard version is based on Linear Discriminant Analysis (LDA). Soft and hard PLS-DA do not require a large computational effort for implementation and work differently by not using PLS scores and loadings in the modeling (Pomerantsev and Rodionova, 2018; Zontov et al., 2020). They are relatively new chemometric tools, therefore, further research must explore their performance, especially in coffee classification context.

Therefore, important coffee parameters, including geographical origin, variety, species, and authenticity were investigated with direct solid sample analysis by NIR technology combined with spectral data analysis by chemometrics. A representative set composed of 527 samples was analyzed, including Robusta Amazônico from indigenous and non-indigenous producers, Conilon from Espírito Santo and Bahia, specialty Arabica, and low-quality Canephora. No reference has been found on the discrimination and classification of Canephora coffees of different qualities, as well as GI Canephora from those without GI in coffee analytical chemistry. In addition, this study highlights important responses to the use of NIR spectroscopy in large-scale agronomic research to directly differentiate Canephora cultivars from Western Brazilian Amazon (Teixeira et al., 2020) using a non-targeted analysis simpler than those (Faria et al., 2022) used in genetic breeding programs for the identification of coffee genotypes.

2. Materials and methods

2.1. Coffee samples

A total of five hundred and twenty-seven genuine roasted and ground

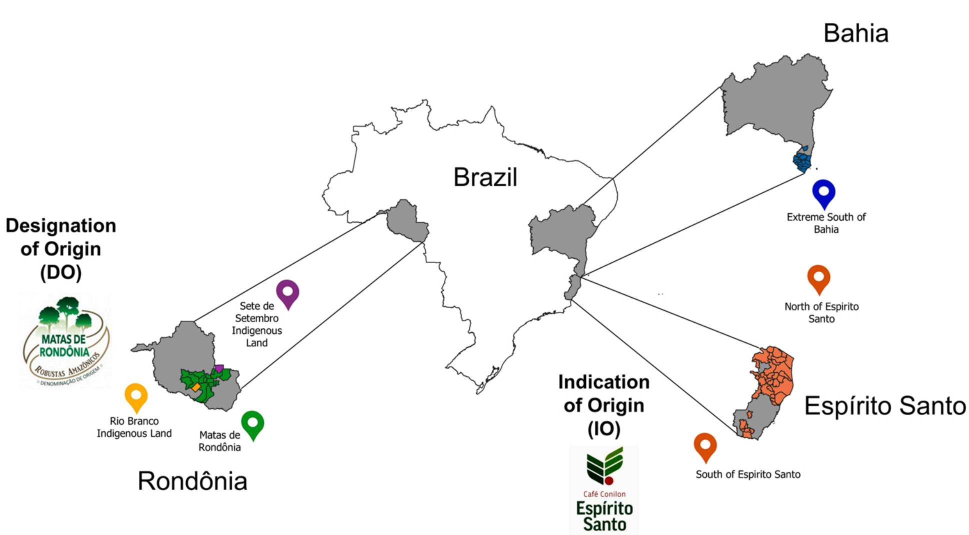

Fig. 1. Graphical representation showing main Brazilian geographical regions responsible for Canephora production (in gray) and the places of origin of the samples (colored highlights) and respective GI seals: Espírito Santo (Conilon) highlighted in orange, Matas de Rondônia (Robusta) highlighted in green, where the "Sete de Setembro" Indigenous Land is highlighted in purple, and "Rio Branco" Indigenous Land is highlighted in yellow, and Bahia (Conilon) highlighted in blue.

coffees were collected, including Canephora and Arabica samples. All samples were harvested in 2020 and roasted to a medium degree for standardization. More information about the samples is shown in Table 1.

2.1.1. GI Canephora samples

Among the 232 samples collected from Rondônia, 99 were from indigenous producers, comprising 78 from "Sete de Setembro" and 21 from "Rio Branco",133 were from non-indigenous producers. It is worth mentioning that the indigenous samples were exclusively processed through induced fermentation, as they are usually produced. Rondônia sample accessions from non-indigenous producers were not known, but some varietal characteristics were known a priori and may be of interest for study and discrimination: 22 were Robusta, 17 were Conilon, and 56 were Conilon and Robusta hybrids. In addition, 38 were Robusta Apoatã, which is of agronomic interest for its high yield and resistance.

From Espírito Santo, 126 Conilon samples were collected, of which 61 were from the northern cities. Other samples were from the south or central region of the state (65 samples). Specific sample origins were unknown to those in the north because they were provided by a local cooperative; however, the others were sourced directly from the producers, as follows: Conceição de Castelo - Central region (8 samples), Nova Venécia - Northwestern region (33 samples), Alegre - Southern region (2 samples), Atílio Vivácqua - Southern region (1 sample), Jerônimo Monteiro - Southern region (2 samples), Mimoso do Sul - Southern region (5 samples), Muqui - Southern region (11 samples), and São Domingos do Norte - Northwestern region (3 samples).

Robusta Amazônico from Rondônia and Conilon from Espírito Santo were high-quality Canephora coffees and had GI specifications. Therefore, GI Canephora mention throughout the text refers to them. These samples were provided by the EMBRAPA Rondônia, which guaranteed their authenticity.

Table 1
Coffee samples considered in the study.

States of Canephora production		n° of samples	Variety	Origin
Rondônia		78	Robusta	Sete de Setembro indigenous land
		21	Robusta	Rio Branco indigenous land
	Total	99		
		38	Robusta Apoatã	Matas de Rondônia
		8	Robusta	Matas de Rondônia
		7	Hybrid	Matas de Rondônia
		25	Hybrid	Matas de Rondônia
		14	Robusta	Matas de Rondônia
		22	Hybrid	Matas de Rondônia
		2	Hybrid	Matas de Rondônia
		17	Conilon	Matas de Rondônia
	Total	133		
Espírito Santo		61	Conilon	Cities in the North of Espírito Santo
		8	Conilon	Conceição de Castelo - Central Region
		33	Conilon	Nova Venécia - Northwestern Region
		2	Conilon	Alegre - Southern Region
		1	Conilon	Atílio Vivácqua - Southern Region
		2	Conilon	Jerônimo Monteiro - Southern Region
		5	Conilon	Mimoso do Sul - Southern Region
		11	Conilon	Muqui - Southern Region
		3	Conilon	São Domingos do Norte - Northwestern Region
	Total	126		
Bahia		6	Conilon	Teixeira de Freitas
		9	Conilon	Eunápolis
		27	Conilon	Extreme South of Bahia
		33	Conilon	Blended coffees
	Total	75		
São Paulo	Total	7	Conilon	Adamantina
Arabica		42	Unknown variety	Pure origin
		33	Unknown variety	Blended coffees
	Total	75		
Low-quality Canephora		6	Robusta	Unknown origin
		6	Conilon	Unknown origin
	Total	12		

2.1.2. Non-GI Canephora samples

Conilon coffees from Bahia of high or intermediate quality were collected directly from different producers or companies. Among the 75 samples, 27 were pure origin and 33 blended. In addition, two producers had a history of specialty Conilon production, where one from Teixeira de Freitas provided 6 samples and another from Eunápolis provided 9 samples.

For further investigation and comparison, 7 Conilon from São Paulo state, located in the southeast of Brazil, were collected in Adamantina city, and had unknown quality grade. Also, 6 low quality Robusta and 6 low quality Conilon, called low-quality Canephora throughout the text, were collected as a control to compare them with specialty Canephora.

2.1.3. Specialty Arabica samples

A total of 75 specialty Arabica coffees from different Brazilian origins, qualities and sensory characteristics were purchased in local markets or provided by companies/producers for species discrimination between Arabica and Canephora. They were pure origin (42 samples) or blended (33 samples) and had chocolate, nuts, floral, or fruity characteristics.

2.2. NIR spectroscopic analysis

Roasted coffee beans were milled and then sieved through a 20-mesh sieve for particle size standardization. NIR spectroscopic fingerprints were obtained from the solid coffees, in ground form, in reflectance mode, using a Perkin Elmer Fourier Transform NIR spectrophotometer, Spectrum 100 N, equipped with a glass cuvette. Each spectrum was digitized with 32 scans from 1000 to 2500 nm with a resolution of 4 nm. Roasted and ground coffee samples were analyzed without any laborious sample pre-treatment in a random sequence at room temperature (22 °C) by placing them directly on the instrument. Three different sample aliquots were used, and the spectrum of each aliquot was recorded, resulting in 1581 (527 × 3) spectral profiles acquired. Before analysis, the blank was evaluated using a NIR reflectance standard.

2.3. Data processing and exploratory analysis

Each sample's average NIR spectrum was calculated and imported in Matlab R2019a (The Mathworks, Natick, MA) with the PLS_Toolbox 8.6 computational package. The 527 original spectroscopic profiles with 6001 variables per spectrum were transformed into pseudo-absorbance (log 1/R). Different pre-processing methods were studied before chemometric data analysis individually or in combination. However, the combination of Savitzky–Golay smoothing and first derivative (5 points window) (Savitzky and Golay, 1964), and multiplicative scatter correction (MSC) (Geladi et al., 1985) was the most effective pre-process to correct baseline variations and the different light scattering of granulated samples. A spectrum segment was removed from the whole wavelength range, and the most informative region between 1100 and 2500 nm with 5093 variables was used for analysis. Moreover, mean centering was performed on the spectra. First of all, several Principal

Component Analysis (PCA) models were built selecting the principal components (PCs) that could extract relevant chemical information about the samples.

2.4. Discrimination and classification methods

For PLS-DA, the samples belonging to each dataset were selected by the Kennard-Stone algorithm (Kennard and Stone, 1969). Calibration datasets (training sets) comprised 75% of the samples selected from each class. Prediction datasets (test sets) were composed of 25% of the remaining samples to evaluate the predictive ability of the models, and they were only used in the final model evaluation. PLS-DA models were built using the same pre-processing used in PCA.

The first objective was a comparison among the three multi-class PLS-DA versions to predict five coffee classes: Robusta Amazônico from indigenous producers (class 1 – 99 samples), Robusta Amazônico from non-indigenous (class 2 – 133 samples), Conilon from Espírito Santo (class 3 – 126 samples), Conilon from Bahia (class 4 – 75 samples) and specialty Arabica (class 5 – 75 samples). As the samples from "Rio Branco" Indigenous Land (21 samples) were inevitably limited by their production and availability and their post-harvest processing were similar, they were grouped with "Sete de Setembro" (78 samples) into a single class, totalizing 99 samples. Conilon from São Paulo (7 samples) and low-quality Canephora (12 samples) were not considered in the PLS-DA models because they were not enough to build models.

To build the 5-class discrimination models, multi-class PLS-DA based on PLS scores, named as traditional PLS-DA, was implemented using PLS_Toolbox in the Matlab, according to previous studies (Baqueta et al., 2021). Internal model validation was performed using venetian blinds cross-validation with five samples. In addition, samples with high leverage and high Q residuals simultaneously were removed since they could be outliers. Hard and soft multi-class PLS-DA models were built using a graphical interface available at https://github.com/yzontov/pls-da. Algorithm descriptions and instructions for implementing them can be found in the literature (Pomerantsev and Rodionova, 2018; Zontov et al., 2020). Monte Carlo cross-validation, error type I and outlier level of 0.01 were applied for both models.

The second objective was to develop binary PLS-DA models to discriminate GI Canephora and non-GI Canephora and test the models, evaluating the ability of NIR spectroscopy and chemometrics. These models were built with the multi-class PLS-DA that performed best.

Sensitivity and specificity for calibration (CAL), cross-validation (CV), and prediction (PRED) determined the quality and reliability of the PLS-DA models. They were also used to select the model complexity, determining the number of latent variables (LVs). Finally, a chemical interpretation of the most discriminant variables was performed by analyzing VIP (variable importance in projection) scores. NIR-VIP-bands associated with coffee compounds listed in the literature were investigated (Baqueta et al., 2021; Barbin et al., 2014; Pires et al., 2021; Ribeiro et al., 2011).

3. Results and discussion

3.1. Spectrum visualization

Fig. 2A shows the original spectra, Fig. 2B the pre-treated spectra, and Fig. 2C the average spectra of the coffee classes. When analyzing the average spectra (Fig. 2C), NIR fingerprints were very similar over most

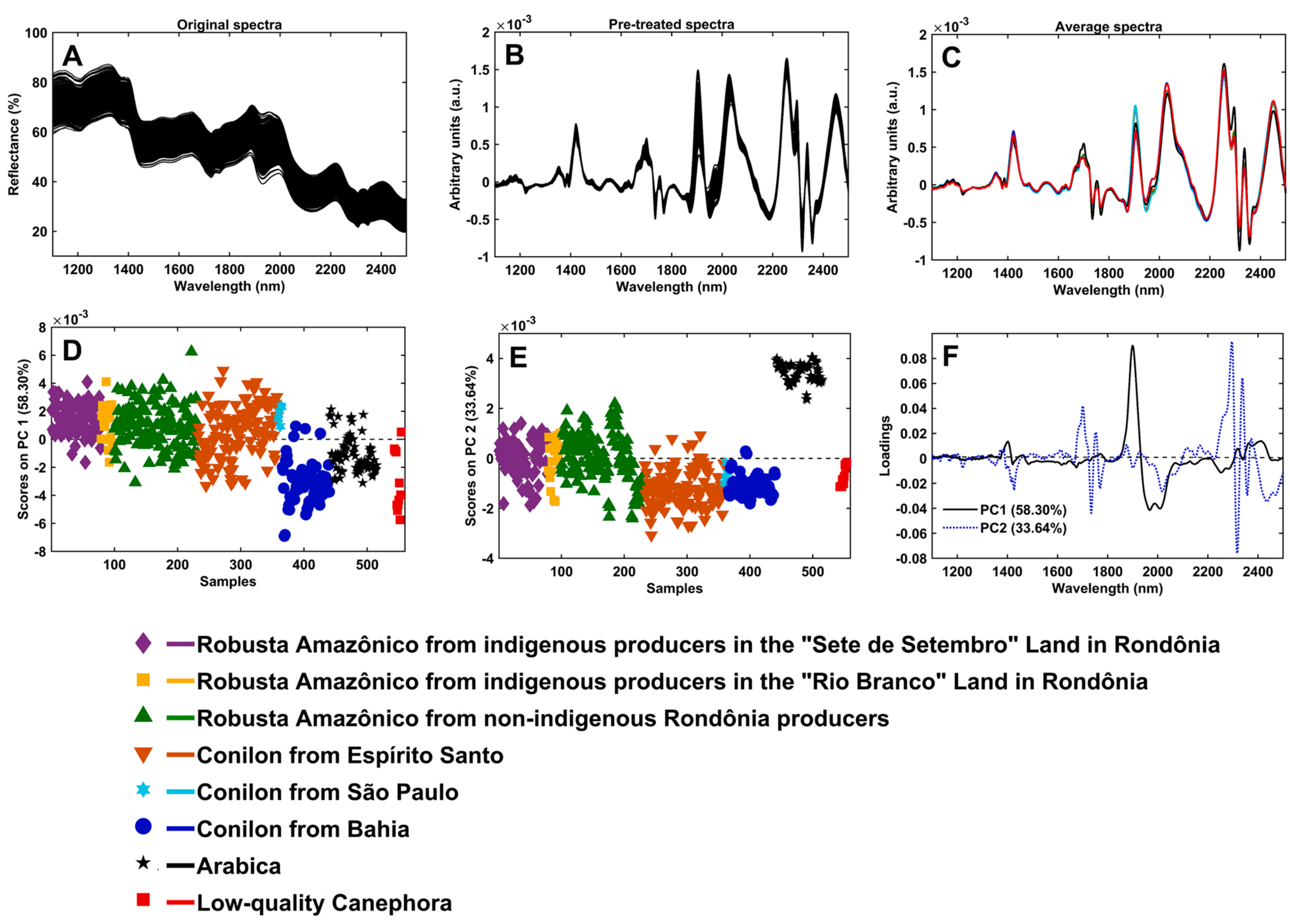

Fig. 2. Original (A), pre-treated (B), and average (C) NIR fingerprints, as well as PCA with data of all coffees showing scores on PC1 (D) and PC2 (E) and respective loadings on PC1 and PC2 (F).

of the range, except in some regions where the differences were visible. While the average spectra of the Robusta Amazônico and Conilon from Espírito Santo overlap (all classes with GI specifications), Conilon from São Paulo and Bahia (both without GI specifications) showed some differences. The botanical difference between the Robusta and Conilon varieties was not evident at first glance, because their spectra overlapped, indicating that they had similar spectrochemical characteristics.

GI Canephora coffees spectra had a slightly different behavior from the others. They exhibited chemical characteristics in common possibly associated with their special production mode (Fig. 2C). Low-quality Canephora coffees showed different spectral behavior compared to the other Canephora, indicating how proper bean-processing can improves the quality of Robusta and Conilon commodities.

Another clear difference was between Canephora and Arabica that had their spectrum affected by their compositions (Fig. 2C). This was probably due to a quantitative difference of the metabolites in each species that has been reported in the literature (Fioresi et al., 2021; Lemos et al., 2020). However, comparing the overall spectrum shape in Fig. 2B with those of pure Arabica available in the literature (Ribeiro et al., 2011), they were quite similar.

3.2. Exploratory analysis

3.2.1. PCA for all coffees

A global PCA model evaluated all samples and revealed distinctions according to their origin, quality, species, and variety (Fig. 2D-F). The first two PCs explaining, respectively, 58.30% and 33.64% of variance were extracted in this model. PC1 scores (Fig. 2D) distinguished the samples according to their qualities and origins. Robusta Amazônico, Conilon from Espírito Santo, and Conilon from São Paulo were placed on the most positive side of PC1, while Conilon from Bahia, Arabica, and low-quality Canephora were placed on the most negative side of PC1. This means that Robusta Amazônico and Conilon from Espírito Santo showed similar characteristics, agreeing with their GI specifications, and a distinct behavior when compared to other coffees (Conilon from Bahia, Arabica, and low-quality Canephora). Conilon from São Paulo showed a distribution near Conilon from Espírito Santo, suggesting that they may have similar characteristics.

PC2 scores (Fig. 2E), even explaining a lower percentage of variance (33.64%), contained helpful chemical information. It differentiated the samples by species, botanical variety, quality, and origin. With respect to species discrimination, Canephora (Conilon and Robusta samples) and Arabica were clearly distinguished. Arabica samples were placed on the positive side in a distant group. Regarding the botanical variety discrimination, most of the Conilon samples from Espírito Santo, São Paulo and Bahia presented negative scores, while Robusta coffees presented positive or negative scores. In addition, Robusta Amazônico samples from Rondônia were distributed among Arabica and other coffees, suggesting a differentiation by quality. However, this may not be exclusively in function of coffee quality only, but also of its variety and origin.

3.2.2. PCA for coffees from Rondônia

Fig. 3A-I shows PCA models for samples belonging to Rondônia. Firstly, a PCA was carried out with all samples, however, its first two PCs - PC1 (58.81%) and PC2 (24.25%) - showed no effective sample separation and were therefore not presented. PC3 (7.63%) brought desired results in Fig. 3A-B and discriminated indigenous and non-indigenous Robusta Amazônico. Indigenous coffees were placed on the most negative side of PC3, while the others were placed on the most positive side. This distinction may have occurred due to bean processing reasons. Indigenous people produce their coffees by induced fermentation, while the other producers often produce natural coffee.

Another model (Fig. 3C-D) was developed to try to discriminate only Robusta Amazônico produced by indigenous people. The first three PCs of this model (PC1 – 60.96%, PC2 – 26.19%, and PC3 – 4.70%) were not informative and only PC4 (2.31%) allowed a minimal information extraction (Fig. 3C). Rio Branco samples showed a more "condensed" cluster but were distributed among those from Sete de Setembro, showing positive and negative scores. Even selecting 4 PCs did not allow a clear distinction, indicating that they were chemically similar. Although their origins were different, they were similarly processed via fermentation, which probably influenced their chemical composition and consequently their spectra.

A PCA was performed to explore Rondônia coffees without indigenous samples (Fig. 3E-F). The first PC (60.51%) was not presented because only PC2 (25.56%) discriminated Robusta, Conilon, and hybrids. Robusta coffees were placed on the most positive side of PC2, while hybrid and Conilon samples were placed on the most negative one. Robusta Apoatã seemed not to follow a trend and was distributed among the other coffees.

Conilon samples were the most distinguishable in the above model. Therefore, they were removed, and a new PCA was performed with the remaining samples (Fig. 3G-I). PC1 (62.00%), PC2 (22.34%), PC3 (5.45%), PC4 (3.51%), and PC5 (2.77%) were considered, but only PCs 2 and 5 presented relevant information. On PC2, a discrimination between hybrid and Robusta was observed. On the other hand, PC5 showed a better Robusta Apoatã discrimination. Previous studies (Souza et al., 2021, 2018) have shown that genetic differences between Robusta, Conilon, and hybrids in Rondônia have a greater influence on the beverage nuances. Robusta and hybrids have a higher incidence of fruity, exotic, fine, and soft characteristics. Also, recent studies have shown other geographical discrimination studies for Canephora from Rondônia, but the variety effect has not been discussed (Robert et al., 2022).

3.2.3. PCA for Conilon from Espírito Santo

PCA was carried out to explore Conilon from Espírito Santo (Fig. 3J-L). Only PC4 (2.00%) and PC5 (1.15%) showed a tendency to differentiate the samples according to their producing regions, therefore, PC1 (78.09%), PC2 (10.95%), and PC3 (4.62%) were not displayed. On PC4, Conilon from southern cities called Conceição de Castelo, Atílio Vivácqua, Jerônimo Monteiro, Mimoso do Sul, and Muqui were mostly on the negative side, while Conilon from Nova Venécia and São Domingos do Norte, in the northern region, were mostly on the positive side. Other samples from northern Espírito Santo were distributed along of this PC, but their specific origins were unknown. Two samples from Alegre city, which is in the south of the state, were placed near those from the north. Similarly, a sample belonging to Muqui (south) was placed near those from the north. On PC5, samples from Conceição de Castelo, Mimoso do Sul, and Muqui – south of the state – showed a more 'condensed' cluster on the negative side, and samples from Alegre city, also in the south, were now placed nearby. Recent studies (Correia et al., 2020) have shown Canephora discrimination from different agroforestry systems in southern Espírito Santo. An electrospray mass spectrometry identification of coffee metabolites followed by portable microNIR spectroscopy, sensory analysis, and PCA allowed the authors to differentiate coffee characteristics within each system, even though the city was similar.

3.2.4. PCA for Conilon from Bahia

A PCA was carried out to explore Conilon from Bahia (Fig. 3M-P). They were pure or blended samples, and some of them of known origin. Among PC1 (77.02%), PC2 (9.99%), and PC3 (5.17%), only PC4 (2.79%) allowed a differentiation, distinguishing pure and blended Conilon (Fig. 3M). A new PCA model removing these samples was developed to differentiate only between Teixeira de Freitas and Eunápolis samples, that were from special producers (Fig. 3N). They were placed separately in PC1 (67.92%) versus PC3 (9.68%), indicating a difference in their compositions and probably qualities.

3.2.5. PCA for specialty Arabica coffees

Although the main aim was to explore Canephora coffees, a PCA

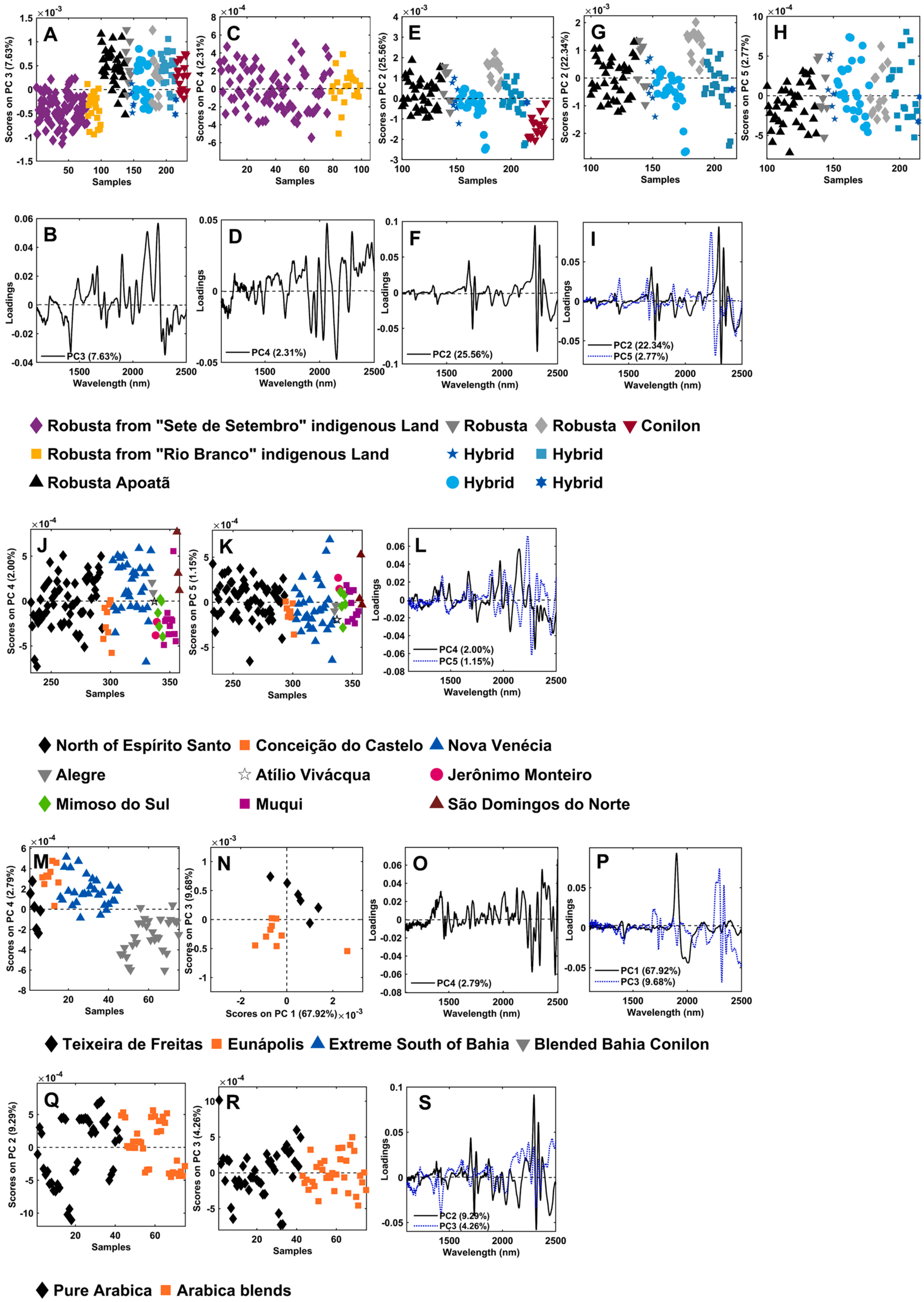

Fig. 3. PCA of Canephora coffees from Rondônia (A-I), Espírito Santo (J-L), Bahia (M-P), and Arabica (Q-S). A and B refer to scores and loadings of PC3, C and D refer to scores and loadings of PC4, E and F refer to scores and loadings of PC2, and G, H, and I refer to scores and loadings of PC2 and PC5.

model was also applied to evaluate Arabica samples (Fig. 3Q-S). With PC2 (9.29%) being more informative than PC1 (77.88%), it showed multiple sample clusters, suggesting that some groups may have similar composition and sensory characteristics. On the other hand, PC3 (4.26%) revealed that blended Arabica coffees were distributed among pure ones.

3.2.6. Considerations for exploratory analysis

The spectral data exploration by several PCA models showed that sample characteristics affected the NIR fingerprints. Origin, quality, species, and variety of coffee have proved to be crucial in distinguishing samples. Spectral differences were observed among the studied coffees, and even after roasting, they carried chemical information associated with their initial green coffee compositions. This information can be helpful for spectroscopic characterization in investigative studies and is of interest for coffee agronomic research, even when obtained through an exploratory analysis of unsupervised pattern recognition. In particular, NIR-based discrimination of Western Brazilian Amazon Canephora coffee cultivars provided an alternative preliminary analysis to be used on a large scale with direct solid analysis of coffee without any laborious procedure. A NIR instrument could be simpler than modern techniques used in genetic breeding programs for coffee genotype identification (Faria et al., 2022) and consequently decreases the sample volume to be analyzed by these expensive and labor-intensive techniques. In addition, the technique provided valuable chemical information not often evaluated in local agronomic studies in Rondônia (Teixeira et al., 2020).

Furthermore, it is worth noting that the PCs with the highest explained variance did not show favorable results for interpretation in most PCA. This achievement is not unusual, because sometimes the valuable chemical information in a data set is not contained in the PC that describes the highest explained variance, but in other PCs besides PC1. Studies have been addressing this situation (de Almeida et al., 2018; dos Santos et al., 2021; Moreira and Scarminio, 2013).

3.3. PLS-DA classification models

3.3.1. Comparison among multi-class PLS-DA versions

The number of selected LVs and sensitivity and specificity values (%) obtained in the traditional, hard, and soft PLS-DA models for multi-class classification are shown in Table 2a. A total of 18 LVs were selected in all models to make the results more comparable. In the 5-class discrimination, at least 17 LVs at are required, which is close to what was obtained. Although 18 seems too high, there is a consensus that this cannot be considered as overfit in PLS-DA, because each class takes around 2–3 LVs for the internal modeling and another 1–2 LVs to describe the external links between classes (Pomerantsev and Rodionova, 2018). Five coffee classes were discriminated and classified, being class 1 composed of Robusta Amazônico from indigenous producers, class 2 Robusta Amazônico from non-indigenous, class 3 Conilon from Espírito Santo, class 4 Conilon from Bahia, and class 5 Arabica.

High classification performances were obtained for all multi-class PLS-DA versions considering that this was a real-world application with a larger number of samples and high variability. Traditional PLS-DA based on PLS scores had the ideal results, with most samples being correctly assigned to their classes having sensitivity and specificity above 90% in prediction. Sensitivity and specificity varied insignificantly in traditional and hard PLS-DA. They perfectly discriminated Conilon from Bahia (class 4) and specialty Arabica (class 5) with values equal to 100% in the prediction but had some difficulties in the discrimination of 100% of the samples on classes 1, 2, and 3. The models discriminated and classified with sensitivity above 91.0% and specificity above 90.2%, which is highly satisfactory for real samples. These three classes (1, 2, and 3) were GI Canephora, and the misclassified samples might indicate that their characteristics and quality agree with GI regulations, not meaning a poor result. In addition, classes 1 and 2 were similarly composed of Robusta Amazônico, but under an attempt to differentiate between their origin from indigenous or non-indigenous producers. A detailed analysis showed that no sample from Bahia or specialty Arabica was belonged to the Robusta Amazônico (classes 1 and 2) and Conilon from Espírito Santo (class 3).

The results show that tracing the identity, quality, variety, and origin of Canephora through its NIR fingerprints is possible, because bean chemical composition affected the spectra. With a more definitive result by using supervised discriminant analysis, indigenous Robusta Amazônico had chemical difference and was spectroscopically distinguishable from the non-indigenous. In addition, Robusta Amazônico, regardless of the producer, as well as Conilon from Espirito Santo, Conilon from Bahia, and specialty Arabica coffees showed chemical

Table 2
Sensitivity, specificity (%), and LVs obtained in the (a) traditional, hard, and soft multi-class PLS-DA models corresponding to five coffees investigated; (b) Binary PLS-DA classification models to differentiate Robusta Amazônico versus Canephora without GI, Conilon from Espírito Santo versus Canephora without GI, and Robusta Amazônico and Conilon from Espírito Santo.

	PLS-score-based traditional version					LDA-based hard version					QDA-based soft version				
(a) Multi-class PLS-DA	Class 1	Class 2	Class 3	Class 4	Class 5	Class 1	Class 2	Class 3	Class 4	Class 5	Class 1	Class 2	Class 3	Class 4	Class 5
Nº LVs	18					18					18				
Sensitivity CAL	100.0	94.0	98.9	100.0	100.0	96.0	93.0	99.0	100.0	100.0	96.0	96.0	95.0	95.0	91.0
Specificity CAL	98.4	95.7	97.2	99.7	100.0	100.0	99.0	98.0	100.0	100.0	98.0	90.0	95.0	100.0	100.0
Sensitivity CV	97.3	89.0	92.6	100.0	100.0	89.0	86.0	95.0	100.0	100.0	91.0	92.0	88.0	91.0	89.0
Specificity CV	96.7	93.6	94.8	99.7	100.0	99.0	96.0	95.0	100.0	100.0	77.0	47.0	63.0	97.0	100.0
Sensitivity PRED	100.0	93.9	96.8	100.0	100.0	96.0	91.0	94.0	100.0	100.0	92.0	97.0	94.0	78.0	89.0
Specificity PRED	97.0	90.2	98.9	100.0	100.0	97.0	97.0	100.0	100.0	100.0	95.0	84.0	95.0	100.0	100.0
(b) GI classification	Model 1[a]			Model 2[b]			Model 3								
	R. A.	Without GI		C. from ES	Without GI		R. A.	C. from ES							
Nº LVs	5			4			13								
Sensitivity CAL	100.0	100.0		100.0	100.0		97.7	98.9							
Specificity CAL	100.0	100.0		100.0	100.0		98.9	97.7							
Sensitivity CV	100.0	98.2		98.9	100.0		94.8	90.5							
Specificity CV	98.2	100.0		100.0	98.9		90.5	94.8							
Sensitivity PRED	100.0	100.0		100.0	100.0		100.0	93.5							
Specificity PRED	100.0	100.0		100.0	100.0		93.5	100.0							

Robusta Amazônico from indigenous producers (class 1), Robusta Amazônico from non-indigenous Rondônia producers (class 2), Conilon from Espírito Santo (class 3), Conilon from Bahia (class 4) and specialty Arabica coffees (class 5). a R. A.is the acronym for Robusta Amazônico coffees from indigenous and traditional Rondônia producers; b C. from ES is the acronym for Conilon from Espírito Santo; Without GI is the acronym for Canephora coffees without GI.

differences, and multi-class PLS-DA proved their robustness to distinguish them.

Soft PLS-DA was worse than the traditional and hard versions. It showed satisfactory sensitivity and specificity for the classification of Robusta Amazônico samples (classes 1 and 2) and Conilon from Espírito Santo (class 3) in the prediction, but inferior results for Conilon from Bahia (class 4) and Arabica (class 5). The nature of this PLS-DA is similar to one-class classifiers (Pomerantsev and Rodionova, 2018). It was observed that many Robusta Amazônico samples (classes 1 and 2) and Conilon from Espírito Santo (class 3) were simultaneously assigned to their classes. In contrast, Conilon from Bahia and Arabica were detected as not members of their classes – an ability of soft PLS-DA. Therefore, the characteristics of Conilon from Bahia and Arabica samples may have negatively affected their classifications.

As a result of this study, traditional PLS-DA showed the best results and was comparable to the hard version that was easier to implement. Both were ways to obtain suitable classification results. There is no criticism for using PLS scores in traditional PLS-DA for classification since appropriate validation is carried out with a relevant test set (Pomerantsev and Rodionova, 2018), as was conducted in this study. Furthermore, the classification results obtained here were comparable and even superior to those of previous studies with coffee analysis (Baqueta et al., 2021; Dias et al., 2018; Robert et al., 2022).

3.3.2. Classifications of Canephora with and without GI

The purpose was to differentiate GI and non-GI Canephora using binary PLS-DA models based on PLS scores, the version chosen based on the previous results. Firstly, Robusta Amazônico was differentiated from the Canephora without GI (model 1). Secondly, Conilon from Espírito Santo was distinguished from the Canephora without GI (model 2).

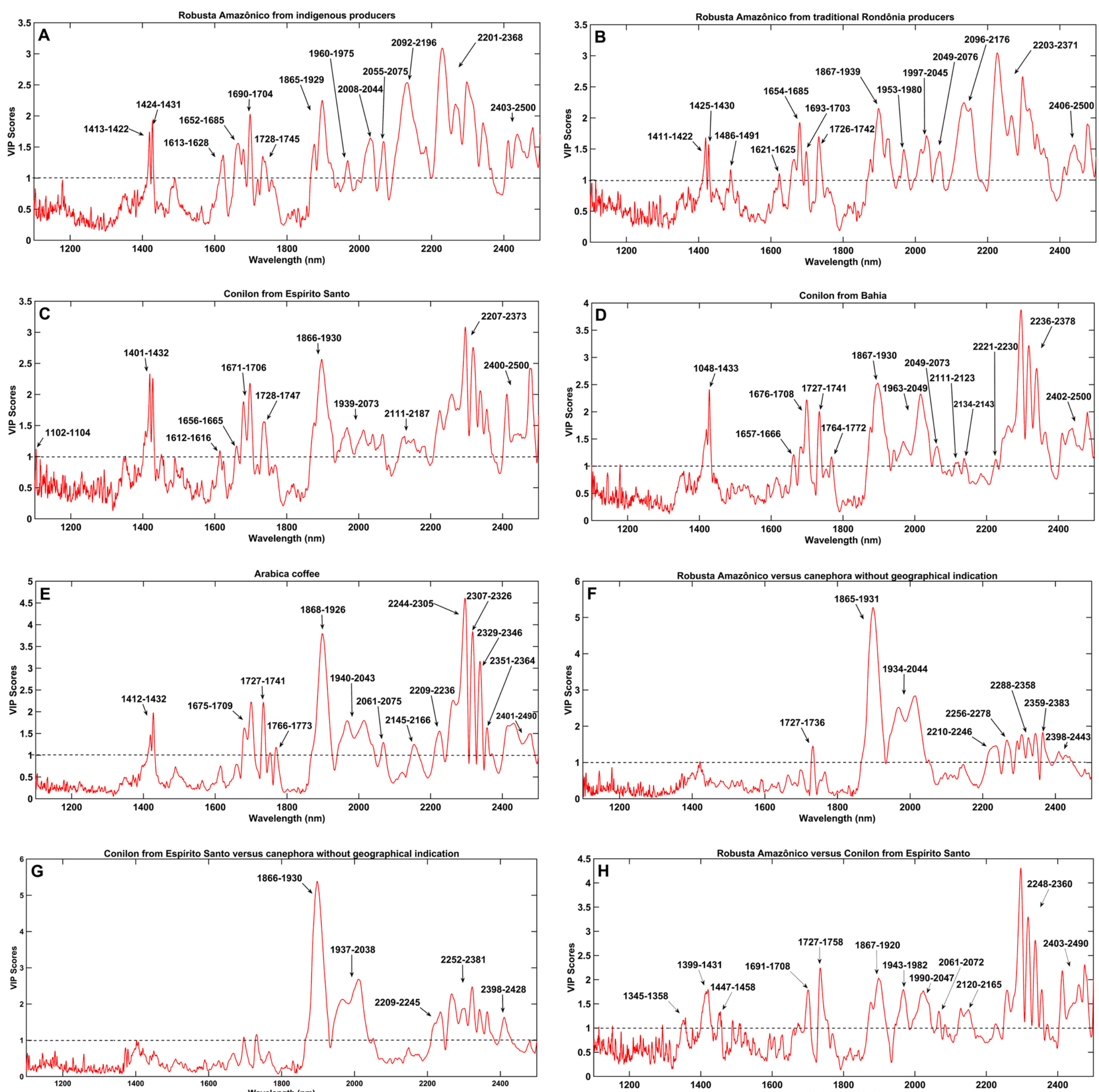

Fig. 4. Chemical interpretation of the PLS-DA models with attribution of NIR-VIP-bands. Traditional multi-class PLS-DA (A – class 1), (B – class 2), (C – class 3), (D – class 4), (E – class 5); Binary PLS-DA models to differentiate Robusta Amazônico versus Canephora without GI (E), Conilon from Espírito Santo versus Canephora without GI (F) and Robusta Amazônico and Conilon from Espírito Santo (H).

Finally, it was tried to differentiate Robusta Amazônico and Conilon from Espírito Santo (both GI Canephora) in a third model (model 3). Robusta Amazônico samples from indigenous and non-indigenous Rondônia producers were grouped into a single class in these models since they had the same GI. Their classification results are shown in Table 2b. Sensitivity and specificity obtained to differentiate Robusta Amazônico (model 1) and Conilon from Espírito Santo (model 2) of non-GI Canephora were perfect in the prediction = 100%. This result pointed out that protection under a GI and its unique production mode influenced the spectra, creating a traceable NIR fingerprint. The third model to differentiate Robusta Amazônico and Conilon from Espírito Santo showed high sensitivity and specificity in the prediction, with percentages ranging from 93.50% to 100%. The misclassified samples indicated that some of them were quite similar and that the model was not able to fully differentiate them.

These models proved that GI Canephora had unique and distinctive characteristics conferred by the green coffee quality and their production systems, making them distinguishable from non-GI Canephora coffees. This result creates opportunities for Canephora producers to exploit the typicity and origin of their coffees as value attributes. The method brought a direct and easy-to-use method for verifying GI coffees. A NIR system instantly evaluates the sample and provides satisfactory classification results that could be applied to the quality control of coffee cooperatives, certification agencies, and associations involved in protecting certified coffees.

3.4. Chemical interpretation of the models by NIR-VIP-bands

NIR spectra are difficult to interpret because they do not have absorption bands with clear chemical assignments. VIP scores analysis was performed for the traditional multi-class PLS-DA model and the three binary models to identify the most discriminant variables (Fig. 4). The highly influential variables are those with VIP scores higher than 1, where this cut-off value is assumed to define which of them are significant. The most informative region exhibiting spectral variables with VIP scores higher than 1 was from 1400 nm onwards for all models. In the multi-class model, several NIR-VIP bands associated with the main coffee compounds were assigned and displayed in Fig. 4A-E. Robusta Amazônico from indigenous (Fig. 4A) and non-indigenous Rondônia producers (Fig. 4B) presented a very similar VIP scores pattern without considerable differences in absorption bands, indicating their chemical composition was quite similar. On the other hand, Conilon from Espírito Santo (Fig. 4C), Bahia (Fig. 4D), and Arabica coffees (Fig. 4E) were more distinguishable.

Several compounds and some unique molecules in coffee show vibration in the NIR region, but the combined absorption of carbohydrates, chlorogenic acids, lipids, caffeine, and proteins is the most important. All these compounds present higher absorption bands from 1400 to 2500 nm. Absorptions around 1400 nm are associated with carbohydrates, chlorogenic acids, and lipids in coffee. The region between 1600 and 1800 nm showed that the variables in this region provided important information to distinguish the samples. This region is mainly associated with caffeine absorption, but carbohydrates can also show absorption bands in this region. In particular, the variables in the range 2000–2500 nm were highlighted as relevant predictors and associated with the combination bands of NH, OH, and C═O bonds. Carbohydrates, lipids, proteins, chlorogenic acids, caffeine, R–OH, and R–NH vibrations were related to higher absorbance in this range previously (Baqueta et al., 2021; Barbin et al., 2014; Pires et al., 2021; Ribeiro et al., 2011).

The discriminant variables seemed the same in the binary models when distinguishing Robusta Amazônico (Fig. 4F) and Conilon from Espírito Santo (Fig. 4G) of non-GI Canephora. NIR-VIP-bands between 1800 and 2100 nm and 2200 and 2400 nm were key to differentiate GI Canephora versus non-GI Canephora with carbohydrates, chlorogenic acids, lipids, caffeine, and proteins exhibiting intense absorption bands at these wavelengths (Barbin et al., 2014; Ribeiro et al., 2011). Fig. 4H shows the VIP scores that differentiate Robusta Amazônico versus Conilon from Espírito Santo, and several absorption bands related to main coffee compounds were assigned. However, Canephora coffee has been gaining consideration, especially after its beverage offers more sweetness and diversity of nuances (Lemos et al., 2020; Souza et al., 2021, 2018). Therefore, the absorption bands associated with sugars are of particular interest.

4. Conclusions

In this work, integrated NIR spectroscopy with chemometrics has been relevant for rapid Brazilian Canephora discrimination in roasted and ground form. NIR fingerprints obtained directly from solid coffees carried information about their origin, quality, species, and variety. An unsupervised pattern recognition chemometric strategy by applying PCA was powerful in providing a global understanding about the samples, pointing out their spectrochemical differences. It was crucial in extracting information from the coffees even after roasting, indicating that they carried chemical information associated with their initial green coffee compositions and had a unique fingerprint for spectroscopic characterization and traceability.

By applying the multi-class PLS-DA models in the first part of the study, they showed promising results for classifying coffee according to origin, quality, species, and variety. Hard PLS-DA offered fast classification without a large computational effort for implementation; however, it was not possible to interpret the discriminant variables that may be of interest in many cases, as when using traditional PLS-DA coupled with VIP scores. Regarding binary PLS-DA models, they allowed 100% classification of GI Canephora versus non-GI Canephora. Sensitivity and specificity obtained in all models indicated that they were robust, considering that this was a real-world application with a considerable volume of samples. Carbohydrates, chlorogenic acids, lipids, caffeine, and proteins were dominant absorption bands in coffee classifications. These metabolites are of particular interest because they are involved in the regulation of coffee quality and flavor.

A NIR system would be an advantageous alternative for obtaining a large-scale fingerprint of Canephora coffees in an origin certification procedure or specialized GI inspection. It allows users to continuously analyze multiple samples directly on the instrument, improving productivity with fast and reproducible sampling. It could be applied in quality control for coffee cooperatives, certification agencies, and associations that protect certified coffees from disposing of a routine, direct, relatively low-cost, sensitive, and easy-to-use method.

CRediT authorship contribution statement

Michel Rocha Baqueta: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. **Enrique Anastácio Alves**: Data curation, Investigation, Methodology. **Patrícia Valderrama**: Conceptualization, Supervision, Formal analysis, Investigation, Methodology, Data curation, Review & editing. **Juliana Azevedo Lima Pallone**: Conceptualization, Supervision, Funding acquisition, Investigation, Methodology, Data curation, Review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported by the São Paulo Research Foundation (FAPESP) (grant #2019/21062-0 and process #2022/03268-3), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (process 306606/2020-8 and process 310982/2020-0), and Agência Brasileira de Desenvolbimento Industrial (process 23200.19/0070-2-01 and process 30.20.90.027.00.00). The authors would like to thank the EMBRAPA Rondônia, cooperatives and coffee producers for their invaluable help with the acquisition of the samples for the study.

References

Alves, E.A., Souza, C.A., de, Rocha, R.B., Pereira, L.L., Lima, P.P., de, Lourenço, J.L.R., 2020. Efeito da fermentação sobre qualidade da bebida do café robusta (Coffea canephora) cultivado na amazônia ocidental. Rev. Ifes Ciênc. 6, 159–170. https://doi.org/10.36524/ric.v6i3.875.

Baqueta, M.R., Coqueiro, A., Março, P.H., Valderrama, P., 2021. Multivariate classification for the direct determination of cup profile in coffee blends via handheld near-infrared spectroscopy. Talanta 222, 121526. https://doi.org/10.1016/j.talanta.2020.121526.

Barbin, D.F., Felicio, A.L., de, 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 Res. Int. 61, 23–32. https://doi.org/10.1016/J.FOODRES.2014.01.005.

Brazil, 2021a. CONAB. Companhia Nacional de Abastecimento. Acompanhamento da safra brasileira: café - Safra 2021 - Quarto levantamento: 8(4) [WWW Document]. URL ⟨https://www.conab.gov.br/info-agro/safras/cafe/boletim-da-safra-de-cafe⟩.

Brazil, 2021b. INPI. Instituto Nacional da Propriedade Industrial. Indicações Geográficas.

Brazil, 2021c. INPI. Instituto Nacional da Propriedade Industrial. Espírito Santo é reconhecido como indicação de procedência para o café conilon [WWW Document]. URL ⟨https://www.gov.br/inpi/pt-br/central-de-conteudo/noticias/espirito-santo-e-reconhecido-como-indicacao-de-procedencia-para-o-cafe-conilon⟩.

Brazil, 2021d. INPI. Instituto Nacional da Propriedade Industrial. Matas de Rondônia é a mais nova Denominação de Origem para café [WWW Document]. URL ⟨https://www.gov.br/inpi/pt-br/central-de-conteudo/noticias/matas-de-rondonia-e-a-mais-nova-denominacao-de-origem-para-cafe⟩.

Brazil, 2021e. Brazilian coffees with geographical indication.

Brazil, 2021f. Empresa Brasileira de Pesquisa Agropecuária. Cafeicultura da Amazônia recebe primeira Denominação de Origem para cafés canéforas sustentáveis do mundo [WWW Document]. URL ⟨https://www.embrapa.br/en/busca-de-noticias/-/noticia/62622381/cafeicultura-da-amazonia-recebe-primeira-denominacao-de-origem-para-cafes-caneforas-sustentaveis-do-mundo#:~:text=Indicação⟩ Geográfica Matas de Rondônia,da Amazônia e do mundo.

Brazil, 2019. Brazilian Agricultural Research Corporation - Embrapa. Pesquisa ajuda indígenas a produzir café com qualidade [WWW Document]. URL ⟨https://www.embrapa.br/en/busca-de-noticias/-/noticia/42727299/pesquisa-ajuda-indigenas-a-produzir-cafe-com-qualidade⟩.

Correia, R.M., Andrade, R., Tosato, F., Nascimento, M.T., Pereira, L.L., Araújo, J.B.S., Pinto, F.E., Endringer, D.C., Padovan, M.P., Castro, E.V.R., Partelli, F.L., Filgueiras, P.R., Lacerda, V., Romão, W., 2020. Analysis of Robusta coffee cultivated in agroforestry systems (AFS) by ESI-FT-ICR MS and portable NIR associated with sensory analysis. J. Food Compos. Anal. 94, 103637 https://doi.org/10.1016/J.JFCA.2020.103637.

Dalazen, J.R., Rocha, R.B., Pereira, L.L., Alves, E.A., Espindula, M.C., de Souza, C.A., 2020. Beverage quality of most cultivated Coffea canephora clones in the western amazon. Coffee Sci. 15, 1–10. https://doi.org/10.25186/.v15i.1711.

de Almeida, M.M.C., Francisco, C.R.L., de Oliveira, A., de Campos, S.S., Bilck, A.P., Fuchs, R.H.B., Gonçalves, O.H., Velderrama, P., Genena, A.K., Leimann, F.V., 2018. Textural, color, hygroscopic, lipid oxidation, and sensory properties of cookies containing free and microencapsulated chia oil. Food Bioprocess Technol. 11, 926–939. https://doi.org/10.1007/s11947-018-2057-x.

Dias, R.C.E., Valderrama, P., Março, P.H., dos Santos Scholz, M.B., Edelmann, M., Yeretzian, C., 2018. Quantitative assessment of specific defects in roasted ground coffee via infrared-photoacoustic spectroscopy. Food Chem. 255, 132–138. https://doi.org/10.1016/J.FOODCHEM.2018.02.076.

dos Santos, V.J., Baqueta, M.R., Março, P.H., Valderrama, P., Visentainer, J.V., 2021. Human milk lactation phases evaluation through handheld near-infrared spectroscopy and multivariate classification. Food Anal. Methods 14, 873–882. https://doi.org/10.1007/s12161-020-01924-y.

Faria, L.De, Emilly, S., Alkimim, R., Ricardo, P., Marques, R., Grespan, B., Ana, L., Andrade, C., Ruane, S., Tiago, S., Sousa, V., Nascimento, M., Teixeira, E., 2022. Genome - wide association study of plant architecture and diseases resistance in Coffea canephora. Euphytica 1–13. https://doi.org/10.1007/s10681-022-03042-8.

Fioresi, D.B., Pereira, L.L., Catarina da Silva Oliveira, E., Moreira, T.R., Ramos, A.C., 2021. Mid infrared spectroscopy for comparative analysis of fermented arabica and robusta coffee. Food Control 121, 107625. https://doi.org/10.1016/J.FOODCONT.2020.107625.

Foschi, M., Biancolillo, A., Vellozzi, S., Marini, F., D'Archivio, A.A., Boqué, R., 2021. Spectroscopic fingerprinting and chemometrics for the discrimination of Italian Emmer landraces. Chemom. Intell. Lab. Syst. 215, 104348 https://doi.org/10.1016/J.CHEMOLAB.2021.104348.

Geladi, P., MacDougall, D., Martens, H., 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. Appl. Spectrosc. 39, 491–500. https://doi.org/10.1366/0003702854248656.

ICO, 2021. International Coffee Organization. Coffee production report.

Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. Technometrics 11, 137–148.

Lemos, M.F., Perez, C., da Cunha, P.H.P., Filgueiras, P.R., Pereira, L.L., Almeida da Fonseca, A.F., Ifa, D.R., Scherer, R., 2020. Chemical and sensory profile of new genotypes of Brazilian Coffea canephora. Food Chem. 310. https://doi.org/10.1016/j.foodchem.2019.125850.

Machado, J.L., Tomaz, M.A., da Luz, J.M.R., Osório, V.M., Costa, A.V., Colodetti, T.V., Debona, D.G., Pereira, L.L., 2021. Evaluation of genetic divergence of coffee genotypes using the volatile compounds and sensory attributes profile. J. Food Sci. https://doi.org/10.1111/1750-3841.15986.

Moreira, I., Scarminio, I.S., 2013. Chemometric discrimination of genetically modified Coffea arabica cultivars using spectroscopic and chromatographic fingerprints. Talanta 107, 416–422. https://doi.org/10.1016/j.talanta.2013.01.053.

Oliveira, E.C., da, S., Filgueiras, P.R., Moreli, A.P., Oliveira, A.C., Venturim, L.H.C., Pereira, L.L., 2020a. Espectroscopia de infravermelho para estudo de café conilon fermentado. Braz. J. Dev. 6, 19248–19259. https://doi.org/10.34117/bjdv6n4-189.

Oliveira, E.C., da, S., Guarçoni, R.C., de Castro, E.V.R., de Castro, M.G., Pereira, L.L., 2020b. Chemical and sensory perception of robusta coffees under wet processing. Coffee Sci. 15, 1–8. https://doi.org/10.25186/.v15i.1672.

Oliveri, P., Downey, G., 2013. Discriminant and class-modelling chemometric techniques for food PDO verification. Comprehensive Analytical Chemistry, first ed. Elsevier B.V. https://doi.org/10.1016/B978-0-444-59562-1.00013-X

Pereira, L.L., Moreli, A.P., Moreira, T.R., Caten, C.S., Ten, Marcate, J.P.P., Debona, D.G., Guarçoni, R.C., 2019. Improvement of the quality of Brazilian Conilon through wet processing: a sensorial perspective. Agric. Sci. 10, 395–411. https://doi.org/10.4236/as.2019.103032.

Pires, F., de, 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 Agtron to predict roasting degrees in coffee beans and ground coffees. Food Chem. 365, 130471 https://doi.org/10.1016/J.FOODCHEM.2021.130471.

Pomerantsev, A.L., Rodionova, O.Y., 2018. Multiclass partial least squares discriminant analysis: taking the right way—a critical tutorial. J. Chemom. 32. https://doi.org/10.1002/CEM.3030.

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, 1352–1358. https://doi.org/10.1016/J.TALANTA.2010.11.001.

Ríos-Reina, R., Camiña, J.M., Callejón, R.M., Azcarate, S.M., 2021. Spectralprint techniques for wine and vinegar characterization, authentication and quality control: advances and projections. TrAC Trends Anal. Chem. 134, 116121 https://doi.org/10.1016/J.TRAC.2020.116121.

Robert, J.V., de Gois, J.S., Rocha, R.B., Luna, A.S., 2022. Direct solid sample analysis using synchronous fluorescence spectroscopy coupled with chemometric tools for the geographical discrimination of coffee samples. Food Chem. 371, 131063 https://doi.org/10.1016/J.FOODCHEM.2021.131063.

Rodionova, O.Y., Pomerantsev, A.L., 2020. Chemometric tools for food fraud detection: the role of target class in non-targeted analysis. Food Chem. 317, 126448 https://doi.org/10.1016/J.FOODCHEM.2020.126448.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639. https://doi.org/10.1021/ac60214a047.

Souza, C.A. de, Alves, E.A., Rocha, R.B., Espindula, M.C., Teixeira, A.L., 2021. Perfis sensoriais dos cafeeiros cultivados na Amazônia Ocidental. In: Partelli, F.ábio Luiz, Pereira, Lucas Louzada (Eds.), Café Conilon: Conilon e Robusta No Brasil e No Mundo. Universidade Federal do Espírito Santo, pp. 187–198.

Souza, C.A., de, Rocha, R.B., Alves, E.A., Teixeira, A.L., Dalazen, J.R., da Fonseca, A.F.A., 2018. Characterization of beverage quality in Coffea canephora Pierre ex A. Froehner. Coffee Sci. 13, 210–218.

Teixeira, A.L., Rocha, R.B., Espindula, M.C., Ramalho, A.R., Júnior, J.R.V., Alves, E.A., Lunz, A.M.P., Souza, F., de, F., Costa, J.N.M., Fernandes, C., de, F., 2020. Amazonian robustas-new Coffea canephora coffee cultivars for the western brazilian amazon. Crop Breed. Appl. Biotechnol. 20, 1–5. https://doi.org/10.1590/1984-70332020v20n3c53.

Zontov, Y.V., Rodionova, O.Y., Kucheryavskiy, S.V., Pomerantsev, A.L., 2020. PLS-DA – a MATLAB GUI tool for hard and soft approaches to partial least squares discriminant analysis. Chemom. Intell. Lab. Syst. 203, 104064 https://doi.org/10.1016/J.CHEMOLAB.2020.104064.