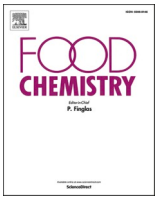
Food Chemistry 463 (2025) 141485

Contents lists available at ScienceDirect

Food Chemistry

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



Multiplatform Path-ComDim study of Capixaba, indigenous and non-indigenous Amazonian Canephora coffees

Michel Rocha Baqueta [a,b], Douglas N. Rutledge [c,d,*], Enrique Anastácio Alves [e], Manuela Mandrone [f], Ferruccio Poli [f], Aline Coqueiro [g], Augusto Cesar Costa-Santos [a], Ana Paula Rebellato [a], Gisele Marcondes Luz [a], Bruno Henrique Fermino Goulart [h], Eduardo Jorge Pilau [h], Juliana Azevedo Lima Pallone [a,*], Patrícia Valderrama [c,d,i,*]

[a] Department of Food Science and Nutrition, School of Food Engineering, Universidade Estadual de Campinas – UNICAMP, Campinas, São Paulo, Brazil
[b] Department of Chemistry, University of Rome "La Sapienza", Piazzale Aldo Moro 5, 00185 Rome, Italy
[c] Muséum national d'Histoire naturelle, MCAM, UMR7245 CNRS, Paris, France
[d] Faculté de Pharmacie, Université Paris-Saclay, Orsay, France
[e] Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA Rondônia, Porto Velho, Rondônia, Brazil
[f] University of Bologna, Department of Pharmacy and Biotechnology (FaBiT), Bologna, Italy
[g] Department of Chemistry, Federal University of Technology – Paraná (UTFPR), Ponta Grossa, PR, 84017-220, Brazil
[h] Chemistry Department, State University of Maringá (UEM), 87020-900, Maringá, Paraná, Brazil
[i] Universidade Tecnológica Federal do Paraná – UTFPR, 87301-899, Campo Mourão, Paraná, Brazil

ARTICLE INFO

Keywords:
Conilon
Robusta
Data integration
Sensory
Fingerprinting techniques

ABSTRACT

Integrating diverse measurement platforms can yield profound insights. This study examined Brazilian Canephora coffees from Rondônia (Western Amazon) and Espírito Santo (southeast), hypothesizing that geographical and climatic differences along botanical varieties significantly impact coffee characteristics. To test this, capixaba, indigenous, and non-indigenous Amazonian canephora coffees were analyzed using nine distinct platforms, including both spectroscopic techniques and sensory evaluations, to obtain results that are more informative and complementary than conventional single-method analyses. By applying multi-block Path-ComDim analysis to the multiple data sets, we uncovered crucial correlations between instrumental and sensory measurements. This integrated approach not only confirmed the hypothesis but also demonstrated that combining multiple data sets provides a more nuanced understanding of coffee profiles than traditional single-method analyses. The results underscore the value of multiplatform approaches in enhancing coffee quality evaluation, offering a more detailed and comprehensive view of coffee characteristics that can drive future research and improve industry standards.

1. Introduction

Canephora (*Coffea canephora*) production and consumption are opening new frontiers in coffee research, yet knowledge remains limited. Brazil, a key player in promoting this species, cultivates two distinct botanical varieties in separate regions: conilon Capixaba in Espírito Santo state (southeast) and Amazonian robusta in Rondônia state (north), in Amazon region. In Rondônia, Amazonian robusta is produced by both indigenous and non-indigenous farmers, with the indigenous crops following more sustainable agricultural practices in a delimited region of origin (Baqueta et al., 2024). Despite the growing interest, significant challenges remain in understanding how geographical, climatic, and agricultural practices affect its chemical and sensory profiles. It is critical to address these knowledge gaps and provide valuable insights into the interplay between region, variety and sustainability that could enhance the appreciation, cultivation and

Abbreviations: NIR, Near Infrared; UV–Vis, Ultraviolet-Visible; ^{1}H NMR, Proton Nuclear Magnetic Resonance; FAAS, Flame Atomic Absorption Spectroscopy; UHPLC-ESI-MS-QTOF, UltraHigh Performance Liquid Chromatography-Electrospray Ionization Quadropole Time of Flight Mass Spectroscopy; CCSWA, Common Component and Specific Weight Analysis; ComDim, Common Dimensions; Path-ComDim, Path modeling Common Dimensions.

* Corresponding authors at: Muséum national d'Histoire naturelle, MCAM, UMR7245 CNRS, Paris, France.
E-mail addresses: douglas.neil.rutledge@gmail.com (D.N. Rutledge), jpallone@unicamp.br (J.A.L. Pallone), patriciav@utfpr.edu.br (P. Valderrama).

https://doi.org/10.1016/j.foodchem.2024.141485
Received 15 February 2024; Received in revised form 20 September 2024; Accepted 28 September 2024
Available online 2 October 2024

market value of Brazilian Canephora coffees (Baqueta, Diniz et al., 2024).

Existing literature in Canephora coffee have focused on individual aspects, such as chemical composition, sensory characteristics, or cultivation practices, using isolated analytical techniques like molecular, electronic, or atomic spectroscopy, as well as separation methods such as liquid and gas chromatography. These studies typically analyze individual datasets and apply basic multivariate methods, such as principal component analysis (PCA). A range of traditional single-method analyses have been employed in Canephora research, including various analytical approaches (Baqueta et al., **2024**; Brand, Silva, Garrett, & Rezende, **2024**; de Souza Costa et al., **2024**; do Rosário, da Silva Mutz, Vieira, Schwan, & Bernardes, **2024**; Lemos et al., **2020**; Luna, da Silva, da Silva, Lima, & de Gois, **2019**; Moraes-Neto et al., **2024**; Robert, de Gois, Rocha, & Luna, **2022**; Viencz et al., **2023**; Zani Agnoletti et al., **2023**, Zani Agnoletti et al., **2022**). By using multiplatform approach, it is possible to connect information about the samples, such as agricultural practices, geographical origin and botanical variety with chemical and sensory traits, offering novel insights into how they influence coffee quality. This aspect is little underexplored in existing literature, adding a new dimension to coffee research.

Previous research in coffee science often applies traditional statistical methods that are limited in their ability to identify complex interactions between chemical and sensory variables. While recent publications reflect growing interest in Canephora, there remains a significant gap in comprehensive multiplatform analyses that integrate both chemical and sensory data, providing a holistic understanding of this coffee species. Some previous examples involved the use of the original ComDim algorithm (originally called Common Components and Specific Weights Analysis - CCSWA) (Hanafi & Qannari, **2008**; Qannari, Wakeling, Courcoux, & MacFie, **2000**) in Canephora multiplatform characterization with data coming from molecular spectroscopy (Baqueta, Valderrama, Alves, Pallone, & Marini, **2023**), atomic and molecular spectroscopy (Baqueta et al., **2023**), and also from molecular spectroscopy, mass spectrometry and sensory analysis (Baqueta, Marini et al., **2024**). ComDim method determines a common space for all the data tables, with each matrix having a specific contribution ("salience") to the definition of the orthogonal directions of this common space. The coordinates of the observations on the ComDim directions (Common Components) are the 'Global Scores' and the contributions of the variables within each of the tables are the 'Loadings'.

This classical ComDim method can show the contribution of each table to the Common Components but does not give any direct information about the relationships between tables. To do this, ComDim was extended to a path modeling approach, called Path-ComDim (Cariou, Qannari, Rutledge, & Vigneau, **2018**). Path-ComDim is a novel multiblock chemometric method that allows for simultaneous analysis of multiple data sources. This can provide a more holistic understanding of the direct relationships between various factors influencing Canephora coffee quality, which the original ComDim could not reveal. Path-ComDim is able to highlight connections between chemical profiles, sensory properties, and environmental factors, offering a deeper understanding of how these components interact. This novel approach allows for the detection of subtle yet significant correlations that have been overlooked in prior research.

The use of Path-ComDim in food analytical chemistry is relatively new and has not been extensively applied to coffee research. Studies that have applied multi-block analysis methods are rare, and even fewer have used the original ComDim to integrate such a broad range of analytical techniques. This is one of the first studies to apply Path-ComDim specifically to Canephora coffee, making it pioneering in the field. Applying this method offers new analytical depth and can serve as a model for future studies in coffee science. This study applies the novel Path-ComDim method to analyze multi-block datasets from advanced spectroscopy and sensory platforms, offering unprecedented insights into the relationships between chemical composition, sensory characteristics, and cultivation practices, with a focus on Brazilian Canpehora producers. This holistic approach reveals new dimensions of Canephora coffee quality, setting it apart from previous studies that focused on isolated traits.

In this context, the objective of this research was to apply the multiblock data analysis method Path-ComDim to evaluate the relationships between the multiple data blocks acquired on conilon Capixaba and indigenous and non-indigenous Amazonian robusta coffees from Brazil. In addition to sensory analysis, spectral responses were measured using Near and Mid-infrared (NIR and MIR, respectively), Ultraviolet-Visible (UV–Vis), Nuclear Magnetic Resonance (^{1}H NMR), UltraHigh Performance Liquid Chromatography-Electrospray Ionization Quadropole Time of Flight Mass Spectroscopy (UHPLC-ESI-MS-QTOF), and Flame Atomic Absorption Spectroscopy (FAAS). Canephora coffees have not been as well studied chemically and sensorially as Arabica coffees. There is still much to be learned about their origin, botanical variety, and their chemical and sensory characteristics, especially those of special provenances cultivated in Brazil. Therefore, the amount of chemical and sensory information obtained by different techniques and methods in this study is exceptional, comprehensive, and challenging.

2. Materials and methods

2.1. Canephora coffee samples

A set consisting of a total of seventy-five Canephora samples from different Brazilian producers was obtained. In this data set, twenty-five samples were Conilon produced in the state of Espírito Santo in Brazil. The other fifty samples were from Rondônia state (Western Amazon), where half (25 samples) were Robusta Amazônico samples from the indigenous producers and the other half (25 samples) were Robusta Amazônico samples from the non-indigenous producers. The samples had their authenticity guaranteed by EMBRAPA Rondônia, which donated the samples. The coffee fruits were harvested from the 2020 crop, ensuring that at least **80 %** were ripe cherries. The harvested cherries were carefully selected and thoroughly washed to remove any floating fruit and impurities such as leaves, stones, sticks and earth. Each sample of coffee beans came from a different producer, each employing their own specific processing methods, which we did not detail as our focus was on representing the major producers. Despite the variations in post-harvest processing - whether natural, fermented, or involving natural or artificial drying - the green coffee beans consistently had a moisture content of **11 %–12 %**. They were mechanically peeled, sorted through sieves of size 15 and larger, and then stored in paper packaging.

The processed green beans were roasted to a medium degree in a Probat sample roaster according to the Uganda Coffee Development Authority protocol (UCDA, **2010**). The initial temperature was 160 °C and 190 °C at the end, with a time ranging from 7:30 min to 9 min. The roasted samples were cooled, milled, and sieved through a 20-mesh sieve for particle size standardization for the chemical analyses. For sensory analysis, the roasted coffee was prepared following the cupping protocol (UCDA, **2010**). Fig. 1 schematically shows the samples and their geographical identities.

2.2. Instrumental analyses

2.2.1. UV–vis spectroscopy

An aqueous extract of the roasted ground coffee powder samples was obtained for UV–Vis analysis. The coffee extracts were prepared following the proportion of 1.0 g of roasted and ground coffee for 10 mL of distilled water at 92 °C. Each sample was weighed and placed in a glass beaker and then distilled water was added to the coffee powder, followed by manual shaking and resting for 10 **min**. After cooling them to room temperature (around 10 min), coffee extracts were filtered using a 25-mm pore-sized quantitative filter paper on an Erlenmeyer flask. The extracts were diluted in a 1:20 (mL:mL) proportion with distilled water

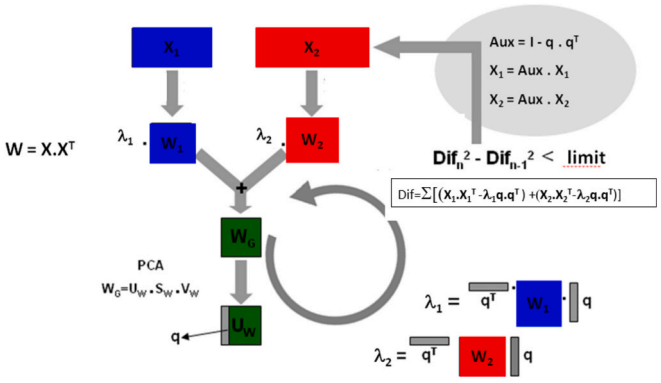


Fig. 1. Coffee samples and their geographical indication identity.

for the analysis.

Absorbance spectra were acquired at room temperature using an omega fluostar microplate reader (BMG Labtech, Ortenberg, Germany) equipped with a quartz microplate with optical path of 10 mm, and spectral resolution of 1 nm in the range of 200–1000 nm. Distilled water was used as blank.

2.2.2. Portable NIR spectroscopy

A MicroNIR spectrometer (microNIR™ 1700) from JDSU Uniphase Corporation (Milpitas, California, USA) analyzed directly the samples of ground roasted coffee powder. The conditions were: spectral region from 906 to 1676 nm, reflectance mode, 50 automatic scans, and 6.25 nm of resolution. Three aliquots of each sample were analyzed and then averaged. The blank was obtained using a standard NIR reflectance (Spectralon™).

2.2.3. Benchtop NIR spectroscopy

A Spectrum 100 N Fourier Transform NIR (Perkin Elmer, Waltham, Massachusetts, USA) spectrophotometer analyzed directly the roasted and ground coffee powder samples. The equipment was configured to reflectance, thirty-two scans per sample, and a resolution of four nm, operating in the region from 1000 to 2500 nm. Three different sample aliquots were taken and then the spectra averaged. The blank was evaluated using a NIR reflectance standard.

2.2.4. ATR-FTIR-MIR spectroscopy

An IRAffinity-1S spectrometer (Shimadzu, Kyoto, Japan) coupled with a horizontal ATR (attenuated total reflectance) accessory of zinc selenide (ZnSe) crystal at a 45° angle (PIKE Technologies, Madison, USA) was used. Coffee powder samples were directly analyzed between 4000 and 600 cm^{-1}, with a resolution of 4 cm^{-1} and 32 scans. The samples were analyzed in triplicate and then the spectra averaged. Background correction was performed using ambient air.

2.2.5. UHPLC-ESI-QTOF-MS

The samples were prepared by weighing 50 mg of each coffee in an Epperndorf and adding 200 µL of distilled water, 200 µL of methanol (Merck, Londrina, Paraná, Brazil) and 400 µL of chloroform (Merck, Londrina, Paraná, Brazil). Duplicate extraction was performed. The mixture was mixed in a vortex for 2 **min** and centrifuged for 5 **min**, separating the aqueous (upper) and organic (lower) phases. Afterward, 80 µL of each phase was filtered through a syringe filter (hydrophobic PTFE membrane, 0.22 µm, (Merck, Londrina, Paraná, Brazil) and pipetted into two separate vials for dilution. The aqueous extract was diluted with 1 mL of water/acetonitrile (1:1). In the vial with the organic phase, 1 mL of pure acetonitrile (Merck, Londrina, Paraná, Brazil) was added.

The extracts were analyzed by direct infusion by electrospray ionization (ESI) in a hybrid quadrupole time-of-flight high-resolution mass spectrometer (Q-TOF, Impact II, Bruker Daltonics Corporation, Bremen, Germany) operating in positive mode. The injection was carried out using an ultra-high-performance liquid chromatograph (Shimadzu, Nexera X2, Yokohama, Japan), but without the use of a separation column. For the aqueous extracts, a mobile phase consisting of water (solvent A) and acetonitrile (solvent B) was used, each added to 0.**1 %** formic acid (Merck, Londrina, Paraná, Brazil) with a constant %B of **50 %** for 1.5 **min**. The organic phase was also injected with water (solvent A) and acetonitrile (B), both with 0.**1 %** formic acid, with %B maintained at **98 %** also for 1.5 **min**. For both extracts, the data were obtained in a range of *m/z* 50 to 1300, with a flow rate of 0.2000 mL/min, an injection volume of 2 µL and an oven temperature of 30 °C.

2.2.6. FAAS

Calcium (Ca), magnesium (Mg), zinc (Zn), iron (Fe), manganese (Mn), copper (Cu) (LabSynth, Diadema, São Paulo, Brazil), and potassium (K) (SpecSol, Quimlab, Jacareí, São Paulo, Brazil) at 1000 mg L^{-1} were used as standard mineral solutions. Hydrogen peroxide **30 %** (*v*/v) (LabSynth, Diadema, São Paulo, Brazil), lanthanum oxide (Sigma Chemical Co., St. Louis, USA), commercial diluted nitric acid **50 %** (v/v) (Merck, Darmstadt, Germany), ultrapure water (Sartorious, Goettingen, Germany), acetylene gas (Messer Gases, São Paulo, Brazil), filter paper of 9 cm diameter (Nalgon, São Paulo, Brazil), ultrasonic bath (model 1400, Unique, Indaiatuba, São Paulo, Brazil), and digester block (model M242, Quimis, São Paulo, Brazil) were also used for the analysis.

The samples (0.6 g of roasted and ground coffee) were mineralized in an open system (digester block) using 6 mL of diluted nitric acid and 2 mL of hydrogen peroxide for four hours. Then, they were filtered and diluted with ultrapure water. The equipment was a PerkinElmer AAnalyst 200 (Norwalk, Connecticut, USA) equipped with a deuterium lamp for correction of the background radiation. The samples were nebulized and mixed with air-acetylene flame (2.5/10 l h^{-1}) at about 2000 °C. Hollow cathode lamps (Perkin Elmer, Norwalk, USA) for Fe – 248.3 nm, Ca – 422.67 nm, Cu – 324.75 nm, Mg – 279.48 nm, Mn – 285.21 nm, and Zn – 213,86 nm were used. The equipment was configured for atomic emission for the determination of K. Each sample was mineralized in triplicate and then analyzed.

The method was verified through some validation parameters. The relative standard deviation (RSD) values were below **10 %**. RSD values (%) were: 3.9 for Ca, 7.2 for Mg, 7.4 for Zn, 6.7 for Fe, 4.8 for Mn, 8.4 for Cu, and 8.6 for K. The correlation coefficients were always high: > 0.9997 for Ca; > 0.9996 for Mg; > 0.9991 for Zn; > 0.9996 for Fe; > 0.9995 for Mn; > 0.9999 for Cu; and > 0.9901 for K. The LOD (limit of detection) and LOQ (limit of quantification) (mg/100 g) for Ca were (LOD: 6.74, LOQ: 11.23); Mg (LOD: 24.63, LOQ: 41.06); Zn (LOD: 0.12, LOQ: 0.20); Fe (LOD: 0.63, LOQ: 1.06); Mn (LOD: 0.03, LOQ: 0.05); Cu (LOD: 0.17, LOQ: 0.28); K (LOD: 250.31, LOQ: 417.19). The recovery (%) was 105 for Ca, 104 for Mg, 86 for Zn, 84 for Fe, 85 for Mn, 83 for Cu, and 103 for K.

2.2.7. ^{1}H NMR spectroscopy

Coffee extracts were prepared using a previous protocol (Baqueta et al., **2021**). Deuterated solvents such as H_2O-d_2 and 3-(trimethylsilyl)-propionic-2,2,3,3-d4 acid sodium salt–TMSP were purchased from Eurisotop (Cambridge Isotope Laboratories, Inc., Tewksbury, Massachusetts, USA). The samples (0.1 g of roasted and ground coffee) were mixed with 1.5 mL of phosphate buffer (90 mM, pH 6.0) in H_2O-d_2 containing 0.**01 %** of TMSP as standard during one hour in a bath maintained to 90 °C. Extracted samples were centrifuged for 10 **min** (17,000 ×g) to obtain the supernatants to be analyzed.

A Varian 14.4 T NMR instrument (600.13 MHz operating at ^{1}H frequency, Agilent Technologies, Santa Clara, USA) equipped with a high-field triple resonance probe and H_2O-d_2 for the internal lock was used. The conditions were: spectra acquisition at 298 K, relaxation delay of 2.0 s, observed pulse of 5.80 µs, 256 scans, acquisition time of 16 min and spectral width of 16.00 ppm. Residual water signal at 4.83 ppm

(power = 22 Hz, presaturation delay = 2 s) was suppressed using a presaturation sequence. The free induction decays (FIDs) were Fourier transformed, and the resulting spectra were phased, baseline-corrected, and calibrated for TMSP at 0.00 ppm. The spectral intensities were reduced to integrated regions of equal width (0.04 ppm) corresponding to the interval of 0.00 to 10.00 ppm after normalization with respect to the signal of the standard at 0.00 ppm using the NMR MestReNova software (Mestrelab Research, Santiago de Compostela, Spain). The regions from 5.00 to 4.50 ppm were excluded from the analysis due to residual water signals. Chemical shifts, coupling constants, and data available in the literature on coffee (Coqueiro, Baqueta, Mandrone, Poli, & Valderrama, **2021**; Zani Agnoletti et al., **2022**) were considered to identify the metabolites.

2.2.8. Sensory analysis

The sensory analysis was conducted at the specialized coffee sensory laboratory of SINDICAFESP - Sindicato da Indústria de Café do Estado de São Paulo, located in São Paulo, Brazil. The analysis was carried out under standardized white lighting conditions. The cupping procedure followed the international Robusta Cupping Protocol established by the Coffee Quality Institute (CQI) (UCDA, **2010**). A panel of six trained and certified cuppers, all men, conducted the evaluations. The panel consisted of four Q-Graders, who were specialized in arabica coffee, and two R-Graders, who were specialized in robusta coffee. Each panelist holds certification from the Specialty Coffee Association. This diverse expertise and formal certification were employed to minimize subjectivity in the analysis and ensure a universally consistent evaluation language. The panelists' ages ranged from 25 to 60 years. To ensure reliable and consistent evaluations, all panelists were trained in accordance with the official protocol (UCDA, **2010**). The training encompassed detailed instruction on evaluation criteria, sensory attributes, and the application of the numerical scoring system, equipping the panelists with the necessary skills for accurate and objective assessments. Five cups of each coffee sample were prepared, following the ratio of 8.25 g of roasted and ground coffee, with granulometry between 70 and **75 %** of the particles passing through a sieve 20 mesh, to 150 mL of water, heated to 94–95 °C. The sensory attributes assessed included fragrance/aroma, flavor, aftertaste, salt/acidity, bitter/sweet, mouthfeel, uniform cups, balance, clean cups, and overall quality. Each attribute was rated on a numerical scale from 0 to 10 to quantify the intensity of each sensory attribute, where 0 denotes the complete absence of the attribute and 10 represents its maximum intensity. This sensory evaluation was conducted in compliance with ethical standards and was approved by Plataforma Brazil and the Research Ethics Council under protocol CAAE: 40806520.3.0000.5404. In addition, all data presented in this study were collected with the informed consent of the participants, and the participants were informed of how the data would be used.

2.3. Multi-block Path-ComDim data analysis

Before the multi-block data analysis, the data sets acquired by the different techniques/methods were organized into nine matrices or blocks (one per analytical technique or method) – UV–Vis (1), portable NIR (2), benchtop NIR (3), MIR (4), UHPLC-ESI-MS-QTOF of organic extracts (5), UHPLC-ESI-MS-QTOF of water extracts (6), FAAS (7), ^{1}H NMR (8), and sensory (9). The signals/results were imported into the Matlab R2019a environment (The Mathworks, Natick, Massachusetts, USA). The samples were sequentially organized from 1 to 75 in each matrix. From sample 1 to 25 were Robusta Amazônico samples from indigenous producers (Group 1), from sample 26 to 50 were Robusta Amazônico samples from non-indigenous producers (Group 2), and from 51 to 75 were Conilon samples from Espírito Santo (Group 3). Preprocessing techniques were applied on the different data blocks according to Table 1. The choice of the most suitable preprocessing for the data blocks was made based on previous experience with similar data sets.

Table 1
Preprocessing techniques applied on the different data blocks.

Technique	Preprocessing
UV-Vis	Baseline correction + Savitzky–Golay smoothing + mean centering
Portable NIR	Multiplicative scatter correction + Savitzky–Golay smoothing + mean centering
Benchtop NIR	Multiplicative scatter correction + Savitzky–Golay smoothing + mean centering
MIR	Spectrum cut + Multiplicative scatter correction + Savitzky–Golay smoothing + mean centering
UHPLC-ESI-MS-QTOF of organic extracts	log 10 + mean centering
UHPLC-ESI-MS-QTOF of water extracts	log 10 + mean centering
FAAS	Autoscaling
^{1}H NMR	Mean centering
Sensory	None

As data obtained by UV–Vis, portable NIR, benchtop NIR, MIR, UHPLC-ESI-MS-QTOF of organic and water extracts were composed of continuous variables in a large range, principal components analyses were applied separately to reduce their dimensionality and obtain their scores. The first 20 scores from these chemical analyses were used as input into the subsequent chemometric analysis instead of their preprocessed data.

In the original algorithm for ComDim analysis, each CC is the first normed scores vector of a weighted sum of scalar matrices calculated from all the data tables as shown in Fig. 2, in the simplest case of two centered and normalized data blocks, $\mathbf{X}_1$ and $\mathbf{X}_2$. A weighted sum $\mathbf{W}_G$ of the samples-based variance–covariance matrices, $\mathbf{Wi} = \mathbf{Xi} \times \mathbf{Xi}^T$, is calculated using an initial weighting, or salience, of $\lambda i = 1$ for all tables. The vector of scores of the first normed Principal Component is extracted from $\mathbf{W}_G$ as an initial estimate of the first Common Component (CC). The salience, λi, of each block $\mathbf{Wi}$ is then recalculated from these scores. The estimations of the Global Scores and saliences are optimized by iterative recalculation until convergence. Each original matrix $\mathbf{Xi}$ is then deflated, and the procedure is repeated for the calculation of the second CC, and so on.

In Path-ComDim, instead of doing a PCA on the sum of all the $\mathbf{Xi}*\mathbf{Xi}^T$ matrices, it is applied to the sum of the $(\mathbf{Xi}*\mathbf{Xi}^T) * (\mathbf{Xj}*\mathbf{Xj}^T)$ matrices where 'i' and 'j' correspond to those data tables whose connections are to be characterized. This Path-ComDim extension could highlight relationships between the responses from all the spectroscopy-based analytical techniques and between them and the sensory measurements. Since we do not have any prior knowledge as to which tables present interesting interactions, Path-ComDim was applied to all 'i' and 'j'. Fig. 3 shows the path diagram showing the interactions between the

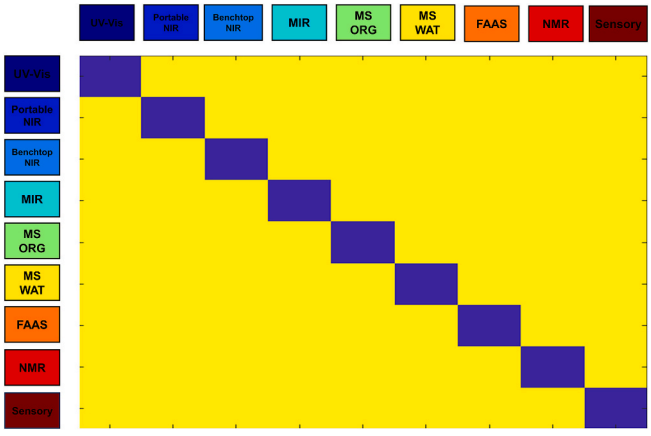


Fig. 2. Schema of the original ComDim algorithm in the case of two data blocks.

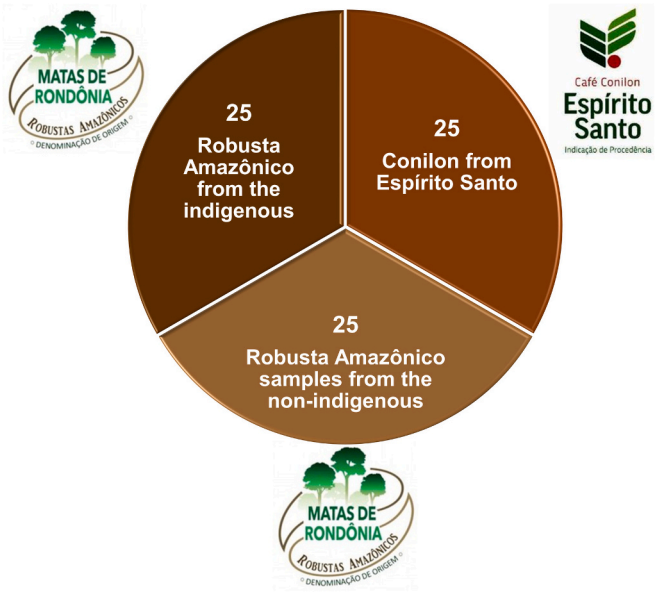


Fig. 3. Path diagram of the interactions between the nine data blocks (UV–Vis spectroscopy, portable NIR spectroscopy, benchtop NIR spectroscopy, MIR spectroscopy, MS ORG acronyms for UHPLC-ESI-MS-QTOF of organic extracts, MS WAT acronyms for UHPLC-ESI-MS-QTOF of water extracts, FAAS, ^{1}H NMR spectroscopy, and sensory). The yellow off-diagonals indicate that the corresponding pairs of tables will be included in the calculations. The blue diagonal indicates that none of the individual tables will be taken into account. **(For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)**

9 analytical techniques, including the sensory data.

Path-ComDim data analysis was then performed on the nine blocks of data, calculating common components (CCs) which could show the saliences/contributions of the different data blocks and relations between data blocks. The plot of the global scores made it possible to identify the informative CCs. Then, the blocks with important contributions to these informative CCs were identified by the salience plots. The corresponding loadings were extracted in order to interpret the variables. More details regarding the Path-ComDim algorithm and its implementation can be found in Cariou et al. (Cariou et al., **2018**). Additional details regarding the properties of ComDim, from which Path-ComDim is derived, can be found in El Ghaziri et al. (El Ghaziri, Cariou, Rutledge, & Qannari, **2016**).

3. Results and discussion

In a study with so many analytical techniques and methods performing chemical measurements and data generated from sensory analysis applied to characterize the same samples comprehensively, the strategy of analyzing the data jointly is ideal for identifying which of them are most informative. It helps to illustrate the specific importance of the chemical and sensory measurements obtained from different platforms and avoids interpreting data that is not relevant and, therefore, does not contribute to the investigation. Therefore, the nine multiblock datasets composed of UV–Vis (1), portable NIR (2), benchtop NIR (3), MIR (4), UHPLC-ESI-MS-QTOF of organic extracts (5), UHPLC-ESI-MS-QTOF of water extracts (6), FAAS (7), ^{1}H NMR (8), and sensory (9) were jointly analyzed by the Path-ComDim, assuming that the spectral techniques could be inter-related and that there is a link between them and also between them and the sensory parameters of the coffees. Connections identified between data blocks can enhance data interpretation and highlight a network of dependent relationships which can be interpreted often as causal effects.

The nine blocks of data contributed differently to the characterization of the samples, but not all of them were relevant considering the Path-ComDim. The Path-ComDim analysis revealed interesting separation of the groups of samples on the first and third common components, CC1 and CC3. To help to understand the results more easily it should be remembered that samples with negative scores are related to negative loadings, whereas samples with positive scores are related to positive loadings.

3.1. CC1 interpretation

Fig. 4 shows the parameters calculated by Path-ComDim for CC1. The first observed result is regarding CC1 saliences (Fig. 4A). Typically, the first CC has a higher importance compared to the subsequent CCs. In this case, it appears that the CC1 mainly highlights the link between UV–Vis, portable NIR, and benchtop NIR, and to a lesser extent benchtop NIR and MIR. This suggests that in addition to these techniques being important in characterizing the samples, there are correlations between the components associated with these three blocks. On the other hand, the data blocks of the organic and aqueous UHPLC-ESI-MS-QTOF extracts, FAAS, ^{1}H NMR and sensory do not seem to be strongly related to the other blocks in CC1.

Portable NIR spectroscopy has been shown to be an important block of information for differentiating between Amazonian Robusta from indigenous and non-indigenous producers, Conilon from Espírito Santo, and Conilon from Bahia, another Canephora producing state in Brazil (Baqueta, Valderrama, Mandrone, et al., **2023**). These results are also seen when applying the benchtop version of NIR for classification (Baqueta, Alves, Valderrama, & Pallone, **2023**). MIR spectroscopy (Baqueta, Valderrama, Mandrone, et al., **2023**) has been shown to be important in distinguishing factors as coffee species (Canephora and Arabica), botanical variety and geographical origin (Rondônia Robusta and Espírito Santo Conilon), and producers (indigenous and non-indigenous coffee producers from Rondônia). The separation of Rondônia Robusta and Espírito Santo Conilon using MIR spectroscopy has also been effective in another study (Zani Agnoletti et al., **2023**).

Fig. 4B illustrates the CC1 scores extracted from the first dimension of the Path-ComDim. CC1 scores clearly separates the indigenous samples (from sample 1 to 25 identified by the number 1) from the non-indigenous samples (from sample 26 to 50 identified by the number 2), and Conilon from Espírito Santo (from 51 to 75 identified by the number 3). The link between UV–Vis and NIR (both portable and benchtop instruments) and between benchtop NIR and MIR distinguished the indigenous samples through negative scores. Only two indigenous samples behaved differently, showing positive scores. Non-indigenous and Conilon from Espírito Santo samples presented positive scores in general. Four non-indigenous and nine Conilon samples from Espírito Santo showed a different trend from their group members. From a spectrochemical point of view, this indicates that some samples may be chemically similar even though they come from different botanical origins and varieties. Therefore, this observed result may be related to factors besides geographical origin and botanical variety that are known for the set of samples.

To better understand the results, we need to look together at the scores jointly with the respective loadings of the important data blocks. Fig. 4C-F shows the variables in each block important for CC1, therefore, they are loadings of the UV–Vis, portable NIR, and benchtop NIR spectra. UV–Vis loadings (Fig. 4C) exhibited peaks and troughs related to the absorbance of some coffee compounds. UV region (200–400 nm) was more informative than the visible one (400–800 nm). The regions between 257 and 275 nm and between 377 and 430 were important to non-indigenous samples and Conilon from Espírito Santo. The regions between 230 and 255 nm and between 276 and 372 nm were important to indigenous samples.

Trigonelline (272–275 nm), caffeine (276–280 nm), caffeic acid (320–325 nm), and melanoidins (400–405 nm) were previously identified in coffee UV–Vis spectra (Teixeira, Polari, Ferreira, Araújo, &

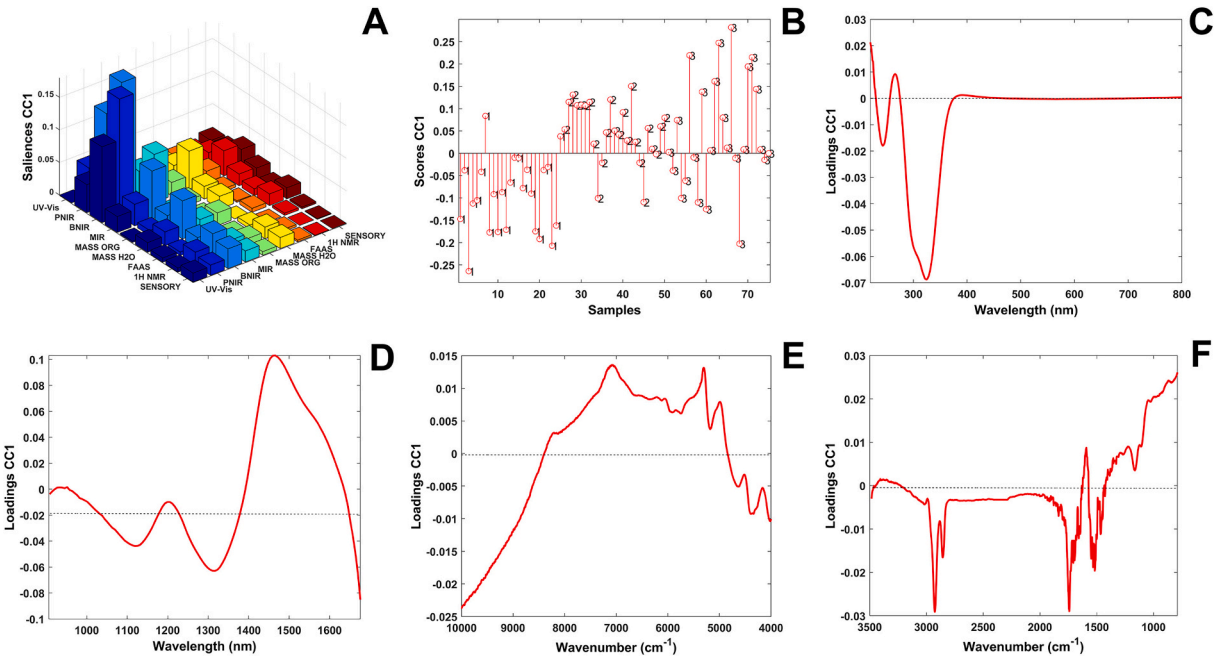


Fig. 4. First global component (CC1) of the Path-ComDim analysis showing the importance/saliences of the nine data blocks (A); Scores (B) with indigenous samples (from 1 to 25 identified by the number 1), from the non-indigenous samples (from 26 to 50 identified by the number 2), and Conilon from Espírito Santo (from 51 to 75 identified by the number 3); Loadings of the UV–Vis (C); Loadings of the portable NIR (D); Loadings of the benchtop NIR (E); Loadings of the MIR (F).

Cirino, **2015**). Some factors in this analytical technique may cause displacement on the maxima electronic absorptions of these compounds, differing slightly from what was observed previously. For example, caffeine absorbance was related to the wavelength 270 nm previously, while 310 nm was associated with the absorbance of caffeic acid and 350 nm with the absorbance of chlorogenic acids (Belay, Ture, Redi, & Asfaw, **2008**; Suhandy & Yulia, **2017**). A study has shown that the trough at 256 nm was related to the absorbance of vanillic acid (Suhandy & Yulia, **2017**). These results suggest that trigonelline and melanoidins could be relevant for the screening of non-indigenous samples and Conilon from Espírito Santo, while caffeine, caffeic acid, chlorogenic acids, and vanillic acid could be considered important for the identification of indigenous samples.

Portable NIR loadings (Fig. 4D) and benchtop NIR loadings (Fig. 4E) are the results of a number of absorptions of several coffee compounds. An intense absorption region of portable NIR (Fig. 4D) from 1400 to 1650 nm was related to non-indigenous samples and Conilon from Espírito Santo. Also, the other two less intense absorption bands, between 906 and 1150 nm and around 1200 nm, were important for these coffees. From another perspective, two troughs around 1110 and 1310 nm were related to indigenous samples. In the benchtop NIR loadings (Fig. 4E), a large region between 8000 and 5000 cm^{-1} was associated with non-indigenous samples and Conilon from Espírito Santo, while the region before 8000 cm^{-1} and after 5000 cm^{-1} were important to indigenous samples.

Caffeine, trigonelline, sugars, lipids, and chlorogenic acids present in the coffee beans absorb in almost the entire analyzed NIR region (906–1676 nm for portable NIR, and 10,000–4000 cm^{-1} for benchtop NIR) (Barbin, Felicio, Sun, Nixdorf & Hirooka, **2014**; Ribeiro, Ferreira, & J. G., **2011**). Both NIR loadings (portable and benchtop) exhibited a high absorption on the final portion of the spectra, starting from 1400 nm for the portable instrument and from 8000 cm^{-1} (or 1250 nm) for the benchtop instrument. Analytical standards of caffeine, chlorogenic acid, caffeic and ferulic acids, and theobromine analyzed by the NIR spectroscopy showed characteristic absorption bands in their spectra associated with coffee spectra (Baqueta, Valderrama, Alves, et al., **2023**). Absorptions around 1400 nm were previously associated with carbohydrates, chlorogenic acids, and lipids in coffee (Correia et al., **2018**). NIR absorption bands of pure caffeine were previously (Zhang et al., **2013**) identified in four regions of 4018–4196 cm^{-1}, 4412–5056 cm^{-1}, 5577–6105 cm^{-1}, and 6784–7706 cm^{-1} (or 1300–1470 nm, 1640–1790 nm, 1980–2270 nm, and 2380–2490 nm). Theobromine, sometimes present in specialty coffees, can be associated with wavenumbers at 7353 cm^{-1}, 5865 cm^{-1}, and 4383 cm^{-1} (Huck, Guggenbichler, & Bonn, **2005**). Between 1600 and 1800 nm, caffeine absorption is identified, but carbohydrates can also show absorption bands in this region. Variables in the range 2000–2500 nm were associated with the combination bands of NH, OH, and C═O bonds, which could be related to carbohydrates, lipids, proteins, chlorogenic acids, caffeine, R–OH, and R–NH vibrations. Other absorptions between 1800 and 2100 nm and between 2200 and 2400 nm could be linked to carbohydrates, chlorogenic acids, lipids, caffeine, and protein absorption bands (Baqueta, Coqueiro, Março, Mandrone, et al., **2021**; Baqueta, Coqueiro, Março, & Valderrama, **2021**).

Specific compounds are not as well identified by NIR as by MIR. In this sense, MIR loadings (Fig. 4F) exhibit more specific absorption bands. The spectral intervals between 3200 and 1630 cm^{-1} and between 1570 and 1430 cm^{-1} can be associated with indigenous samples because these variables were on the negative side. The dominant bands in 2924 and 2850 cm^{-1} are related to the stretching of CH bonds in the CH_2 and CH_3 groups of lipids and caffeine, respectively (Munyendo, Njoroge, & Hitzmann, **2022**). The third important signal was in 1745 cm^{-1} and corresponds to vibrations of the C═O group in triglyceride esters. The fourth signal, in 1643 cm^{-1}, is where the C═O stretching vibration in caffeine takes place (Craig, Botelho, Oliveira, & Franca, **2018**). Lipids and caffeine were identified as markers for distinguishing Espírito Santo Conilon and Rondônia Amazon Robusta, but it did not include Robusta produced by indigenous people (Zani Agnoletti et al., **2023**). Here the same compounds are found to distinguish between the indigenous samples from the rest. The positive peak between 1631 and 1570 cm^{-1} and the other positive loadings from 1420 to 800 cm^{-1} can be related to non-indigenous samples and Conilon from Espírito Santo. The region from 1130 to 950 cm^{-1} is highly influenced by the absorption of polysaccharides (galactans absorb at 1134 cm^{-1}, cellulose at 1059 cm^{-1} and

1033 cm^{-1}, and arabinogalactans at 1065 cm^{-1} and 1020 cm^{-1} or 1078 cm^{-1} and 1043 cm^{-1}), and it is difficult to attribute precise chemical assignments (Craig et al., **2018**). Additionally, chlorogenic acids exhibit strong absorption in the region of 1300–1150 cm^{-1} (Munyendo et al., **2022**).

3.2. CC3 interpretation

Although the first component (CC1) is often the most important to analyze in many methods from the multi-block family, the following common components can be informative too. This is the case with the third dimension, the CC3 (Fig. 5). The saliences of this component (Fig. 5A) highlight important links between sensory and essential elements (FAAS), as well as between sensory and ^{1}H NMR. Another result is the correlation between elemental and molecular composition as seen through the high salience for the link of FAAS and MIR (Fig. 5A). The relationship between sensory and molecular spectroscopy data was observed in a previous multi-block coffee study (Baqueta, Coqueiro, Março, Mandrone, et al., **2021**), with ^{1}H NMR and NIR showing comparable saliences with sensory data of commercial Brazilian coffees.

Regarding the CC3 scores (Fig. 5B), the coffee samples from Rondônia (both indigenous and non-indigenous producers) are separated from the samples from Espírito Santo. Rondônia samples (from 1 to 50, samples identified by 1 and 2) exhibited generally negative scores, while Espírito Santo samples (from 51 to 75, samples identified by 3) showed mostly positive scores. This analysis indicates a separation by geographical origin (different producing-states – Rondônia and Espírito Santo), but also by botanical origin (Robusta and Conilon).

Fig. 5C-F shows the loadings of the FAAS, ^{1}H NMR, MIR, and sensory analysis which are the important blocks for CC3. However, before considering these results, it is important to check the sensory profile of the coffees classified by the cuppers (Fig. 6). For this purpose, average

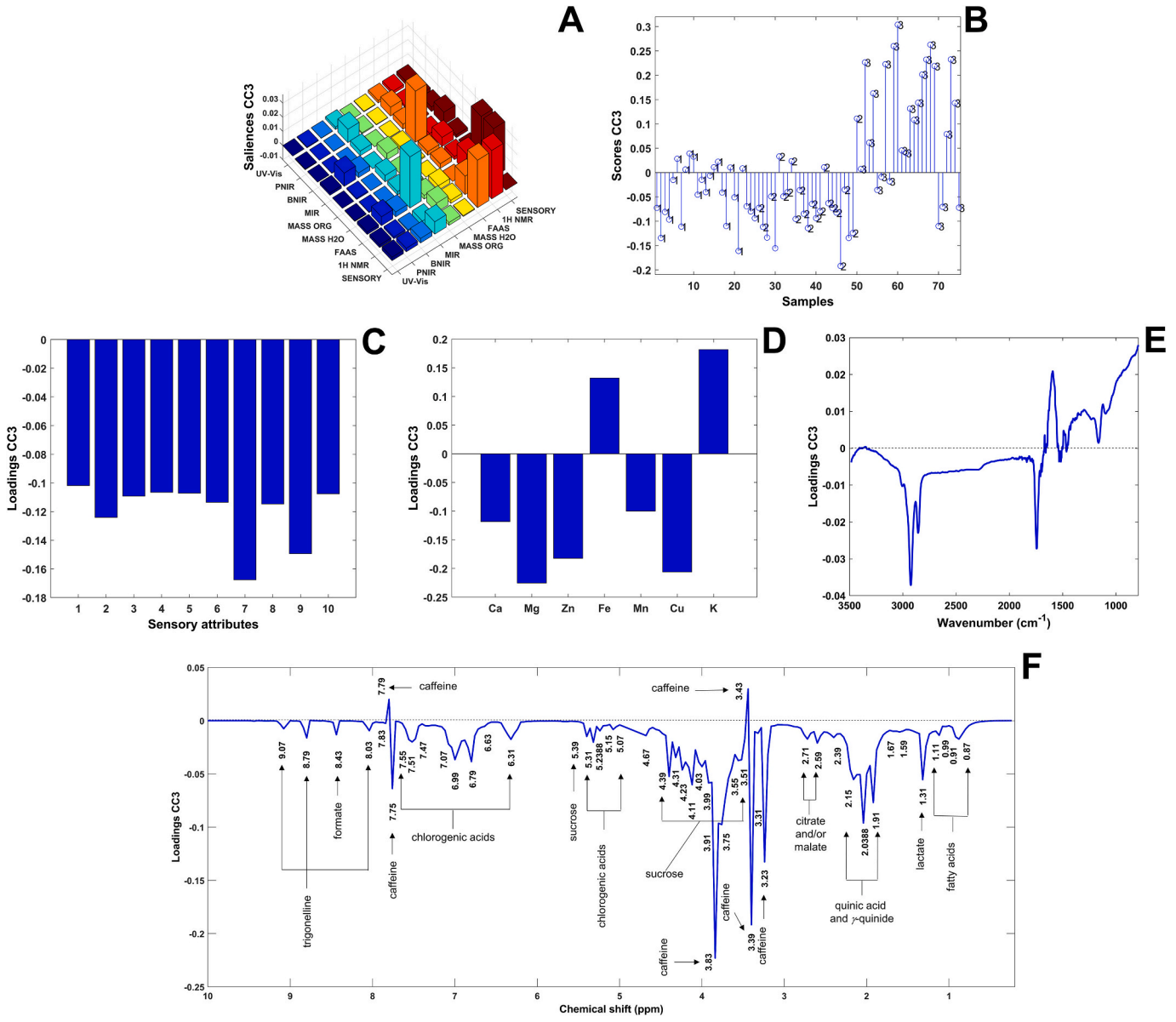


Fig. 5. Third global component (CC3) of the Path-ComDim analysis showing the importance/saliences of the nine data blocks (A); Scores (B) with indigenous samples (from 1 to 25 identified by the number 1), from the non-indigenous samples (from 26 to 50 identified by the number 2), and Conilon from Espírito Santo (from 51 to 75 identified by the number 3); Loadings of the sensory attributes (C), which were codified as (1) fragrance/aroma, (2) flavor, (3) aftertaste, (4) salt/acidity, (5) bitter/sweet, (6) mouthfeel, (7) uniform cups, (8) balance, (9) clean cups, and (10) overall quality; Loadings of the minerals (D); Loadings of the MIR (E); Loadings of the ^{1}H NMR (F).

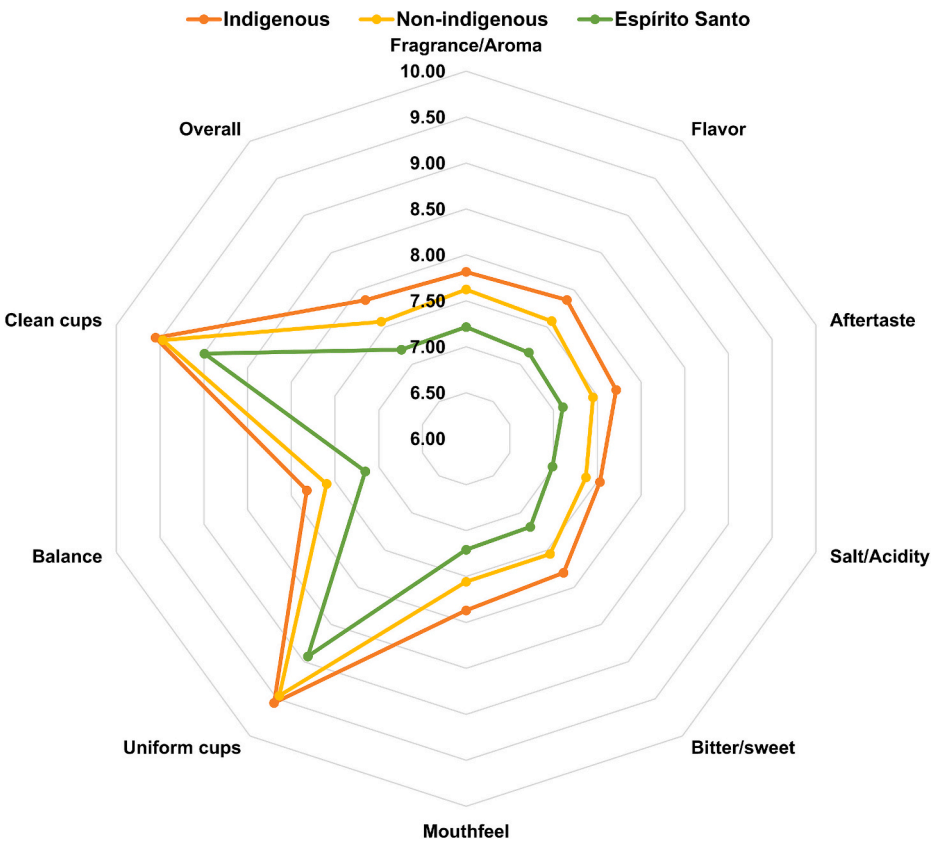


Fig. 6. Intensity of the sensory attributes of Canephora coffee samples evaluated through cupping.

results of each coffee group were considered and displayed in detail in Table 2. Most of the analyzed attributes received a score between 7.00 and 8.00, except for uniformity and clean cup, which ranged from 8.90 to 10.00. It was found that there is a decrease in sensory quality from indigenous Robusta to non-indigenous Robusta and then Conilon from Espírito Santo (Fig. 6 and/or Table 2).

To infer more information about the chemical aspects, we need to look again at the scores in CC3 (Fig. 5B) with their respective loadings. Beginning with sensory loadings (Fig. 5C), all sensory attributes had negative loadings and could be related to Robusta samples from Rondônia. Observing the loadings for essential minerals in CC3 (Fig. 5D), Conilon samples from Espírito Santo could be related with Fe and K, while Robusta samples from Rondônia could be related with Ca, Mg, Zn, Mn, and Cu. Considering only mineral analysis in a previous study (Baqueta, Costa-Santos, et al., **2024**), a generalized interpretation leads to the association of K with the Conilon from Espírito Santo and the other minerals (Ca, Mg, Zn, Mn, Cu, and Fe) with the Robusta coffees from Rondônia, both indigenous and non-indigenous producers.

Considering the MIR loadings (Fig. 5E), the spectral variables from 3500 to 1700 cm^{-1} can be associated with Robusta samples from Rondônia because they were on the negative side, while the region between 1670 and 800 cm^{-1} could be related to the Conilon samples from Espírito Santo. When looking at the peaks between 3500 and 1700 cm^{-1}, the wavenumbers at 3010, 2924, 2852, and 1745 cm^{-1} were important. Wavenumbers between 2840 **and** 2940 cm^{-1} could be associated with lipids and caffeine by the stretching bands of C—H bonds in CH_2 and CH_3 groups (Assis et al., **2019**; Munyendo et al., **2022**). Lipids also can exhibit absorption in 2922, 2852 and 1743 cm^{-1} due to asymmetric and symmetric stretching of CH_2, and stretching of C═O, respectively (Assis et al., **2019**). The wavenumber at 3008 cm^{-1} was also related to lipids (Munyendo et al., **2022**). Lipids and caffeine have been recently identified as markers for distinguishing Conilon coffees from Espírito Santo and Robusta from Rondônia using MIR spectroscopy (Zani Agnoletti et al., **2023**), indicating that the present results are in accordance with the literature. Observing the peaks from 1670 to 800 cm^{-1}, the wavenumbers at 1666, 1595, 1490, and 1120 cm^{-1} were highlighted. Absorption bands in these wavenumbers suggest the presence of chlorogenic acids, carbohydrates, lipids, trigonelline, and caffeine (Assis et al., **2019**; Munyendo et al., **2022**).

The last loadings to examine are those of the ^{1}H NMR block (Fig. 5F). A number of chemical shifts suggesting some important metabolites in coffee with negative loadings were related to Robusta samples from Rondônia (both indigenous and non-indigenous). Based on the literature, these chemical shifts could be attributed to molecules of

Table 2
Sensorial scores of Canephora coffee samples evaluated through cupping, expressed as mean ± standard deviation.

Samples	Fragrance/ Aroma	Flavor	Aftertaste	Salt/ Acidity	Bitter/ sweet	Mouthfeel	Uniform cups	Balance	Clean cups	Overall
Indigenous	7.82 ± 0.26	7.86 ± 0.31	7.72 ± 0.31	7.53 ± 0.33	7.80 ± 0.23	7.87 ± 0.31	9.55 ± 0.37	7.82 ± 0.31	9.55 ± 0.33	7.86 ± 0.36
Non-indigenous	7.62 ± 0.28	7.58 ± 0.38	7.45 ± 0.37	7.37 ± 0.35	7.55 ± 0.36	7.56 ± 0.36	9.46 ± 0.29	7.60 ± 0.37	9.47 ± 0.36	7.57 ± 0.42
Espírito Santo	7.22 ± 0.59	7.16 ± 0.64	7.11 ± 0.60	6.99 ± 0.50	7.19 ± 0.63	7.21 ± 0.65	8.93 ± 1.00	7.15 ± 0.61	8.99 ± 1.00	7.20 ± 0.62

trigonelline, formate, caffeine, chlorogenic acids, sucrose, citrate and/or malate, quinic acid and γ-quinide, lactate, and fatty acids. The two chemical shifts with positive loadings are associated with Conilon samples from Espírito Santo and could be attributed to caffeine. Espírito Santo Conilon coffees were considered to be more bitter in this study (see Fig. 6 and Table 2 with sensorial results) and were related to caffeine by ^{1}H NMR analysis, helping to better understand this result. Previous studies (Zani Agnoletti et al., **2022**) have demonstrated that Conilon coffee ^{1}H NMR spectra exhibited chlorogenic acids, trigonelline, caffeine, lactates, quinic acid, citrate, γ-butyrolactone, lipids, formic acid, sucrose, *N*-methyl-pyridinium, hydroxymethylfurfural, and acetic acid when submitted to different types of fermentation, highlighting lipids as a marker of quality.

In addition to analyzing the loadings separately as was above, it is possible to identify and discuss relationships between the variables within each block when they provide important information for the study. Correlations between data blocks are possible because greater and similar saliences proportions were found for them initially. The first one is an attempt to establish a relationship between the minerals and sensory attributes of the coffee beverage. It is an interesting outcome, since is known that minerals have an influence on perception of specific sensory attributes of coffee (Lingle & Menon, **2017**; UCDA, **2010**). For example, high potassium content make Canephora coffee less pleasant, increasing the perception of salinity and bitterness (Lingle & Menon, **2017**). Although it was expected to observe a multivariate relation between the K and some sensory attributes such as salt/acidity, aftertaste, or/and bitter/sweet that are influenced by it according to the Fine Robusta protocol (Lingle & Menon, **2017**; UCDA, **2010**), this was not found analyzing the data by chemometrics. However, there is a relationship between minerals Ca, Mg, Zn, Mn, and Cu with sensory attributes, allowing for a new discussion. This has not yet been the subject of discussion from a sensory point of view, as has been the influence of K on sensory aspects (Lingle & Menon, **2017**; UCDA, **2010**). Unfortunately, the present multi-block data analysis does not make it possible to identify the relationship between each mineral and each sensory attribute, because they were all analyzed together.

Another attempt to correlate the information was using the metabolites identified through ^{1}H NMR with sensory attributes. The overall quality is a combination of all sensory attributes and as all of they had the same behavior, it is expected that many metabolites are important for various sensory attributes. The caffeine identified may be related to the bitter/sweet and salt/acid aspect rations, as indicated in the Fine Robusta protocol (Lingle & Menon, **2017**; UCDA, **2010**). Moreover, the bitterness in Canephora coffee is also affected by chlorogenic acid, also identified by ^{1}H NMR, and potassium levels, whereas the sweetness is derived from the fruit acids and sugar levels (Lingle & Menon, **2017**; UCDA, **2010**). Formate, citrate and/or malate, lactate, and sucrose were also identified by ^{1}H NMR being able to contribute with sweetness and salt/acid aspect ratio. In the composition of organic acids in coffee, citric, malic, formic, chlorogenic, lactic, and quinic acids have been constantly present in high-quality coffees from Brazil and other countries (Rune et al., **2023**), as well as for Conilon from Espírito Santo (Lemos et al., **2020**; Zani Agnoletti et al., **2022**). However, no information is available about the Amazonian Robusta grown in Rondônia. Fatty acids contribute to aroma formation, texture, and mouthfeel or "body" of the beverage (Lingle & Menon, **2017**; Zani Agnoletti et al., **2023**). A previous study (Baqueta, Coqueiro, Março, Mandrone, et al., **2021**) using the classical ComDim analysis demonstrated that quinic acids, myo-inositol, and mainly *N*-methylpyridinium and lipids were related to the sensory attribute body in commercial Brazilian coffees, while trigonelline, formate, caffeine, chlorogenic acids, choline, lactate, citrate, and lipids were related to powder fragrance, drink aroma, acidity, bitterness, flavor, astringency, residual flavor, and overall quality. In general, chlorogenic acids, carbohydrates, proteins, trigonelline and caffeine are the main responsible for regulating coffee beverage quality (Ribeiro et al., **2011**).

Another interesting result observed was between FAAS and MIR, suggesting a correlation between elemental and molecular composition of the coffee samples. Trying to interpret these results, it is possible to suggest a multivariate correlation between minerals Fe and K with MIR bands between 1670 and 800 cm^{-1}, which are usually associated with chlorogenic acids, carbohydrates, lipids, trigonelline, and caffeine absorption bands (Munyendo et al., **2022**). Other correlations can exist between the minerals Ca, Mg, Zn, Mn, and Cu with MIR bands between 3500 and 1700 cm^{-1}, which are mainly associated with lipids and caffeine molecular vibrations (Munyendo et al., **2022**). This type of correlation is still little investigated in the literature but was recently found in another case study (Sushkov, Galbács, Fintor, Lobus, & Labutin, **2022**), where *potassium*-lithium were correlated with Raman spectroscopy bands of carotenoids and tryptophan, as well as lithium with fatty acids and valine absorption bands. This shows that the FAAS and MIR chemical information are complementary from a data analysis perspective, but it is a topic that is not yet well understood from a chemical perspective and could be used as an example for further studies.

Future research would benefit significantly from incorporating additional physicochemical parameters—such as color, pH, acidity, sugars, and lipid profiles, which are commonly analyzed in coffee studies—alongside the spectral data we have collected. Including these measurements could provide a more comprehensive understanding of coffee samples, enhancing the depth of analysis and creating a richer dataset. This approach would complement the spectrochemical information presented in our study and potentially uncover deeper insights into the relationships between these parameters, sensory attributes, and spectral data from various analyses. Such integration represents a promising new avenue for research, offering the potential to discover multivariate relationships that could advance the analysis and interpretation of specialized studies in coffee science. This holistic approach could greatly enrich future investigations in the field.

4. Conclusions

This study exemplifies the power of a comprehensive multiplatform characterization approach for Brazilian *Coffea canephora*, specifically conilon Capixaba and Amazonian robusta coffees, from both indigenous and non-indigenous sources. By integrating diverse analytical techniques with the Path-ComDim multi-block data analysis method, we have demonstrated the effectiveness of combining analytical and sensory data to gain a nuanced understanding of coffee composition and sensory attributes. This novel approach, which merges atomic and molecular information from spectroanalytical techniques, represents a significant advancement in food chemistry. Our findings underscore the importance of selecting suitable analytical methods to address specific research questions and reveal how combining multiple techniques can provide deeper insights into the relationships between chemical composition and sensory attributes. Notably, the observed correlations between elemental and molecular compositions highlight the interconnectedness of various compositional aspects of coffee. Overall, this study showcases the value of a multidimensional analytical strategy in coffee research, suggesting that the integration of diverse spectroscopic techniques and sensory analysis with advanced data analysis methods can significantly enhance coffee characterization. The study also highlights which are the most informative techniques that could be chosen to implement in new studies, for example, studying the correlation between information to build predictive models for a parameter of interest, whether chemical or sensory. Future research should build on these insights by incorporating additional analytical techniques, broadening the range of coffee samples, and further exploring the correlations identified. This integrated approach will not only advance our understanding of coffee chemistry but also facilitate more detailed and informed sensory analysis.

Future research direction could greatly benefit from a multifaceted

approach that integrates additional physicochemical analyses and chromatographic results alongside the spectral data obtained in our study. This integration would provide more detailed insights into the specific analytes present in Brazilian *Coffea canephora*, enhancing the overall interpretability of the coffee samples. By leveraging such comprehensive data analysis, researchers can explore an extensive array of data points, leading to more nuanced understandings of coffee characteristics. Moreover, incorporating agronomic and genetic parameters, such as phenotypic characterization, could provide valuable context to the spectral and physicochemical data. This holistic approach would allow for the exploration of how genetic and environmental factors influence coffee quality and composition.

Expanding the study to include new genotypes of Brazilian *Coffea canephora*—such as different botanical varieties (conilon, robusta, and hybrids), as well as emerging cultivars from the Western Brazilian Amazon and lesser-known conilon genotypes with unexplored chemical profiles—could reveal novel insights into the diversity of canephora coffees. This would contribute to a deeper understanding of the genetic and environmental influences on coffee chemistry and quality. Additionally, comparative studies between canephora and arabica coffees, particularly those of specialty grade, could offer valuable benchmarks and insights into the unique attributes of each type. Such comparisons would provide a broader context for interpreting the findings from canephora studies and guide future research directions. Integrating these diverse datasets and perspectives will not only enhance the depth of coffee analysis but also pave the way for innovative approaches in coffee science. This expanded scope promises to uncover new relationships and trends, ultimately advancing the field and offering practical insights for growers, roasters, and consumers alike.

CRediT authorship contribution statement

Michel Rocha Baqueta: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Douglas N. Rutledge:** Writing – review & editing, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Enrique Anastácio Alves:** Resources. **Manuela Mandrone:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis. **Ferruccio Poli:** Writing – review & editing, Methodology, Investigation, Formal analysis. **Aline Coqueiro:** Writing – original draft, Methodology, Investigation, Formal analysis. **Augusto Cesar Costa-Santos:** Writing – original draft, Methodology, Investigation, Formal analysis. **Ana Paula Rebellato:** Writing – original draft, Methodology, Investigation, Formal analysis. **Gisele Marcondes Luz:** Writing – original draft, Methodology, Investigation, Formal analysis. **Bruno Henrique Fermino Goulart:** Methodology, Formal analysis. **Eduardo Jorge Pilau:** Validation, Methodology, Investigation, Formal analysis. **Juliana Azevedo Lima Pallone:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Patrícia Valderrama:** Writing – review & editing, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Conceptualization.

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 grant #2022/04068-8, 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 402441/2022-2, and process 310933/2023-4), Agência Brasileira de Desenvolvimento 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

Assis, C., Pereira, H. V., Amador, V. S., Augusti, R., de Oliveira, L. S., & Sena, M. M. (2019). Combining mid infrared spectroscopy and paper spray mass spectrometry in a data fusion model to predict the composition of coffee blends. *Food Chemistry, 281* (December 2018), 71–77. https://doi.org/10.1016/j.foodchem.2018.12.044

Baqueta, M. R., Alves, E. A., Valderrama, P., & Pallone, J. A. L. (2023). Brazilian Canephora coffee evaluation using NIR spectroscopy and discriminant chemometric techniques. *Journal of Food Composition and Analysis, 116*(November 2022), 0–2. https://doi.org/10.1016/j.jfca.2022.105065

Baqueta, M. R., Coqueiro, A., Março, P. H., Mandrone, M., Poli, F., & Valderrama, P. (2021). Integrated 1H NMR fingerprint with NIR spectroscopy, sensory properties, and quality parameters in a multi-block data analysis using ComDim to evaluate coffee blends. *Food Chemistry, 355*(September 2020), Article 129618. https://doi.org/10.1016/j.foodchem.2021.129618

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*, Article 121526. https://doi.org/10.1016/J.TALANTA.2020.121526

Baqueta, M. R., Costa-Santos, A. C., Rebellato, A. P., Luz, G. M., Pallone, J. A. L., Marini, F., … Valderrama, P. (2024). Independent components–discriminant analysis for discrimination of Brazilian Canephora coffees based on their inorganic fraction: A preliminary chemometric study. *Microchemical Journal, 196*(October 2023), Article 109603. https://doi.org/10.1016/j.microc.2023.109603

Baqueta, M. R., Diniz, P. H. G. D., Pereira, L. L., Almeida, F. L. C., Valderrama, P., & Pallone, J. A. L. (2024). An overview on the Brazilian Coffea canephora scenario and the current chemometrics-based spectroscopic research. *Food Research International, 194*(July), Article 114866. https://doi.org/10.1016/j.foodres.2024.114866

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

Baqueta, M. R., Valderrama, P., Alves, E. A., Pallone, J. A. L., & Marini, F. (2023). Discrimination of Robusta Amazônico coffee farmed by indigenous and non-indigenous people in Amazon: **C**omparing benchtop and portable NIR using ComDim and duplex. *Analyst, 148*(7), 1524–1533. https://doi.org/10.1039/d3an00104k

Baqueta, M. R., Valderrama, P., Mandrone, M., Poli, F., Coqueiro, A., Costa-Santos, A. C., … Marini, F. (2023). 1H NMR, FAAS, portable NIR, benchtop NIR, and ATR-FTIR-MIR spectroscopies for characterizing and discriminating new Brazilian Canephora coffees in a multi-block analysis perspective. *Chemometrics and Intelligent Laboratory Systems, 240*(June), Article 104907. https://doi.org/10.1016/j.chemolab.2023.104907

Barbin, D. F., Felicio, A. L. D. 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

Belay, A., Ture, K., Redi, M., & Asfaw, A. (2008). Measurement of caffeine in coffee beans with UV/**V**is spectrometer. *Food Chemistry, 108*(1), 310–315. https://doi.org/10.1016/j.foodchem.2007.10.024

Brand, A. L. M., Silva, A. C. R., Garrett, R., & Rezende, C. M. (2024). Quantitative lipidomics in green robusta coffees from the Brazilian Amazon by LC-HRMS. *Food Bioscience, 57*, Article 103472. https://doi.org/10.1016/J.FBIO.2023.103472

Cariou, V., Qannari, E. M., Rutledge, D. N., & Vigneau, E. (2018). ComDim: From multiblock data analysis to path modeling. *Food Quality and Preference, 67*, 27–34. https://doi.org/10.1016/j.foodqual.2017.02.012

Coqueiro, A., Baqueta, M. R., Mandrone, M., Poli, F., & Valderrama, P. (2021). Coffee assessment using 1H NMR spectroscopy and multivariate data analysis: **A** review. In F. Atta-ur-Rahman, & M. I. Choudhary (Eds.), *Applications of NMR spectroscopy* (pp. 35–58). Bentham Science Publishers. https://books.google.com.br/books/about/Applications_of_NMR_Spectroscopy_Volume.html?id=u2lZEAAAQBAJ&redir_esc=y.

Correia, R. M., Tosato, F., Domingos, E., Rodrigues, R. R. T., Aquino, L. F. M., Filgueiras, P. R., … Romão, W. (2018). Portable near infrared spectroscopy applied to quality control of Brazilian coffee. *Talanta, 176*, 59–68. https://doi.org/10.1016/J.TALANTA.2017.08.009

Craig, A. P., Botelho, B. G., Oliveira, L. S., & Franca, A. S. (2018). Mid infrared spectroscopy and chemometrics as tools for the classification of roasted coffees by

cup quality. *Food Chemistry, 245*, 1052–1061. https://doi.org/10.1016/J.FOODCHEM.2017.11.066

Moraes-Neto, V. F.d., Baqueta, M. R., Caramês, E. T. D. S., Santana, F. B.d., 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

El Ghaziri, A., Cariou, V., Rutledge, D. N., & Qannari, E. M. (2016). Analysis of multiblock datasets using ComDim: Overview and extension to the analysis of (K + 1) datasets. *Journal of Chemometrics, 30*(8), 420–429. https://doi.org/10.1002/cem.2810

Hanafi, M., & Qannari, E. M. (2008). Nouvelles propriétés de l'analyse en composantes communes et poids spécifiques. *Journal de La Société Française de Statistique & Revue de Statistique Appliquée, 149*, 75–79.

Huck, C. W., Guggenbichler, W., & Bonn, G. K. (2005). Analysis of caffeine, theobromine and theophylline in coffee by near infrared spectroscopy (NIRS) compared to high-performance liquid chromatography (HPLC) coupled to mass spectrometry. *Analytica Chimica Acta, 538*(1–2), 195–203. https://doi.org/10.1016/j.aca.2005.01.064

Lemos, M. F., Perez, C., da Cunha, P. H. P., Filgueiras, P. R., Pereira, L. L., Almeida da Fonseca, A. F., … Scherer, R. (2020). Chemical and sensory profile of new genotypes of Brazilian Coffea canephora. *Food Chemistry, 310*(July 2019). https://doi.org/10.1016/j.foodchem.2019.125850

Lingle, T. R., & Menon, S. N. (2017). Cupping and **g**rading—Discovering **c**haracter and **q**uality. *The Craft and Science of Coffee*, 181–203. https://doi.org/10.1016/B978-0-12-803520-7.00008-6

Luna, A. S., da Silva, A. P., da Silva, C. S., Lima, I. C. A., & de Gois, J. S. (2019). Chemometric methods for classification of clonal varieties of green coffee using Raman spectroscopy and direct sample analysis. *Journal of Food Composition and Analysis, 76*(June 2018), 44–50. https://doi.org/10.1016/j.jfca.2018.12.001

Munyendo, L., Njoroge, D., & Hitzmann, B. (2022). The **p**otential of **s**pectroscopic **t**echniques in **c**offee **a**nalysis—A **r**eview. *Processes, 10*(1), 1–25. https://doi.org/10.3390/pr10010071

Qannari, E. M., Wakeling, I., Courcoux, P., & MacFie, H. J. (2000). Defining the underlying sensory dimensions. *Food Quality and Preference, 11*(1–2), 151–154. https://doi.org/10.1016/S0950-3293(99)00069-5

Ribeiro, **J. S.**, **Ferreira**, **M. M. C.**, & Salva, T. de **J. G.** (2011). Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. Talanta, 83(5), 1352–1358. doi:https://doi.org/10.1016/J.TALANTA.2010.11.001.

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 Chemistry, 371*, Article 131063. https://doi.org/10.1016/J.FOODCHEM.2021.131063

do Rosário, D. K. A., da Silva Mutz, Y., Vieira, K. M., Schwan, R. F., & Bernardes, P. C. (2024). Effect of self-induced anaerobiosis fermentation (SIAF) in the volatile compounds and sensory quality of coffee. *European Food Research and Technology, 250*(2), 667–675. https://doi.org/10.1007/s00217-023-04393-9

Rune, C. J. B., Giacalone, D., Steen, I., Duelund, L., Münchow, M., & Clausen, M. P. (2023). Acids in brewed coffees: Chemical composition and sensory threshold. *Current Research in Food Science, 6*(March), Article 100485. https://doi.org/10.1016/j.crfs.2023.100485

de Souza Costa, A. M., Lirio Soares, K., de Souza Silveira, L., Carlos Verdin Filho, A., Louzada Pereira, L., Moreira Osório, V., … Scherer, R. (2024). Influence of maturation and roasting on the quality and chemical composition of new conilon coffee cultivar by chemometrics. *Food Research International, 176*, Article 113791. https://doi.org/10.1016/J.FOODRES.2023.113791

Suhandy, D., & Yulia, M. (2017). Peaberry coffee discrimination using UV-visible spectroscopy combined with SIMCA and PLS-DA. *International Journal of Food Properties, 20*(1), S331–S339. https://doi.org/10.1080/10942912.2017.1296861

Sushkov, N., Galbács, G., Fintor, K., Lobus, N., & Labutin, T. (2022). A novel approach for discovering correlations between elemental and molecular composition using laser-based spectroscopic techniques. *Analyst, 147*(14), 3248–3257. https://doi.org/10.1039/D2AN00143H

Teixeira, U., Polari, D. C., Ferreira, M., Araújo, D., & Cirino, E. (2015). Food Science and Technology Identi fi cation of adulteration in ground roasted coffees using UV e Vis spectroscopy and SPA-LDA. *LWT, 63*, 1037–1041. https://doi.org/10.1016/j.lwt.2015.04.003

UCDA. (2010). UCDA - Uganda **c**offee **d**evelopment **a**uthority and CQI - **c**offee **q**uality **i**nstitute. Fine Robusta **s**tandards and **p**rotocols: **T**echnical standards, evaluation procedures and reference materials for quality-differentiated Robusta coffee. **https://www.coffeestrategies.com/wp-content/uploads/2015/04/compiled-standards-distribute1.1.pdf**.

Viencz, T., Acre, L. B., Rocha, R. B., Alves, E. A., Ramalho, A. R., & Benassi, M. D. T. (2023). Caffeine, trigonelline, chlorogenic acids, melanoidins, and diterpenes contents of Coffea canephora coffees produced in the Amazon. *Journal of Food Composition and Analysis, 117*(January). https://doi.org/10.1016/j.jfca.2023.105140

Zani Agnoletti, B., dos Santos Gomes, W., Falquetto de Oliveira, G., Henrique da Cunha, P., Helena Cassago Nascimento, M., Cunha Neto, Á., Louzada Pereira, L., Vinicius Ribeiro de Castro, E., Catarina da Silva Oliveira, E., & Roberto Filgueiras, P. (2022). Effect of fermentation on the quality of conilon coffee (Coffea canephora): Chemical and sensory aspects. *Microchemical Journal, 182*, Article 107966. https://doi.org/10.1016/J.MICROC.2022.107966

Zani Agnoletti, B., Louzada Pereira, L., Anastácio Alves, E., Barros Rocha, R., Gracieri Debona, D., Valério Vieira Lyrio, M., … Roberto Filgueiras, P. (2023). The terroir of Brazilian Coffea canephora: **C**haracterization of the chemical composition. *Food Research International, 113814*. https://doi.org/10.1016/J.FOODRES.2023.113814

Zhang, X., Li, W., Yin, B., Chen, W., Kelly, D. P., Wang, X., … Du, Y. (2013). Improvement of near infrared spectroscopic (NIRS) analysis of caffeine in roasted arabica coffee by variable selection method of stability competitive adaptive reweighted sampling (SCARS). *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy, 114*, 350–356. https://doi.org/10.1016/j.saa.2013.05.053