Food Chemistry 362 (2021) 129716

Contents lists available at ScienceDirect

Food Chemistry

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

Ecometabolic mixture design-fingerprints from exploratory multi-block data analysis in *Coffea arabica* beans from climate changes: Elevated carbon dioxide and reduced soil water availability

Gustavo Galo Marcheafave [a,*], Cláudia Domiciano Tormena [a], Amelia Elena Terrile [b], Carlos Alberto Rossi Salamanca-Neto [c], Elen Romão Sartori [c], Miroslava Rakocevic [d], Roy Edward Bruns [e], Ieda Spacino Scarminio [a,*], Elis Daiane Pauli [e,*]

[a] Laboratory of Chemometrics in Natural Sciences (LQCN), Department of Chemistry, State University of Londrina, CP 6001, 86051-990 Londrina, PR, Brazil
[b] Department of Chemistry, Federal University of Technology – Paraná, Av. dos Pioneiros 3131, 86036-370 Londrina, PR, Brazil
[c] Laboratory of Electroanalytical and Sensors (LAES), Department of Chemistry, State University of Londrina, CP 6001, 86051-990 Londrina, PR, Brazil
[d] Northern Rio de Janeiro State University - UENF, Plant Physiology Lab, Av. Alberto Lamego 2000, 28013-602 Campos dos Goytacazes, RJ, Brazil
[e] Institute of Chemistry, State University of Campinas, CP 6154, 13083-970 Campinas, SP, Brazil

ARTICLE INFO

Keywords:
Coffee beans
Statistical mixture design
ComDim
k-NN
Fingerprints
Ecometabolomics

ABSTRACT

Ecometabolic mixture design-fingerprinting in coffee cultivated under climate change was chemically explored using ComDim. Multi-blocks were formed using UV, NIRS, ^{1}H NMR, SWV, and FT-IR data. ComDim investigated all these different fingerprints according to the extractor solvent and in virtue of atmospheric CO_2 increase. Ethanol and ethanol-dichloromethane showed the best separations due to CO_2 environment. ^{1}H NMR loading indicate increases of fatty acids, caffeine, trigonelline, and glucose in beans under current CO_2 levels, whereas quinic acid/chlorogenic acids, malic acid, and kahweol/cafestol increased in beans under elevated CO_2 conditions. SWV indicated quercetin and chlorogenic acid as important compounds in coffee beans cultivated under current and elevated CO_2, respectively. Based on the ethanol and ethanol-dichloromethane fingerprints, *k*-NN correctly classified the beans cultivated under different carbon dioxide environments and water availabilities, confirming the existence of metabolic changes due to climate changes. SWV proved to be promising compared with widely used spectrometric methods.

1. Introduction

The high-throughput plant ecometabolomics uses hyphenated analytical strategies to perform global analysis of complex biological systems, where the integration of datasets acquired from various analytical instrumental sources is difficult to accomplish (Jan & Ahmad, 2019). In general, scientific studies in ecometabolomics are based on the use of hyphenated separation techniques with the most diverse types of detectors. Examples of hyphenated analytical techniques used are liquid chromatography-mass spectrometry (LC-MS) and gas chromatography-mass spectrometry (GC–MS) and their instrumental derivations (Nagler et al., 2018; Peters et al., 2018). In principle, when investigations are based on non-target analysis, the objective is to detect as many classes or metabolites as possible, seeking to trace the phenotypic expression at the metabolic level and provide an instant metabolic capture of the biological system (Jan & Ahmad, 2019). On the other hand, target analysis of ecometabolic fingerprints can be applied for screening biomarker substances for their efficient characterization of plant samples of interest (Jan & Ahmad, 2019).

Based upon the ideas concerning metabolomics and the advantages and successes obtained with fingerprinting techniques, researchers have configured strategies that apply instrumental methods that are not expensive, have quick analytical acquisition, and low technical demand. Among these techniques, we can highlight nuclear magnetic resonance (NMR) (Marcheafave et al., 2020a), Fourier-transform infrared spectroscopy (FT-IR) (Afonso et al., 2019), near-infrared spectroscopy (NIRS) (Marcheafave et al., 2020b), ultraviolet spectrophotometry (UV) (Marcheafave et al., 2020c; Tormena et al., 2019a) and more recently, electroanalytical methods that include square-wave voltammetry (SWV) (Salamanca-Neto et al., 2020) employed for numerous applications in

* Corresponding authors.
E-mail addresses: ggalo@uel.br (G.G. Marcheafave), ieda@uel.br (I.S. Scarminio), elispauli@hotmail.com (E.D. Pauli).

https://doi.org/10.1016/j.foodchem.2021.129716
Received 6 January 2021; Received in revised form 26 February 2021; Accepted 21 March 2021
Available online 1 April 2021

the food and biological sciences, process and product controls. Clearly, the limitations of structural determination of metabolites from complex biological systems are evident for specific molecules in target analysis. In most cases, the detailed molecular determination is not the aim of chemical analyses, but rather a snapshot of metabolic profile as a holistic strategy.

Spectroscopic methods are valuable tools to solve problems of discrimination or classification in systems subjected to environmental disturbance, where the purpose is not only to define chemical changes properly, but to predict the existence of chemical changes based on the quantitative and qualitative variations of the analytical signals using chemometric methods (Marcheafave et al., 2020a). It is worth mentioning that, these methods can be successfully applied in cases of chemical screening with a target as in the case of biomarkers (Tormena et al., 2020). However, based on non-target strategies, the chemical domain of the metabolic profile detected by these techniques, including chromatographic methods, is not able to represent the entire metabolome, due to limitations of the extraction process (Soares et al., 2018). Analytically, it is impossible to establish any sample preparation method that can extract all metabolomic substances from the investigated material. Alternatives are investigated to reduce this experimental void, such as the systematic application of statistical mixture design to obtain multifaceted fingerprint systems in ecometabolomic investigations by tracking the different metabolites between perturbed and controlled systems (Delaroza et al., 2014).

In addition to the mixture design, several analytical techniques of metabolic detection can be used or compared to increase the range of the metabolic classes extracted in the different solvents and their mixtures (Marcheafave et al., 2019a). Owing to their broad metabolic inclusion, their biological significance can be identified with the minimal presumption on the phenotypic expression at the metabolic level. Most of the analytical methodologies of metabolic univariate extraction are biased towards different classes and specific metabolites model only a fraction of the metabolome of the biological object (Jan & Ahmad, 2019). We hypothesize that multi-block exploratory analysis with variations of extractor solvent and analytical acquisition techniques can assist the coherent assembly of different metabolomic sets, and allow the expansion of propositions by inductive reasoning among investigated biological classes, reducing the effect of biased extractions and detections for certain classes. Here it is also hypothesized that metabolic or specific classes could possibly be pointed out as indicators of the environmental disturbance even though this variation is not known previously.

In ecometabolic studies, the greater the combination of multiple environmental changes that experimental plants undergo, the greater the understanding of how those environmental changes affect the chemical and biochemical aspects of the metabolome, and the biological risks associated with these changes (Jan & Ahmad, 2019; Peters et al., 2018). In Brazil, coffee currently represents the greatest commodity. Due to the risks associated with the production, quality, and commercialization, the climatic change scenario predicted for the end of the XXI century, has been implemented in the Brazilian Agricultural Research Corporation (EMBRAPA, Jaguariúna-SP). It is considered to be the first Free-Air Carbon dioxide Enrichment (FACE) facility in South America, established in 2011 (Ghini et al., 2015). It was also the first in the world that operated during 5 years with atmospheric carbon dioxide increase and during the last year variable water availability applied to coffee culture. The FACE experiments allowed exploration of the development, productivity, ecophysiology and phytosanitary aspects of coffee culture under increased CO_2 concentrations (Damatta et al., 2016; Ghini et al., 2015; Rakocevic et al., 2020; Tormena et al., 2019b). The long-term exposure to the levels of CO_2 projected for the climate around the end of the century, associated with the variation of water availability, provides valuable samples for ecometabolic studies with robust metabolic changes.

Consolidating different analytical gaps of ecometabolomics, the present study aims to compare the fingerprints of coffee beans subjected to a 2^2 factorial experiment with two factors (atmospheric CO_2 concentration and soil water availability) and two levels (elevated and current CO_2/irrigated and unirrigated coffee trees). The novelty of this study is a methodological search for coffee fingerprints using a statistical mixture design of four components (ethanol, dichloromethane, hexane, and ethyl ether) with five instrumental acquisitions (NMR, FT-IR, NIRS, UV, and SWV). The Common Dimension (ComDim) chemometric method for multi-block data analysis is applied to compile all variations of the ecometabolic fingerprints. The FD-MD-ComDim (factorial design/mixture design/Common Dimension) configuration allowed the simultaneous investigation of the variations of the extraction systems, the acquisition of signals from the extracted metabolic systems as well as multi-block exploration of the ecometabolic dataset for chemical changes in coffee during the FACE experiments. The supervised method of *k*-Nearest Neighbors (*k*-NN) allowed confirmation of the ComDim results, which showed the separation between coffee beans grown under different environmental conditions, using square-wave voltammetry.

2. Material and methods

2.1. Coffee bean collection

The augmented atmospheric carbon dioxide experiments was performed at a FACE facility located in Jaguaríuna – Brazil (22°43′S, 47°01′W), using *Coffea arabica* cultivar Catuaí Vermelho IAC 144. The FACE facility consisted of six ring plots for elevated CO_2 treatment and six control ring plots under current CO_2 conditions, each with 10 m diameter. The twelve ring plots were located in a 7 ha coffee field. The twelve ring plots were spaced at least 70 m apart. The injection of CO_2 to the atmosphere began on August 25, 2011, and ended on June 30, 2016 (Rakocevic & Matsunaga, 2018). At the beginning of the experiment, the concentration of CO_2 was approximately 390 μL L^{-1}. In the plots under elevated CO_2, the direct injections of pure CO_2 were 150–200 μL L^{-1} above this current level. Further details on the configuration of the FACE experiments on coffee trees can be seen at Ghini et al. (2015). Irrigation occurred in half of the coffee trees in the period between October 2015 and June 2016.

The coffee fruits were harvested in May 2016 in the 5th year of the FACE experiment. They were collected randomly in the rings from sixteen coffee plants stratified in four 50 cm-thick plant layers over the vertical tree profile (Marcheafave et al., 2020c). Four plants were grown with irrigation at current CO_2 levels, four in rainfed conditions under current CO_2, four irrigated plants under an elevated CO_2 atmosphere, and four unirrigated plants (rainfed conditions) at elevated CO_2. Green and red berries were collected, dried, processed with the separation of the bark and parchment, grounded, and stored in an ultra-freezer at −60 °C in paper bags.

2.2. Extraction strategy

2.2.1. Statistical mixture design in ecometabolomics

The principal limitation of the fingerprint approach corresponds to the process of extracting the set of metabolites from the plant tissue. Solvents used in the preparation of the samples will provide only a fraction of the original metabolic set and metabolic trends for only some specific classes or metabolites (Marcheafave et al., 2020b) will be observed (Marcheafave et al., 2019b). This limitation in the extraction process can be minimized with the systematic use of several solvents with different solvent/solute interaction characteristics (Barwick, 1997). The statistical mixture design strategically applies combinations of different solvents with different chemical properties providing an opportunity to systematically optimize the extractor system for specific metabolites (Silveira et al., 2020), multiple metabolites or even to maximize the exploratory fingerprint analysis. To maximize the extraction of the metabolites from *C. arabica* beans, a mixture design of

four components was used in this research with ethanol (Anidrol – Diadema, Brazil), dichloromethane (Anidrol – Diadema, Brazil), hexane (Anidrol – Diadema, Brazil), and ethyl ether (Vetec – Rio de Janeiro, Brazil). All organic solvents were of analytical grade. The four solvent design resulted in four extracts from the pure solvents, six extracts from the binary mixtures, four extracts from the ternary mixtures, and one extract from the quaternary mixture (Table 1), resulting in fifteen varied extracts or seventeen extracts on performing a triplicate at the central point. The extracts made by combinations of various solvents do not necessarily provide the fingerprint that contains the most important information in the process of discrimination or classification. The combination of different solvents allows antagonistic or synergistic interactions to occur, modifying the metabolic profiles of these extracts (Pauli, Bruns, & Scarminio, 2016).

2.2.2. Solvent-selectivity triangle for plant extractions

In recent years, our research group has reinforced the idea that experimental extractive procedures with biological materials, mainly vegetative, when properly designed, can serve as an effective systematic strategy to increase the number of metabolic classes investigated simultaneously, eliminating the bias of extractor solvent when univariate experiments are performed (Marcheafave et al., 2020a; Passari, Scarminio, Marcheafave, & Bruns, 2019). Thus, solvent classification systems based on similar properties are interesting for choosing the systems that compose the mixture design. In this work, the choice of solvents was based on the reclassification of Rutan et al. (1989) for the selectivity parameters (x_i) - dipolarity (x_n), hydrogen bond acidity (x_d), and basicity (x_e) from Rohrschneider-Snyder solvent-selectivity triangle (SST). Snyder's selectivity parameters (x_e, x_d, and x_n) are 0.53, 0.13, 0.34 for ethyl ether; 0.52, 0.19, 0.29 for ethanol; 0.27, 0.33, 0.40 for dichloromethane and 0, 0, 0 for hexane, respectively (Rutan et al., 1989). The magnitudes of these selectivity parameter values indicate the importance of different kinds of intermolecular interactions (Snyder, Carr, & Rutan, 1993).

2.2.3. Preparation of the extracts

The ultrasound-assisted extraction was performed in an ultrasonic bath from Unique model Ultracleaner 1600 operating at 40 kHz frequency and constant temperature of 15 °C. The erlenmeyers containing 2.5 g of grounded green coffee beans with 60 mL of each solvent (Table 1), were immersed in a sonication water bath for 60 min. After this time, the solution was filtered through a qualitative filter from Unifil, Brazil (0.16 mm thickness and 80 g m^2) for approximately 5 min.

Table 1
Statistical mixture design for four components: ethanol, dichloromethane, hexane, and ethyl ether. The proportions used for extraction given in mL.

Design point	Extract notation	Volume of solvents (mL) Ethanol (*E*)	Dichloromethane (*D*)	Hexane (*H*)	Ethyl ether (*e*)
1	*E*	60	0	0	0
2	*D*	0	60	0	0
3	*H*	0	0	60	0
4	*e*	0	0	0	60
5	*ED*	30	30	0	0
6	*EH*	30	0	30	0
7	*Ee*	30	0	0	30
8	*DH*	0	30	30	0
9	*De*	0	30	0	30
10	*He*	0	0	30	30
11	*EHe*	20	0	20	20
12	*EDH*	20	20	20	0
13	*EDe*	20	20	0	20
14	*DHe*	0	20	20	20
15	*EDHe*	15	15	15	15
16	*EDHe*	15	15	15	15
17	*EDHe*	15	15	15	15

The filtered solution was stored in closed erlenmeyers in a fume hood at 20 °C. The beans were again immersed in 60 mL of solvent and filtered 4 more times. At the end of the 5th extraction process, 300 mL of metabolic extract were obtained for samples from the four cultivation conditions in 17 extractive solvent systems, including the central point repetitions (Table 1). The liquid extract was used for UV, NIRS, and SWV analysis. For NMR and FT-IR the solutions were evaporated under forced air circulation and lyophilized to constant weight.

For the exploratory stage of the extracts, the beans collected at the three strata were homogenized. In the validation stage, the extractions were performed in the same way as described above for samples from sixteen plants and beans collected from three layers over the vertical tree profile, excluding the lowest layer (0–50 cm) that did not bear any fruit in the 4th production year (Rakocevic et al., 2021). This strategy included different layers of coffee fruit growth over the vertical tree profile, tending to increase the analytical robustness of the method, including variabilities of the experimental field.

2.3. Methods of measurements

2.3.1. Ultraviolet spectrophotometry measurements

The UV absorption fingerprint spectra were recorded using a UV–vis spectrophotometer Thermo Scientific model Evolution 60S, using quartz cuvettes (1 cm × 1 cm × 4.5 cm) with a 1 cm optical path length, coupled with VISION Lite software. The spectra were obtained from 190 to 400 nm with a 1 nm resolution at 22 ± 1 °C. The acquisitions were made from the diluted extracts in the proportion of 1:9 (1 mL) with its respective extracting solvent. Datasets were collected in absorbance units and saved in txt' data file format with 211 data points per sample.

2.3.2. Near-infrared transmittance spectroscopy measurements

NIR absorption measurements were carried out using a DLP NIR scan Nano software (Texas Instruments Inc., Dallas, USA) in transmittance mode equipped with a quartz cuvette (1 cm × 0.2 cm × 2.8 cm) with an optical path of 2 mm. NIR spectra were obtained between 1700 and 900 nm with a 3.9 nm resolution at 22 ± 1 °C and each mixture design-fingerprint spectra corresponded to an average of 99 scans. The acquisitions were made with 1 mL of extract immediately after extraction without dilution. NIR spectra were saved in absorbance units in txt' data file format with 215 data points per sample.

2.3.3. Proton nuclear magnetic resonance spectroscopy measurements

For NMR analysis, 50.0 mg of lyophilized extract were dissolved in 0.5 mL of DMSO-d_6 with TMS by using a vortex mixer. Then, the solubilized extract was put into a 5 mm NMR tube. The one-dimensional ^{1}H NMR measurements were performed at 28 °C on a Bruker Model Avance III spectrometer operating at 400 MHz, equipped with a 5 mm broadband inverse (BBI) multinuclear probe. Spectra were acquired with 64 scans, an acquisition time of 10 min with 1 s of waiting time. For each sample, the acquisition and processing of ^{1}H NMR spectra were performed, using the standard Bruker parameter (pulse sequence zg30). The analyses were processed by the TopSpin 3.6.1 software (Bruker). The resulting NMR spectra were manually phased, baseline-corrected, and referenced to TMS at a chemical shift (δ) of 0.0 ppm and saved in txt' data file format retaining 32,768 data points per sample. The residual δ of water and DMSO in 3.2867–3.4608 and 2.4703–2.5306 ppm regions, respectively, were removed retaining 17,590 data points per sample.

2.3.4. Square-wave voltammetry measurements

The SWV measurements were carried out using a three-electrode single-compartment glass cell coupled to an Autolab PGSTAT101 potentiostat/galvanostat (Metrohm; Schiedam, Netherlands) controlled by NOVA 2.1 software. A platinum plate was used as the auxiliary electrode, an Ag/AgCl (3.0 mol L^{-1} KCl) as the double-junction reference electrode, and a boron-doped diamond electrode (BDDE; 8000 ppm;

7.07 mm^2 exposed area; Windsor Scientific, Slough, United Kingdom) as working electrode. Before each analysis, the BDDE was manually polished using aqueous slurry of alumina powder (0.01 μm diameter) on a smooth polishing cloth (Metrohm; Schiedam, Netherlands), cleaned in ethanol under ultrasound bath during 30 s, and left to dry under room temperature. All experiments were performed at room temperature.

A solution of 0.1 mol L^{-1} perchloric acid in methanol was used as supporting electrolyte in the voltammetric measurements. To obtain the voltammetric profile of each extract, 500 μL of the extract were transferred to the electrochemical cell containing 4.5 mL of the supporting electrolyte. The cell content was stirred to homogenization during 30 s, with 5 s of pause prior placing the working electrode and starting either cathodic or anodic potential scanning, using SWV (amplitude: 20 mV; potential increment: 4 mV; frequency: 30 Hz). The voltammograms were baseline corrected, using the moving average mode. Datasets were collected in microamperes units and saved in txt' data file format with 374 data points per sample for cathodic SWV and 531 ones for anodic SWV.

Qualitative voltammetric profiling was obtained using chlorogenic acid, palmitic acid, D(-)-quinic acid, caffeic acid, ferulic acid, theobromine, p-coumaric acid, caffeine, quercetin, theophylline, trigonelline hydrochloride standards all from Sigma-Aldrich, D(+)-mannose, sucrose, D(+)-glucose, D(+)-xylose, D(-)-mannitol all from Merck, D-fructose from Vetec, and gallic acid from Reatec, at a concentration of 20 μg L^{-1} at the same electrochemical conditions as the extracts.

2.3.5. Fourier-transform infrared spectroscopy measurements

FT-IR spectral analysis of the extracts was carried out using a PerkinElmer Spectrum Two with a universal ATR accessory equipment (UATR) with the diamond crystal. The ATR-FTIR spectra were obtained in transmittance mode, in the wavenumber range of 4000–450 cm^{-1}, with a resolution of 1 cm^{-1} and force gauge of 60 (pressure arm) at room temperature. The spectroscopy analysis of extract were performed with solid extracts after drying. FT-IR spectra were collected by data acquisition Spectrum 10™ software (PerkinElmer) and saved in txt' data file format with 3551 data points per sample. For chemometric analysis, the dataset were converted to absorbance units.

2.4. Data analysis and software

2.4.1. Common dimension (ComDim) chemometrics method for multi-block data analysis in ecometabolomics

ComDim analysis is a particular implementation of common component and specific weights analysis (CCSWA) (Bouveresse et al., 2011). Its algorithm is fully described by Qannari et al. (2000). It belongs to the family of multi-block component analyses and consists of describing several data tables with, not necessarily having the same number of observed variables, on the same *n* samples (Makimori & Bona, 2019; Mazerolles et al., 2006). This method iteratively calculates, for each successive common dimension, a series of score vectors, which represent a direction of maximum dispersion of the samples in a space common to all the data tables. Each table has a specific weight (or 'salience') associated with each dimension for this common space, reflecting the importance that various tables of variables attach to the components (Cariou, Qannari, Rutledge, & Vigneau, 2018; Karoui, Dufour, & De Baerdemaeker, 2006). The difference between the values of the specific weights for a given dimension may be due to the fact that the dimension contains information which is present in some tables, but not in others. For each Common Component, the loading values of the variables in each table are also calculated (Rosa et al., 2017; Vieira et al., 2020), helping in the interpretation of results.

To maximize extraction of metabolic classes and minimize instrumental bias in the ecometabolic exploratory data analysis associated with mixture design-fingerprint, ComDim allowed the study of the fingerprints acquired by different instrumental techniques, indicating possible relationships between these tables, and discriminating the samples using global information included in all data tables.

The data for ComDim were organized in six blocks: UV spectra (block 1), FT-IR spectra (block 2), NIR spectra (block 3), ^{1}H NMR spectra (block 4), cathodic SW voltammograms (block 5), and anodic SW voltammograms (block 6) (Table 2). Each data block was properly pretreated, in which the parameters were empirically defined in preliminary tests (Table 2). Before the analysis by ComDim, each block was column centered and then normalized, using division by its Frobenius norm to obtain a scaled matrix (Qannari et al., 2000).

2.4.2. k-Nearest Neighbors (k-NN) algorithm in ecometabolomics

As a first attempt, linear classification methods (soft independent modeling of class analogies – SIMCA; linear and quadratic discriminant analysis – LDA and QDA– on the principal component scores) were applied, but they did not provide satisfactory results in terms of classification rates (data not shown), indicating a non-linear method needs to be tried. So, the *k*-NN method was chosen for the classification of coffee beans cultivated under current/elevated CO_2 levels with and without irrigation to confirm the results obtained by the unsupervised exploratory method (ComDim), which showed separations owing to different environmental conditions. This method is a supervised machine learning algorithm known as non-parametric classifiers and is indicated when the input dataset is not preassigned a parametric function ($\mathbf{Y} = f(\mathbf{X})$) - this limitation is configured by the complexity of the relationship between **X** and **Y** and so, linearity is not adequate to explain the relationship between input and output variables (Mezquita et al., 2020). As commonly described, the input data test is projected into a multidimensional space and is classified based on the calculation of the distance or metric dissimilarity using Manhattan, Euclidian, Minkowski, or Mahalanobis distances or by correlation calculations (Mezquita et al., 2020). The class dependency is assigned accordingly to the number of training samples, closer to the test sample, *i.e.*, the class with the greatest number of neighboring samples will determine the class of the test sample. Furthermore, it is one of the methods less prone to overfitting thanks to the simplicity of its algorithm (Malegori et al., 2018).

In ecometabolomics, many data sets have a non-linear distribution and, consequently, when linear models are used in mathematical modeling, they can show failures or overfitting in the predictive model. The non-linearity in the spectroscopic or voltammetric datasets of plants originates from several abiotic or biotic factors, as atmospheric CO_2 concentration and the emisson of fruits on the vertical profile of the plant. In this case, different stratification receive different influences from the environment: light, shade, temperature (Rakocevic et al., 2021) are examples of factors that influences changes through the vertical profile, changing the chemical composition of different plant organs, such as leaves, buds (Tormena et al., 2019a) and beans (Rakocevic et al., 2021).

Anodic SW voltammograms for *k*-NN analysis were organized in four matrices according to their respective class: current CO_2 unirrigated, current CO_2 irrigated, elevated CO_2 unirrigated, and elevated CO_2 irrigated. Then, the whole data set, made up of 25 voltammograms for each class, was split into a training set of 18 samples and a test set of 7 samples by the Kennard–Stone algorithm (Kennard & Stone, 1969). Finally, the matrices were again organized according to the objective of classification by the *k*-NN analysis.

2.4.3. Data processing

Preprocessing, analysis, and visualization of the spectral data was performed with the chemometric tools Gamma 3.0 (Bona, E. GAMMA-Routines for Multivariate Analysis. Campo Mourão 2017; Galvan - drop) for ComDim and PLS_Toolbox 8.7 (Eigenvector, US) for de *k*-NN analysis, applied with Matlab 2016 (Mathworks, US). ComDim was implemented in MATLAB R2008b according to the algorithm presented in Ghaziri, Cariou, Rutledge, & Qannari (2016) and described in detail by Bouveresse et al. (2011).

Table 2
Data blocks of the *C. arabica* bean fingerprints obtained from different instrumental techniques for ComDim analysis.

Block	Measurement ranges of the fingerprints	Dimension matrix(samples × variables)	Preprocessing
UV	190–400 nm	68 × 211	Savitzky–Golay filter(9-point window)
FT-IR	4000–450 cm^{-1}	68 × 3551	Multiplicative Scattering Correction (MSC)
NIR(transmittance)	900–1700 nm	68 × 215	Standard NormalVariate (SNV)
^{1}H NMR	10.9997–0.0177 ppm	68 × 17590	Pareto-Scaling
Cathodic SWV	1.5–0.0 V	68 × 374	Moving Average
Anodic SWV	−0.8–1.3 V	68 × 531	Moving Average

3. Results and discussion

The multi-block analysis by ComDim was carried out with spectral and voltammetric datasets, in which the fingerprints were obtained from the mixture extract solutions. As different extraction solvents and analytical techniques were applied, the necessity for further comparative analysis was imposed. ComDim was applied to calculate the global scores for the samples and the loadings of the variables in each block (UV, FT-IR, NIR, NMR, Cathodic SWV and Anodic SWV) for each Common Dimension (CD), also providing the salience values (Rosa et al., 2017). ComDim analysis was performed using 8 common dimensions, considered sufficient to contemplate all important sources of variation in the data. However, only four were shown (Fig. 1 and Fig. S1) since the other CDs did not provide significant additional information to separate the samples (Fig. S2).

Score plots CD1 and CD2 were depicted (Fig. S1). The separation of samples related to the characteristics of the solvents used in the extraction of metabolites from *C. arabica* beans was observed. The CD1 score plot showed the discrimination between ethanol-containing extracts, from other extracts. According to the saliences (Table 3), all analytical techniques contributed to this separation, where the most influenced ones were anodic and cathodic SWV, followed by FT-IR and NMR, UV and NIRS. CD2 showed the separation of pure ethyl ether and the binary mixture ethanol-ethyl ether extracts from other extracts. Two blocks contributed to this separation, UV spectra, with high salience values, and FT-IR spectra. Further information and discussions of the variables that most contributed to the discrimination of extraction solvents can be found in the Supplementary material (Fig. S3).

The projection of the ComDim scores along CD4 and CD8 showed the discrimination of the *C. arabica* beans according to cultivation conditions (Fig. 1). CD4 permitted separation of most extracts into two groups related to the beans cultivated under different CO_2 environments. The *C. arabica* beans collected from plants grown in elevated CO_2 environment were grouped at the positive side of CD4, while those from current CO_2 environment were at the negative side of CD4. The saliences for CD4 (Table 3) showed that the predominant contribution to the dispersion of the *C. arabica* beans was due the metabolic changes observed on the anodic SW voltammograms. By assessing the CD4 loadings plot

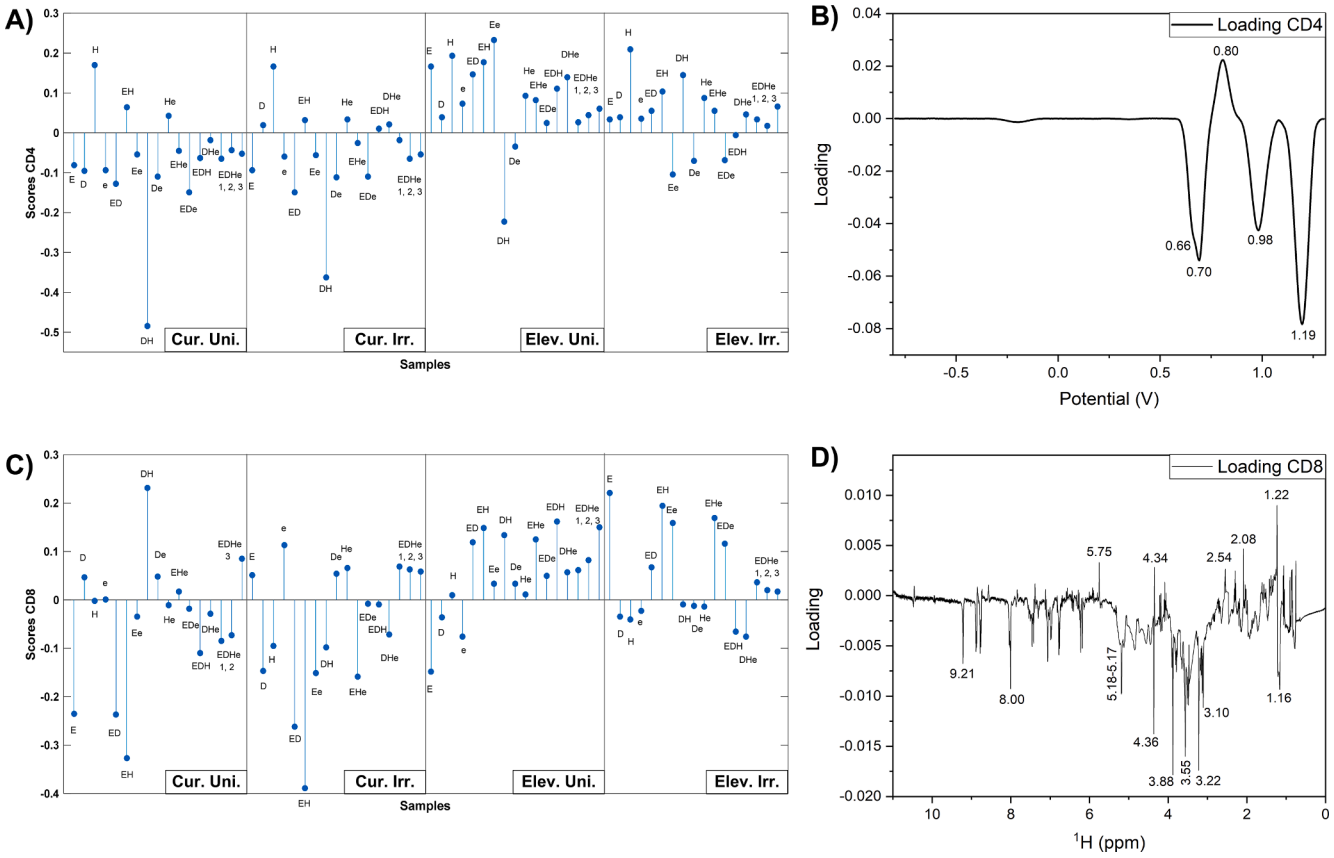

Fig. 1. Scores of the ComDim analysis for CD4 (A) and CD8 (C) for the fingerprints of the four classes of *Coffea arabica* beans and loading plots for the anodic SW voltammograms in CD4 (B) and the ^{1}H NMR in CD8 (D); Cur. Uni: Current CO_2 unirrigated; Cur. Irr.: Current CO_2 irrigated; Elev. Uni.: Elevated CO_2 unirrigated; Elev. Irr.: Elevated CO_2 irrigated. The score identifications correspond to the extraction medium of the statistical mixture design used (Table 1): *E* = ethanol; *D* = dichloromethane; *H* = hexane; *e* = ethyl ether.

Table 3
Saliencies of each block of applied methodology and for each common dimension (CD).

Block	CD1	CD2	CD3	CD4	CD5	CD6	CD7	CD8
UV	**0.2658**	**0.4893**	0.0283	0.0365	0.0041	0.0017	0.1249	0.0005
FT-IR	**0.3600**	**0.1092**	0.3107	0.0025	0.0096	0.1429	0.0010	0.0006
NIR	**0.1941**	0.0274	0.1075	0.0259	0.2022	0.0274	0.0298	0.0027
NMR	**0.3394**	0.0082	0.0465	0.0468	0.0111	0.0056	0.0093	**0.1317**
Cathodic SWV	**0.6325**	0.0786	0.0494	0.0620	0.0097	0.0066	0.0076	0.0070
Anodic SWV	**0.6038**	0.0344	0.0601	**0.1653**	0.0055	0.0038	0.0096	0.0081
CD variance (%)	76.51	15.23	3.49	1.46	1.21	0.36	0.41	0.33
Accumulated variance (%)	76.51	91.74	95.23	96.69	97.90	98.26	98.67	99.00

*The highest salience values for Common Dimensions (CD) with significant results are marked in bold.

(Fig. 1B), it appears that the beans collected from plants cultivated under current CO_2 environment were discriminated by the anodic voltammetric peaks at 0.66, 0.70, 0.98, and 1.19 V and those from the elevated CO_2 by 0.80 V. Quercetin was one of the important chemical compounds attributed to coffee beans collected from the current CO_2 environment, being confirmed by spiking ethanol extracts with the quercetin analytical standard (Fig. S4). The chlorogenic acid was one of the important metabolites for the separation of coffee beans collected from plants cultivated under elevated CO_2, also confirmed by spiking with its analytical standard (Fig S4).

The score projection for CD8 (Fig. 1C) indicated the discrimination of most *C. arabica* beans into two clusters, in which beans from plants cultivated under elevated CO_2 environment were grouped into positive scores and those from plants cultivated under current CO_2 environment into negative ones. The salience values (Table 3) pointed out that this separation was better expressed in the NMR spectra (Block 4). ComDim loadings (Fig. 1D) revealed that the signals at 1.22; 2.08; 2.54; 4.34; and 5.75 ppm had positive values and contributed to discriminate beans from elevated atmospheric CO_2. The signals at 1.16; 3.10; 3.22; 3.55; 3.88; 4.36; 5.17–5.18; 8.00; and 9.21 ppm had negative values and were relevant to discriminate beans of *C. arabica* plants cultivated under current CO_2 conditions. The metabolic separation due to CO_2 environment has also been found when ^{1}H NMR was associated with other chemometric tools (principal component analysis (PCA), analysis of variance - simultaneous component analysis (ASCA), and partial least squares - discriminant analysis (PLS-DA)) (Marcheafave et al., 2020a). Spectral profiles for fatty acids, caffeine, trigonelline, and glucose increase in beans under current atmospheric CO_2, while spectral profiles for quinic acid/chlorogenic acids, malic acid, and kahweol/cafestol increase in coffee beans grown in elevated CO_2. Furthermore, by ^{1}H NMR mixture design-fingerprints, carbon dioxide has the only significant effect on the chemical change in the metabolome of coffee beans, reinforcing actual results found in CD4 and CD8. No separation relative to water availability of the *C. arabica* beans in the computed CDs was observed. This was due to the high variability of solvents and mixtures, where the chemical variation related to them in the samples was greater than the variation caused by different water regimes.

To our knowledge, the SWV technique has not yet been explored to classify coffee beans cultivated under different CO_2 levels. This instrumental technique was applied to analyse new samples of *C. arabica* extracted with two solvents: pure ethanol (*E*) and the binary mixture ethanol-dichloromethane (*ED*). Two points were taken into account when choosing these solvents. First, the extracts obtained with them were correctly separated according to the levels of atmospheric CO_2 in which the coffee plants were cultivated (Fig. 1A and 1C). The second point is that the original voltammograms from *ED* and *E* extracts (Fig. S5) presented well-defined anodic peaks at 0.66, 0.70, 0.80, 0.98 and 1.19 V, which were important for the class separation, as indicated by the loading ComDim values (Fig. 1B and 1D).

New extractions were performed from beans collected from three 50 cm-thick-layers of the vertical profile of the plants, to assist the classification of beans from the coffee plants cultivated at the different CO_2 levels. Berries from lower positions are shaded, receiving low photosynthetically active radiation (PAR) (Rakocevic et al., 2021). At midday, during the harvest period, the PAR incident on the top of the canopy is about 1700 μmol m^{-2} s^{-1}, diminishing the available average irradiance to 439, 122 and 4 μmol m^{-2} s^{-1} for layers 3, 2 and 1, respectively. The temperature varies along the vertical profile, but less so than PAR. It ranges from 32.5 °C to 27.4 °C in the morning, from 34.2 °C to 31.2 °C by midday, and from 32.3 °C to 28.0 °C in the afternoon, for the upper and the lowest layers, respectively. These microclimate differences are expected to provoke metabolic changes over the vertical profile of the same coffee plant, as shown in Rakocevic et al. (2021) using NIRS analysis. The anodic SW voltammograms could provide non-linear distribution data due to environmental variations related to stratifications and even about the homogenization of CO_2 gas among heights in which the berries are located. PCA results in Marcheafave et al. (2020c) for UV spectra of coffee beans cultivated under different CO_2 levels and collected at different strata, show that extracts of the same class had some separate clusters and could be explained owing to stratifications. So, in order to confirm the separation of beans cultivated at different CO_2 levels the anodic SW voltammogram matrices to *ED* and *E* extracts were submitted to classification analysis using the *k*-NN algorithm, separately for each solvent. Here *k*-NN was chosen because it is non-probabilistic and free from any assumptions about variable distributions (Oliveri et al., 2019). Then, the whole data set, made up of 100 voltammograms for each solvent, was divided into a training set of 72 samples (36 current and 36 elevated) and a test set of 28 samples (14 current and 14 elevated) by the Kennard–Stone algorithm (Kennard & Stone, 1969). Three nearest neighbors were selected for the two classes as they provided the best classification as indicated by cross-validation results. The quality performance parameters found for *k*-NN models for classification between CO_2 levels were calculated from the confusion matrix. (Table 4).

Good classification performance was achieved for the *ED* extracts of coffee beans from the training set, with sensitivity (rate of true positives) and specificity (rate of true negatives) of 91.67% for elevated and current CO_2 classes (Table 4). For the prediction set, the model showed sensitivity of 100% and specificity of 92.86% for current CO_2 class and

Table 4
Confusion matrix and quality performance parameters for *k*-NN model built from anodic SWV voltammograms of ethanolic and ethanol-dichloromethane binary mixture extracts for discrimination between CO_2 levels of coffee beans collected from plants grown in different environments. Cur.: Current CO_2; Elev.: Elevated CO_2; Sens: Sensitivity; Spec: Specificity.

	Training set				Test set			
$k = 3$	Cur	Elev	Sens	Spec	Cur	Elev	Sens	Spec
Ethanol:Dichloromethane (*ED*)								
Cur.	33	3	91.67	91.67	14	1	100	92.86
Elev.	3	33	91.67	91.67	0	13	92.86	100
Ethanol (*E*)								
Cur.	34	0	94.44	100	13	2	92.86	85.71
Elev.	2	36	100	94.44	1	12	85.71	92.86

92.86% and 100%, respectively, for the elevated CO_2 class. In the case of pure ethanol extracts, the training model was 94.44% sensitive and 100% specific for the coffee beans from the current CO_2 class, and had 100% sensitivity and 94.44% specificity for beans from elevated CO_2 class. In the prediction set, the sensitivity was 92.86% and specificity 85.71% for the current CO_2 class. The prediction rates for beans cultivated under elevated atmospheric CO_2 had a sensitivity of 85.71% and a specificity of 92.86%.

The results of the *k*-NN models confirmed that coffee beans were classified according to CO_2 environments, even when collected from different layers of the coffee trees, by using SWV profiling. In order to assess whether *k*-NN is able to classify coffee beans from plants grown in the two CO_2 environments while taking into account the water regime, four classes were investigated using the *ED* and *E* extracts. Each set relative to each solvent was divided into a training set of 72 samples (18 current CO_2 unirrigated, 18 current CO_2 irrigated; 18 elevated CO_2 unirrigated, and 18 elevated CO_2 irrigated) and a test a set of 28 samples (7 current CO_2 unirrigated, 7 current CO_2 irrigated; 7 elevated CO_2 unirrigated, and 7 elevated CO_2 irrigated), using the Kennard–Stone algorithm (Table 5).

For the training set for *ED* extracts of coffee beans collected from irrigated plants cultivated under current CO_2 environment, the sensitivity and specificity were 100% (Table 5) whereas the prediction set results were 100% sensitive and 90.48% specific. The beans from the class formed by unirrigated plants at the current CO_2 level had a training set sensitivity of 83.33% and specificity of 94.44%. For the prediction set, the sensitivity was 85.71% and specificity 100%. The sensitivity and specificity results found by *k*-NN for beans from irrigated plants cultivated under elevated CO_2 were the same as for the class of beans from unirrigated plants cultivated under current CO_2. The training results for coffee beans from the elevated CO_2 unirrigated class was 94.44% sensitive and 98.15% specific, whereas the prediction set showed a sensitivity of 100% and specificity of 100%.

The extracts obtained using pure ethanol for coffee beans from irrigated trees cultivated under current CO_2 had 88.89% sensitivity and 100% specificity for *k*-NN training model, with 100% and 85.71%, respectively, for the prediction set (Table 5). Beans from both unirrigated plants grown under current CO_2 and irrigated plants grown under elevated CO_2 showed sensitivity and specificity of 100% for the training set. However, the prediction model for the class of beans from unirrigated trees grown in the current CO_2 was only 71.43% sensitive, while for sensitivity of the model for the class of beans from the irrigated plants cultivated in elevated CO_2 was 85.71%. Both models had specificity higher than 95%. Finally, for the prediction dataset of the beans from unirrigated plants grown in elevated CO_2, the model presented sensitivity and specificity of 85.71% and 100%, respectively.

The advancement of analytical techniques for metabolic detection and quantification has deepened our knowledge about ecometabolic changes in plants due to environmental disturbances (Jan & Ahmad, 2019). However, as we have shown, each instrumental technique contributes with different chemical information about metabolic fingerprints. The ComDim method has proven to be an advantageous tool to investigate instrumental responses pertinent to the interest of the investigation. In this study, ComDim provided information on the extraction process in relation to the solvent used, beyond the metabolic alterations owing to climate change. It is worth mentioning that FD-MD-ComDim is promising in exploratory analysis processes when the knowledge about any experimental factor is unknown, such as altered metabolites, optimal extractor solvent and instrumental technique that best enhances the discrimination of the investigated classes, and even determine which environmental disturbances have major effects on the modification of the chemical composition of the plant organ.

The first CD showed that all techniques are capable of exploring the effect of the extractor solvent on the composition of the fingerprint. The process of identifying the similarity or dissimilarity between the scores is fundamental to the choice of suitable extracting solvents that assign different chemical information to the fingerprinting method. This type of systematic study allows increases in the number of metabolic classes explored in the classification process or by means of comparisons among different spectral profiles, optimizes the figures of merit for classification models (Marcheafave et al., 2020a).

CD4 and CD8 indicated the separation between the scores due to the concentration of atmospheric carbon dioxide. ComDim highlights the potential of anodic SWV analysis and ^{1}H NMR as spectral profiles best indicating the biological significance due to climate change effects on the chemical composition of coffee beans. ^{1}H NMR was the most suitable spectrometric technique for obtaining metabolomic profiles, but requires expensive equipment and analysis. In spite of its low cost and relative simplicity, SWV metabolic classifications coupled with suitable chemometric methods of pattern recognition are still rarely used in ecometabolomics studies with few examples in food quality control (Salamanca-Neto et al., 2020). Through multi-block methods, it was possible to compare SWV with widely used techniques (*e.g.* LC-MS, GC–MS, NMR, FT-IR) and evaluate their use and implementation, as applied in our study. SWV analysis is easy to operate being fast, inexpensive, sensitive and the respective instrumentation is in full scientific development.

Finally, *k*-NN successfully classified *C. arabica* beans owing to the concentration levels of atmospheric CO_2, showing high specificity and sensitivity for *E* and *ED* extracts, as indicated by ComDim scores. The classification by *k*-NN model showed that the distribution of the beans in the vertical profile of the plant provide non-linear fingerprints due to the environmental variations. Among the 15 extracts obtained from the statistical mixture design, those obtained with ethanol and ethanol-dichloromethane contained the best metabolic sets that indicating chemical change due to the increase in atmospheric carbon dioxide. These same solvents are promising in the discrimination of coffee beans grown under different CO_2 levels, when analysed by other instrumental

Table 5
Confusion matrix and quality performance parameters for *k*-NN model of the anodic SWV voltammograms of ethanolic and ethanol-dichloromethane binary mixture extracts for discrimination between the CO_2 levels of coffee beans collected from irrigated and unirrigated trees. Cur. Uni.: Current CO_2 unirrigated; Cur. Irr.: Current CO_2 irrigated; Elev. Uni.: Elevated CO_2 unirrigated; Elev. Irr.: Elevated CO_2 irrigated; Sens: Sensitivity; Spec: Specificity.

	Training set						Test set					
$k = 3$	Cur.Irr.	Cur.Un.	Elev.Irr.	Elev.Un.	Sens	Spec	Cur.Irr.	Cur.Un.	Elev.Irr.	Elev.Un.	Sens	Spec
Ethanol:Dichloromethane (*ED*)												
Cur. Irr.	18	0	0	0	100	100	7	1	1	0	100	90.48
Cur. Un.	0	15	3	0	83.33	94.44	0	6	0	0	85.71	100
Elev. Irr.	0	2	15	1	83.33	94.44	0	0	6	0	85.71	100
Elev. Un.	0	1	0	17	94.44	98.15	0	0	0	7	100	100
Ethanol (*E*)												
Cur. Irr.	16	1	0	0	88.89	100	7	1	1	1	100	85.71
Cur. Un.	0	17	0	0	100	100	0	5	0	0	71.43	100
Elev. Irr.	0	0	18	0	100	100	0	1	6	0	85.71	95.24
Elev. Un.	2	0	0	18	100	96.30	0	0	0	6	85.71	100

and chemometric methods (Marcheafave et al., 2020b, 2020c), which is confirmed by our results.

4. Conclusions

The Factorial design/Mixture design/ComDim strategy showed good results for the multi-block analysis of the data obtained from different analytical methods and multifaceted fingerprints. ComDim allowed the discrimination of the coffee beans cultivated under two CO_2 levels by anodic SWV and ^{1}H NMR, according to CD4 and CD8 parameters, and indicated the solvent extractors promoting better separation between classes. This study highlighted the potential of anodic SWV analysis, coupled with suitable chemometric methods for the classification of coffee samples cultivated under two different atmospheric CO_2 levels. According to CD4 loading plot, the SW voltammetric fingerprints indicated that the beans cultivated at the current CO_2 level were discriminated by the anodic peaks at 0.66, 0.70, 0.98, and 1.19 V, and the beans under elevated CO_2 by the peak at 0.80 V. Whereas quercetin was one of the important chemical compounds in coffee beans cultivated under current CO_2 environmental conditions, chlorogenic acid was predominant for elevated CO_2. From the CD8 parameters, ^{1}H NMR fingerprints for fatty acids, caffeine, trigonelline, and glucose increased in beans at current CO_2 levels, while quinic acid/chlorogenic acids, malic acid, and kahweol/cafestol increased in coffee beans from elevated CO_2. The extracts obtained in pure ethanol and ethanol-dichloromethane binary mixtures indicated the best separations between the beans due to the alteration of carbon dioxide. On the other hand, ComDim showed no pattern of separation between the scores due to water availability, indicating that the CO_2 factor had greater impact on the metabolome of *Coffea arabica* beans than water availability. The *k*-NN classification confirmed the ComDim results, discriminating and predicting beans cultivated under different CO_2 environments. In addition, through the selection of extracting solvents, the *k*-NN algorithm was able to classify ecometabolic fingerprints due to the water regime. As such, electroanalytical techniques are promising tools to explore changes in environmental conditions for coffee beans and other plant matrices complementing other analytical techniques.

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.

Acknowledgements

The authors acknowledge financial support of the Consórcio Brasileiro de Pesquisa e Desenvolvimento de Cafés (no.° 02.13.02.042.00.00), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 (GGM, CDT and CARSN). ISS (PQ, Process: 302204/2018-0), ERS (PQ, Process: 305320/2019-0), EDP (PDJ, Process: 151843/2019-8) and MR (PV, Process: 312959/2019-2) thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for granted fellowships. The authors are thankful to Dr. Evandro Bona, Universidade Tecnológica Federal do Paraná - UTFPR, Câmpus Campo Mourão, Brazil, for providing the ComDim algorithm.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2021.129716.

References

Afonso, S., Silva, F. B., Marcheafave, G. G., Hatumura, P. H., Bruns, R. E., & Scarminio, I. S. (2019). Influence of seasonality and sunlight effects on *Rollinia mucosa* leaves fingerprint. *Journal of the Brazilian Chemical Society, 30*(5), 968–977. https://doi.org/10.21577/0103-5053.20180242.

Barwick, V. J. (1997). Strategies for solvent selection - A literature review. *TrAC - Trends in Analytical Chemistry, 16*(6), 293–309. https://doi.org/10.1016/S0165-9936(97)00039-3.

Bouveresse, D. J., Pinto, R. C., Schmidtke, L. M., Locquet, N., & Rutledge, D. N. (2011). Identification of significant factors by an extension of ANOVA – PCA based on multi-block analysis. *Chemometrics and Intelligent Laboratory Systems, 106*(2), 173–182. https://doi.org/10.1016/j.chemolab.2010.05.005.

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.

Damatta, F. M., Godoy, A. G., Menezes-Silva, P. E., Martins, S. C. V., Sanglard, L. M. V. P., Morais, L. E., Torre-Neto, A., & Ghini, R. (2016). Sustained enhancement of photosynthesis in coffee trees grown under free-air CO_2 enrichment conditions: Disentangling the contributions of stomatal, mesophyll, and biochemical limitations. *Journal of Experimental Botany, 67*(1), 341–352. https://doi.org/10.1093/jxb/erv463.

Delaroza, F., Rakocevic, M., Malta, G. B., Bruns, R. E., & Scarminio, I. S. (2014). Spectroscopic and chromatographic fingerprint analysis of composition variations in *Coffea arabica* leaves subject to different light conditions and plant phenophases. *Journal of the Brazilian Chemical Society, 25*(11), 1929–1938. https://doi.org/10.5935/0103-5053.20140172.

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.v30.810.1002/cem.2810.

Ghini, R., Torre-Neto, A., Dentzien, A. F. M., Guerreiro-Filho, O., Iost, R., Patrício, F. R. A., Prado, J. S. M., Thomaziello, R. A., Bettiol, W., & DaMatta, F. M. (2015). Coffee growth, pest and yield responses to free-air CO_2 enrichment. *Climatic Change, 132*(2), 307–320. https://doi.org/10.1007/s10584-015-1422-2.

Jan, S., & Ahmad, P. (2019). Introducing Metabolomics. In *Ecometabolomics: Metabolic Fluxes versus Environmental Stoichiometry* ((1st ed.,, pp. 1–56). Academic Press. https://doi.org/10.1016/b978-0-12-814872-3.00001-1.

Karoui, R., Dufour, É., & De Baerdemaeker, J. (2006). Common components and specific weights analysis: A tool for monitoring the molecular structure of semi-hard cheese throughout ripening. *Analytica Chimica Acta, 572*(1), 125–133. https://doi.org/10.1016/j.aca.2006.04.089.

Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics, 11*(1), 137–148. https://doi.org/10.2307/1266770.

Makimori, G. Y. F., & Bona, E. (2019). Commercial instant coffee classification using an electronic nose in tandem with the ComDim-LDA Approach. *Food Analytical Methods, 12*(5), 1067–1076. https://doi.org/10.1007/s12161-019-01443-5.

Malegori, C., Grassi, S., Ohm, J.-B., Anderson, J., & Marti, A. (2018). GlutoPeak profile analysis for wheat classification: Skipping the refinement process. *Journal of Cereal Science, 79*, 73–79. https://doi.org/10.1016/j.jcs.2017.09.005.

Marcheafave, G. G., Tormena, C. D., Mattos, L. E., Liberatti, V. R., Ferrari, A. B. S., Rakocevic, M., Bruns, R. E., Scarminio, I. S., & Pauli, E. D. (2020a). The main effects of elevated CO_2 and soil-water deficiency on ^{1}H NMR-based metabolic fingerprints of *Coffea arabica* beans by factorial and mixture design. *Science of The Total Environment, 749*, 142350. https://doi.org/10.1016/j.scitotenv.2020.142350.

Marcheafave, G. G., Pauli, E. D., Tormena, C. D., Mattos, L. E., de Almeida, A. G., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2020b). Irrigated and CO_2 level effects on metabolism in *Coffea arabica* beans from mixture design–near infrared fingerprints. *Microchemical Journal, 152*, 104276. https://doi.org/10.1016/j.microc.2019.104276.

Marcheafave, G. G., Pauli, E. D., Tormena, C. D., Ortiz, M. C. V., de Almeida, A. G., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2020c). Factorial design fingerprint discrimination of *Coffea arabica* beans under elevated carbon dioxide and limited water conditions. *Talanta, 209*, 120591. https://doi.org/10.1016/j.talanta.2019.120591.

Marcheafave, G. G., Tormena, C. D., Pauli, E. D., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2019a). Experimental mixture design solvent effects on pigment extraction and antioxidant activity from *Coffea arabica* L. leaves. *Microchemical Journal, 146*, 713–721. https://doi.org/10.1016/j.microc.2019.01.073.

Marcheafave, G. G., Tormena, C. D., Afonso, S., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2019b). Integrated chemometric approach to optimize sample preparation for detecting metabolic changes provoked by abiotic stress in Coffea arabica L. leaf fingerprints. *Journal of the Brazilian Chemical Society, 30*(10), 2085–2094. https://doi.org/10.21577/0103-5053.20190068.

Mazerolles, G., Hanafi, M., Dufour, E., Bertrand, D., & Qannari, E. M. (2006). Common components and specific weights analysis: A chemometric method for dealing with complexity of food products. *Chemometrics and Intelligent Laboratory Systems, 81*(1), 41–49. https://doi.org/10.1016/j.chemolab.2005.09.004.

Mezquita, Y., Alonso, R. S., Casado-Vara, R., Prieto, J., & Corchado, J. M. (2020). A review of *k*-NN algorithm based on classical and quantum machine learning. In J. Kacprzyk (Ed.), *Advances in Intelligent Systems and Computing* (pp. 189–199). Poland: Springer.

Nagler, M., Nägele, T., Gilli, C., Fragner, L., Korte, A., Platzer, A., … Weckwerth, W. (2018). Eco-metabolomics and metabolic modeling: Making the leap from model systems in the lab to native populations in the field. *Frontiers in Plant Science, 9*. https://doi.org/10.3389/fpls.2018.01556.

Oliveri, P., Malegori, C., Casale, M., Tartacca, E., & Salvatori, G. (2019). An innovative multivariate strategy for HSI-NIR images to automatically detect defects in green coffee. *Talanta, 199*, 270–276. https://doi.org/10.1016/j.talanta.2019.02.049.

Passari, L. M. Z. G., Scarminio, I. S., Marcheafave, G. G., & Bruns, R. E. (2019). Seasonal changes and solvent effects on fractionated functional food component yields from *Mikania laevigata* leaves. *Food Chemistry, 273*, 151–158. https://doi.org/10.1016/j.foodchem.2017.12.075.

Pauli, E. D., Bruns, R. E., & Scarminio, I. S. (2016). UV-Vis spectral fingerprinting and chemometric method applied to the evaluation of *Camellia sinensis* leaves from different harvests. *Analytical Methods, 8*(41), 7537–7544. https://doi.org/10.1039/C6AY02462A.

Peters, K., Worrich, A., Weinhold, A., Alka, O., Balcke, G., Birkemeyer, C., … Dam, N. (2018). Current challenges in plant Eco-Metabolomics. *International Journal of Molecular Sciences, 19*(5), 1385. https://doi.org/10.3390/ijms19051385.

Qannari, E. M., Wakeling, I., Courcoux, P., & MacFie, H. J. H. (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.

Rakocevic, M., Batista, E. R., Pazianotto, R. A. A., Scholz, M. B. S., Souza, G. A. R., Campostrini, E., & Ramalho, J. C. (2021). Leaf gas exchange and bean quality fluctuations over the whole canopy vertical profile of Arabic coffee cultivated under elevated CO_2. *Functional Plant Biology*. https://doi.org/10.1071/FP20298.

Rakocevic, M., Braga, K. S. M., Batista, E. R., Maia, A. H. N., Scholz, M. B. S., & Filizola, H. F. (2020). The vegetative growth assists to reproductive responses of Arabic coffee trees in a long-term FACE experiment. *Plant Growth Regulation, 91*(2), 305–316. https://doi.org/10.1007/s10725-020-00607-2.

Rakocevic, M., & Matsunaga, F. T. (2018). Variations in leaf growth parameters within the tree structure of adult *Coffea arabica* in relation to seasonal growth, water availability and air carbon dioxide concentration. *Annals of Botany, 122*(1), 117–131. https://doi.org/10.1093/aob/mcy042.

Rosa, L. N., de Figueiredo, L. C., Bonafé, E. G., Coqueiro, A., Visentainer, J. V., Março, P. H., Rutledge, D. N., & Valderrama, P. (2017). Multi-block data analysis using ComDim for the evaluation of complex samples: Characterization of edible oils. *Analytica Chimica Acta, 961*, 42–48. https://doi.org/10.1016/j.aca.2017.01.019.

Rutan, S. C., Carr, P. W., Cheong, W. J., Park, J. H., & Snyder, L. R. (1989). Re-evaluation of the solvent triangle and comparison to solvatochromic based of solvent strenght and selectivity. *Journal of Chromatography, 463*, 21–37.

Salamanca-Neto, C. A. R., Marcheafave, G. G., Scremin, J., Barbosa, E. C. M., Camargo, P. H. C., Dekker, R. F. H., Scarminio, I. S., Barbosa-Dekker, A. M., & Sartori, E. R. (2020). Chemometric-assisted construction of a biosensing device to measure chlorogenic acid content in brewed coffee beverages to discriminate quality. *Food Chemistry, 315*, 126306. https://doi.org/10.1016/j.foodchem.2020.126306.

da Silveira, T. F. F., de Andrade Lima, M., Meinhart, A. D., Kuhnle, G. G. C., & Godoy, H. T. (2020). Effect of solvent composition on the extraction of phenolic compounds and antioxidant capacity of bacaba juice (*Oenocarpus bacaba* Mart.). *Food Analytical Methods, 13*(5), 1119–1128. https://doi.org/10.1007/s12161-020-01726-2.

Snyder, L. R., Carr, P. W., & Rutan, S. C. (1993). Solvatochromically based solvent-selectivity triangle. *Journal of Chromatography A, 656*(1–2), 537–547. https://doi.org/10.1016/0021-9673(93)80818-S.

Soares, P. K., Marcheafave, G. G., Gomes, A.d. A., Scarminio, I. S., & Bruns, R. E. (2018). Mixture Design PARAFAC HPLC-DAD metabolomic fingerprints of fractionated organic and basic extracts from *Erythrina speciosa* Andrews leaves. *Chromatographia, 81*(8), 1189–1200. https://doi.org/10.1007/s10337-018-3554-9.

Tormena, C. D., Marcheafave, G. G., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2019a). Sequential mixture design optimization for divergent metabolite analysis: Enriched carbon dioxide effects on *Coffea arabica* L. leaves and buds. *Talanta, 191*, 382–389. https://doi.org/10.1016/j.talanta.2018.09.002.

Tormena, C. D., Marcheafave, G. G., Pauli, E. D., Bruns, R. E., & Scarminio, I. S. (2019b). Potential biomonitoring of atmospheric carbon dioxide in *Coffea arabica* leaves using near-infrared spectroscopy and partial least squares discriminant analysis. *Environmental Science and Pollution Research, 26*(29), 30356–30364. https://doi.org/10.1007/s11356-019-06163-1.

Tormena, C. D., Pauli, E. D., Marcheafave, G. G., Scheel, G. L., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2020). FT-IR biomarkers of sexual dimorphism in yerba-mate plants: Seasonal and light accessibility effects. *Microchemical Journal, 158*, 105329. https://doi.org/10.1016/j.microc.2020.105329.

Vieira, T. F., Makimori, G. Y. F., dos Santos Scholz, M. B., Zielinski, A. A. F., & Bona, E. (2020). Chemometric approach using ComDim and PLS-DA for discrimination and classification of commercial Yerba Mate (Ilex paraguariensis St. Hil.). *Food Analytical Methods, 13*(1), 97–107. https://doi.org/10.1007/s12161-019-01520-9.