The main effects of elevated CO₂ and soil-water deficiency on ¹H NMR-based metabolic fingerprints of Coffea arabica beans by factorial and mixture design

Gustavo Galo Marcheafave a,⁎, Cláudia Domiciano Tormena a, Lavínia Eduarda Mattosa a, Vanessa Rocha Liberatti b, Anna Beatriz Sabino Ferrari b, Miroslava Rakocevic c,d, Roy Edward Bruns e, Ieda Spacino Scarminio a,⁎, Elís Daiane Pauli a

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, State University of Londrina, CP 6001, 86051-990 Londrina, PR, Brazil
c Northern Rio de Janeiro State University - UENF, Plant Physiology Lab, Av. Alberto Lamego 2000, 28013-602 Campos dos Goytacazes, RJ, Brazil
d Embrapa Environment, Rodovia SP 340, Km 1275, 13820-000 Jaguariúna, SP, Brazil
e Institute of Chemistry, State University of Campinas, CP 6154, 13083-970 Campinas, SP, Brazil

HIGHLIGHTS
- ¹H NMR C. arabica bean fingerprint changes evaluated by chemometric methods.
- Elevated CO₂ levels provoke significant changes in ¹H NMR C. arabica bean fingerprints.
- Chemometric models find key metabolites owing to carbon dioxide increase.
- Coffee tree vertical profile does not affect bean classification.
- Strategy to achieve global chemical profile of food has been demonstrated.

GRAPHICAL ABSTRACT

ABSTRACT

The metabolic response of Coffea arabica trees in the face of the rising atmospheric concentration of carbon dioxide (CO₂) combined with the reduction in soil-water availability is complex due to the various (bio)chemical feedbacks. Modern analytical tools and the experimental advance of agronomic science tend to advance in the understanding of the metabolic complexity of plants. In this work, Coffea arabica trees were grown in a Free-Air Carbon Dioxide Enrichment dispositive under factorial design (2²) conditions considering two CO₂ levels and two soil-water availabilities. The ¹H NMR mixture design-fingerprinting effects of CO₂ and soil-water levels on beans were strategically investigated using the principal component analysis (PCA), analysis of variance (ANOVA) - simultaneous component analysis (ASCA) and partial least squares-discriminant analysis (PLS-DA). From the ASCA, the CO₂ factor had a significant effect on changing the ¹H NMR profile of fingerprints. The soil-water factor and interaction (CO₂ × soil-water) were not significant. ¹H NMR fingerprints with PCA, ASCA and PLS-DA analysis determined spectral profiles for fatty acids, caffeine, trigonelline and glucose increases in beans from current CO₂, while quinic acid/chlorogenic acids, malic acid and kahweol/cafestol increased in coffee.
1. Introduction

The variation in the earth’s climate is dependent on several multivariate processes that involve the lithosphere, hydrosphere and atmosphere (Ainsworth et al., 2020). In the last decades, serious climate change has been occurring, partially due to carbon dioxide (CO₂) emissions associated with industrial activity (Cernusak et al., 2019). The consequences of climate changes, such as more frequent drought periods, elevated temperatures and atmospheric gas elevations, have been investigated (Bazzaz, 1990; McGrath and Lobell, 2013; Nagelkerken and Connell, 2015; Rajashekar, 2018). Concerns in raising CO₂ emission are investigated (Bazzaz, 1990; McGrath and Lobell, 2013; Nagelkerken and Connell, 2015; Rajashekar, 2018). In raising the interactions between various environmental factors and thus, the environmental stresses such as CO₂, water and temperature under controlled field systems (Pleijel and Högy, 2015; Rakocевич et al., 2018). Some studies have investigated the chemical responses (i.e. metabolic changes) of cultivation systems under environmental disturbances, although they explored the effects of environmental changes (e.g. limited nutrient stress, high temperature, elevated CO₂) using only metabolic quantifications (Miyagi et al., 2017; Tenne et al., 2019). To study only a few metabolites often complicates understanding the global response of the metabolic system of these plants and does not completely define the relationship between the various environmental perturbances that modified plant organ functioning. To solve those challenges, multivariate fingerprint methodologies have been developed and improved in recent decades as a metabolomic tool through spectroscopic, spectrometric and chromatographic techniques (Alonso et al., 2015; Gallo and Ferranti, 2016).

The plant metabolome is widely considered to be the most complex in the living world, due to the presence of primary and especially secondary metabolites (Dunn and Ellis, 2005). The metabolomic fingerprint of proton nuclear magnetic resonance (¹H NMR) spectroscopy offers a wide variety of metabolites and provides an instant metabolic signature to a plant extract (Kim et al., 2010). This signature can be related to the different climatic conditions to which they are submitted during their plant development. Owing to the great advantages inherent in the NMR technique (Pontes et al., 2017), it is used as one of the main methods to define quality control (Gall et al., 2004; Minoja and Napoli, 2014); search for markers (Gebregiorgis and Powers, 2012); fingerprinting (Locci et al., 2011; Porzel et al., 2014) among other applications. In ¹H NMR-based metabolic fingerprinting, the larger the number of chemical species present in the signature, robustness increases and the experimental domain of metabolic analysis under different cultivation conditions of plant species becomes more informative.

Our research group implemented statistical mixture design as a sample preparation tool for fingerprint extraction coupled with various spectroscopic, spectrometric and chromatographic techniques (Marcheafave et al., 2020a, 2020b; Soares et al., 2018). The variation in the chemical properties of pure solvents and their mixtures provides fingerprints with different characteristics and when combined they can better represent the instantaneous condition of the plant metabolome when compared to a single solvent or mixture, resulting in more complex data sets (Alonso et al., 2019; Tormena et al., 2019a). After the acquisition of fingerprints according to a mixture design the complex metabolomic data set contains a large amount of chemical shifts. To determine the effects of environmental factors and their interactions on metabolome expression multifactorial experiments in the cultivation scenario are needed which increase the size of variations within the data set (Jansen et al., 2005; Zwanenburg et al., 2011). To define the environmental factors that mainly impact the chemical composition of the system, the analysis of variance (ANOVA) - simultaneous component analysis (ASCA) is used as an exploratory tool (Jansen et al., 2005). Additionally, supervised methods, such as partial least squares-discriminant analysis (PLS-DA), can be applied to the data set, classifying samples owing to significant effect history and reducing the complexity of the mathematical modeling. Thus, the set of all these tools (statistical mixture design, NMR fingerprints, ASCA and PLS-DA) overcomes a series of difficulties encountered in the process of detecting chemical changes in complex metabolic systems, expanding the range of variability of chemical profiles, statistically determining significant effects and interactions and properly classifying chemical changes due to climate change.

For this study, Coffea arabica L. (Arabica coffee) green beans were collected in the fourth production year under the FACE experiment in a factorial design (2²) with atmospheric CO₂ and soil-water availability levels. With the aid of statistical mixture designs, NMR fingerprints, PCA, ASCA and PLS-DA, significant changes in coffee beans due to drought and elevated CO₂ levels can be reliably determined. The main objectives here are to 1) apply mixture design as a sample preparation tool to increase the variability of ¹H NMR fingerprints; 2) use PCA and ASCA to explore and quantitatively determine the statistical significance of the main and interaction effects on the ¹H NMR fingerprints of C. arabica beans for the factorial design experiments involving two atmospheric CO₂ levels and two soil-water availabilities; 3) detect the main metabolic changes in C. arabica beans under drought and elevated CO₂ conditions from the classification by PLS-DA and 4) test the prediction ability of this PLS-DA model using the variability of the extracting solvents and different berry locations (containing coffee beans) over the vertical profile of coffee trees.
2. Experimental

2.1. Coffee bean collection

The FACE device with Arabic coffee was located in Embrapa Environment, Jaguariúna (22°43′S, 47°01′W, 570 m of altitude), São Paulo, Brazil. The FACE experiment was carried at 7 ha plantations and was limited to 12 octagon plots. Each octagon contained four rows summing 44 plants for each plot (Marcheafave et al., 2020a; Tormena et al., 2019a; Tormena et al., 2019b). Experiments were conducted between August 25, 2011 until June 30, 2016. The injection of pure CO2 was limited to six octagon plots, maintaining about 590 μL L−1, which was compared to six octagons maintained under the current CO2 conditions (390 μL L−1) (Fig. S1B). In the last production year of the experiment, half of each octagon was irrigated (started in October 2015 to the end of May 2016) and a half was maintained under the rainfall conditions (considered drought conditions), dependent on vapor pressure deficit and rain distributions (Fig. S1A).

Each coffee berry can contain one or two grains. In the fourth production year of Arabica coffee cultivation under the FACE (2015–2016), the berries were collected in the second week of May 2016 (tagged in Fig. S1A), over 50 cm thick layers of the coffee tree vertical profiles (Rakocevic and Matsunaga, 2018). The berry collection contained cherry and green berries together (Rakocevic et al., 2020). Berries were dried and benefited resulting in samples of well-formed beans.

To evaluate whether the effects and/or interaction of two climate factors on the metabolic fingerprint of green beans of C. arabica are statistically significant, the experiments were executed following a two-level full factorial design with two factors (22) (Hibbert, 2012; Neto et al., 2010) (Table S1). The factors were carbon dioxide levels (1) and soil-water supply (2). The carbon dioxide levels are current (−1) and elevated (1) irrigation.

2.2. Reagents

For extraction, hexane, ethanol, dichloromethane (Anidrol, Diadema, Brazil) and ethyl ether (Vetcet, Rio de Janeiro, Brazil) were of analytical grade. For NMR analysis dimethyl sulfoxide-d6 (DMSO-d6, Sigma-Aldrich, St. Louis, USA) with tetramethylsilane (TMS) as an internal reference was used.

2.3. Preparation of the extracts

To maximize the extraction of metabolic classes and obtain a set of fingerprints that determine the maximum variations resulting from environmental change, a statistical mixture design with four components consisting of ethanol (E), dichloromethane (D), hexane (H) and ethyl ether (e) was used to plan the experimental process (Fig. 1). To diversify the chemical interactions between the plant matrix and the extracting solvent, the solvent selection is based on Snyder’s solvent-selectivity triangle (Barwick, 1997). The extraction was carried out according to Marcheafave et al. (2020b) with minor modifications. The extractions correspond to four growing conditions of the coffee plants according to Table S1. The grains were collected from the total vertical profile of the plant and homogenized. The central point of the mixture design was carried out in triplicate, resulting in 17 extracts per cultivation condition. The extractive procedure took place with 2.50 g of the chemical (considered drought conditions), dependent on vapor pressure deficit and rain distributions (Fig. S1A).

Each coffee berry can contain one or two grains. In the fourth production year of Arabica coffee cultivation under the FACE (2015–2016), the berries were collected in the second week of May 2016 (tagged in Fig. S1A), over 50 cm thick layers of the coffee tree vertical profiles (Rakocevic and Matsunaga, 2018). The berry collection contained cherry and green berries together (Rakocevic et al., 2020). Berries were dried and benefited resulting in samples of well-formed beans.

To evaluate whether the effects and/or interaction of two climate factors on the metabolic fingerprint of green beans of C. arabica are statistically significant, the experiments were executed following a two-level full factorial design with two factors (22) (Hibbert, 2012; Neto et al., 2010) (Table S1). The factors were carbon dioxide levels (1) and soil-water supply (2). The carbon dioxide levels are current (−1) and elevated (1) irrigation.

2.4. NMR measurements

For NMR analysis, 0.05 g of lyophilized extract was dissolved by a vortex mixer in 0.5 mL of DMSO-d6 with TMS at 25 °C. Then, the solubilized extract (0.0001 g L−1) was put into a 5 mm NMR tube. The one-dimensional 1H 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 1H 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.

2.5. Spectral processing and statistical analysis

Two-level full factorial design with two factors (22) (Hibbert, 2012; Neto et al., 2010) was performed, considering two CO2 levels (current −1, and elevated 1) and two soil-water supplies (without −1 and with 1 irrigation).

For multivariate analysis, each 1H NMR spectrum in the range between 0.0177 and 10.9997 ppm was segmented into 0.0006-ppm chemical shift bins. 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. So, the dataset was organized in a matrix of 68 extracts (17 extracts for each environmental condition listed in Table S1) × 17,590 ppm obtained by NMR. The chemometric tools were applied to the data matrix containing the spectral fingerprints, but before building models the data were preprocessed appropriately for each multivariate method. The spectra were Pareto scaled for principal component analysis, PCA (Sarabhai and Centre, 1990) and analysis of variance (ANOVA) - simultaneous component analysis (ASCA) (Jansen et al., 2005; Smilde et al., 2005). For the partial least squares-discriminant analysis (PLS-DA) (Barker and Rayens, 2003) the spectra were normalized (inf-Norm, maximum = 1). After that, the Orthogonal Signal Correction (OSC) filter was used (Wold et al., 1998). The spectra were mean-centered previous any multivariate analysis. The experimental and chemometric analysis strategy is summarized in Fig. 2.

PCA was used to explore the adaptation of C. arabica plants to the effects of the CO2 and soil-water supply factors, based on beans...
metabolome. PCA is a valuable tool for exploration revealing dominant sources of variance in the observed data, being frequently used for metabolic fingerprinting (Chen et al., 2008; Worley and Powers, 2013). This method provides a preliminary understanding of the effects provoked on many classes of metabolites and allowing visualization of pertinent data set information with only a few principal components.

ASCA is an exploratory technique for the analysis of multivariate data resulting from an experimental design (Jansen et al., 2005; Luca et al., 2016; Olmos et al., 2019; Smilde et al., 2005). ASCA uses an ANOVA model to first decompose the observed data matrix into a series of additive effect matrices, according to the experimental design. Then a PCA model is fit to each effect matrix to extract and assess the contribution of each experimental factor (Saccenti et al., 2018; Timmerman et al., 2015). So, an ASCA model can be explored and interpreted like a standard PCA model (Saccenti et al., 2018). In this research, the ASCA was applied to statistically assess the significance of the effects of experimental design factors (carbon dioxide levels and soil-water availability) and their interaction on spectral changes of the C. arabica beans. To estimate the significance of each factor, the p-values were calculated based on the permutation test (Vis et al., 2007). More information can be found at Hoefsloot et al. (2009).

PLS-DA is a discriminant chemometric technique which establishes a linear regression between an independent matrix (X) that are the analytical responses, and dependent matrix or vector (Y or y, respectively) that contains binary dummy variables indicating the class to which each sample belongs (Barker and Rayens, 2003; Bevilacqua et al., 2013; Miaw et al., 2018). In the dependent vector (y), as in the case this study, the value 1 is assigned to the sample that belongs to a certain class and the value 0 to the other class. PLS-DA was used to discriminate the 1H NMR spectra between C. arabica beans cultured under elevated and current atmospheric CO2 and to evaluate the effect of this factor. In this work, sensitivity, specificity, accuracy, efficiency and Matthews correlation coefficient (MCC) were used to evaluate the obtained classification models. More details can be found in Oliveira et al. (2020), Garreta-Lara et al. (2018) and Chao et al. (2014). The selection of the optimal number of latent variables in PLS-DA was done using the criterion of lowest root mean square error of cross-validation (RMSECV) for the calibration samples obtained by venetian blinds procedure.

2.6. Test of prediction ability PLS-DA model

We tested the prediction ability of the PLS-DA model for the effects of two CO2 levels and two soil-water availabilities on coffee bean metabolic profiles. This test was performed by using one dataset that had not been a part of the initial dataset used to build the PLS-DA model. The data set from samples of four 50 cm-thick-layers of the vertical profile of the coffee tree plants, as a test to evaluate the environmental variability inherent to the localization of production over the plant architecture.

Using the constructed PLS-DA predictive model, 22 new extractions were performed using only beans collected from C. arabica plants grown in the current CO2 level. These extractions were performed by randomly selecting the extraction solvents from the mixture design to assess the effect of the solvent on the correct classification of these samples by 1H NMR fingerprints. Information about layers of coffee beans and solvents used are shown in Table S2.

2.7. Statistical analyses

Data treatment and chemometric analysis were performed with Matlab 2016b® software (The MathWorks Inc., Natick, USA), using tools of PLS toolbox 8.7® (Eigenvector Research Inc., Manson, WA, USA).

3. Results and discussion

Two CO2 levels and two soil-water availabilities were initially explored by the unsupervised analysis method, in this case by PCA, performed on 1H NMR spectra to observe the natural differences among the C. arabica beans and the relationships between variables and classes. As shown in Fig. 3a principal component 1 (PC1) explained 37.04% of the total variance, and PC3 accounted for 10.52% of the variance. It was not possible to make a distinction between the cultivation experimental factors studied according to PC1, maybe due to different extracted metabolite proportions according to solvents, their characteristics, and/or natural variation among the different samples. Along PC3 the formation of two clusters can be found, a metabolite group characterizing most C. arabica coffee beans from elevated CO2 on the negative side, and current level CO2 on the positive side.

PCA score results showed that the CO2 factor had relevance on the 1H NMR spectral profile of C. arabica beans, and this was due to changes in metabolite responses (Fig. 3b). The signals at 1.16; 1.22; 3.22; 3.87; 5.18–5.19 ppm; and 8.00 ppm had positive values along PC3 as shown in Fig. 3b and were related to beans of C. arabica plants cultivated under current CO2. The region of peaks around 1.16 and 1.22 ppm can be associated with the presence of fatty acids (Barison et al., 2010; Consomni et al., 2012; Knothe and Kenar, 2004; Toci et al., 2018), which are a common constituent in coffee beans (Scholz et al., 2018). Some peaks present in the original spectra (not shown) reinforced the presence of fatty acids, such as those at 0.9; 2.3 and 2.7 ppm (Barison et al., 2010; Knothe and Kenar, 2004). The peaks in the region between 3.22 and 3.87 ppm suggested a caffeine profile (Bosco et al., 1999; del
Campo et al., 2010; Toci et al., 2018), as the original spectra had a peak at 3.50 ppm, possibly produced by the three N-methyl group. Furthermore, the peak at 7.85 ppm probably corresponds to an aromatic proton (del Campo et al., 2010). The loading peak at 8.00 ppm (Fig. 3b), may correspond to trigoneline, being supported by the presence a triplet in the region at 8.00 ppm, two doublets at 8.78 and 8.87 ppm and a singlet at 9.21 ppm, corresponding to aromatic protons, while the signal around 4.36 ppm was related to N-methyl groups observed in the original spectra of some solvents (del Campo et al., 2010; Toci et al., 2018). The peaks at 5.18–5.19 ppm can be due to glucose (Abreu et al., 2018; D’Agostina et al., 2004; Hatatzakis, 2019; Wei et al., 2014), since this region is typical of the anomeric hydrogens (D’Agostina et al., 2004).

The PC3 loadings plot showed the bands that contributed to the grouping of C. arabica beans obtained from plants grown under elevated CO\textsubscript{2} (Fig. 3b). Negative loading values pointed to the signals at 2.08; 2.54; and 5.75 ppm. In this case, the peak at 2.08 ppm indicated compounds belonging to the quinic acid/chlorogenic acid classes (CGA) (Consonni et al., 2018; Wei et al., 2010, 2011). The signal at 2.54 ppm also can be related to organic acids, such as malic (Aguilera-Sáez et al., 2019; Wei et al., 2010), which is present in green coffee (Clarke and Vitzthum, 2001). Finally, the signal at 5.75 ppm can be attributed to diterpenes like kahweol or cafestol, present in C. arabica beans (Medina et al., 2017; Monakhova et al., 2015; Okaru et al., 2020; Toci et al., 2018). This indication was reinforced by the resonance chemical shift (4.43 and 8.00 ppm) (del Campo et al., 2010; Toci et al., 2018), and beyond the peaks at 5.31 and 6.28 ppm, which suggests compounds belonging to the chlorogenic acid class (Wei et al., 2010). For the C. arabica beans obtained from plants cultivated under elevated CO\textsubscript{2}, the 1\textsuperscript{H} NMR signals were related to quinic acid/CGAs (2.08 ppm) (Consonni et al., 2018; Wei et al., 2010, 2011); malic acid (2.54 ppm) (Wei et al., 2010) and diterpenes (5.75 ppm) (Medina et al., 2017; Monakhova et al., 2015; Okaru et al., 2020). These are essentially the same loadings as given by PCA.

To verify whether two experimental factors and their binary interaction have a significant effect in the changes in the 1\textsuperscript{H} NMR spectral fingerprints, a two-way ASCA model with interaction was performed. The variability in the spectral matrix was partitioned by ASCA, being that the total variance of the spectroscopic signal consisted of 5.1% for the CO\textsubscript{2} factor, 1.2% for the soil-water availability factor, 1.6% for the interaction (CO\textsubscript{2} × water), and 92% for the residuals (Table 1). The model showed that the CO\textsubscript{2} factor was significant with p-value equal 0.0042, based on the permutation test, which used 10,000 randomizations. Soil-water availability and the carbon dioxide × soil-water availability interaction were not significant, with p-values larger than 0.05, as showed in Table 1.

The score and loading plots of PCA submodel, obtained by ASCA analysis, corresponding C. arabica beans from plants cultivated under the two CO\textsubscript{2} levels is given in Fig. 4 and showed a very good separation of results for the two CO\textsubscript{2} levels. Most of the scores relative to coffee beans obtained with the current CO\textsubscript{2} level had positive values, while the score of those grown under elevated CO\textsubscript{2} showed negative values (Fig. 4a). The model loadings (Fig. 4b) can be used to relate changes in certain 1\textsuperscript{H} NMR spectral regions to atmospheric CO\textsubscript{2} factor effects, i.e in the most sensitive spectral variables to this effect. The most relevant signals, according to loading values for C. arabica beans obtained from plants cultivated under current CO\textsubscript{2} atmosphere, were in regions associated the fatty acids (1.16 and 1.22 ppm) (Barison et al., 2010; Consonni et al., 2012; Knothe and Kenar, 2004; Toci et al., 2018); caffeine (3.22 and 3.87 ppm) (Bosco et al., 1999; del Campo et al., 2010; Toci et al., 2018); glucose (5.18–5.19 ppm) (Abreu et al., 2018; D’Agostina et al., 2004; Hatatzakis, 2019; Nunes et al., 2008; Wei et al., 2014); trigonelline (4.43 and 8.00 ppm) (del Campo et al., 2010; Toci et al., 2018), and beyond the peaks at 5.31 and 6.28 ppm, which suggests compounds belonging to the chlorogenic acid class (Wei et al., 2010). For the C. arabica beans obtained from plants cultivated under elevated CO\textsubscript{2}, the 1\textsuperscript{H} NMR signals were related to quinic acid/CGAs (2.08 ppm) (Consonni et al., 2018; Wei et al., 2010, 2011); malic acid (2.54 ppm) (Wei et al., 2010) and diterpenes (5.75 ppm) (Medina et al., 2017; Monakhova et al., 2015; Okaru et al., 2020). These are essentially the same loadings as given by PCA.

### Table 1

<table>
<thead>
<tr>
<th>PC</th>
<th>Effect (%)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO\textsubscript{2} levels</td>
<td>1</td>
<td>5.1</td>
</tr>
<tr>
<td>Soil-water availability</td>
<td>1</td>
<td>1.2</td>
</tr>
<tr>
<td>Interaction CO\textsubscript{2} × water</td>
<td>1</td>
<td>1.6</td>
</tr>
<tr>
<td>Residuals</td>
<td>92.0</td>
<td></td>
</tr>
</tbody>
</table>

*Residuals* 92.0

Based on PCA and ASCA models, the supervised pattern recognition method PLS-DA was applied to correlate the changes of 1\textsuperscript{H} NMR spectra of the beans obtained from plants cultivated under the two CO\textsubscript{2} levels, to further check bands coinciding with those indicated by the previously mentioned models. Four latent variables (LV) were necessary to build the optimal model, representing 66.9% of the variance explained in the X and 80.6% of the y-variance. A good classification performance was achieved with sensitivity (rate of true positives) and specificity (rate of true negatives) of 100% for both classes. Accuracy, which shows the proportion of samples identified correctly, which on the tests sets
Figure 4. ASCA analysis of $^1$H NMR fingerprints of Coffea arabica beans grown in a Free-Air Carbon dioxide Enrichment (FACE) and delineated in a factorial design with two CO$_2$ levels (current and elevated) and two soil-water availabilities (with and without irrigation). A) Score plot for the effect of CO$_2$ concentration with projected residuals and B) PCI loading plot.

were equal to 100%, as well as the efficiency index that combines all information carried by the sensitivity and specificity. Matthews correlation coefficient gives an overview of model behavior and was equal to 1.0, showing very good classification for the samples. Validation results, by cross-validation, were also satisfactory once the classification performance did not decreased significantly. In other words, sensitivity equal to 93.9% and 100% for current and elevated classes, respectively; accuracy and efficiency equal or greater than 97% and MCC equal to 0.94.

PLS-DA scores of the first latent variable showed the formation of two distinct groups. The C. arabica beans obtained from plants cultivated under elevated CO$_2$ conditions have positive values, while those at the current level are negative (Fig. 5a). The loading plot shows the $^1$H NMR regions that contributes to the differentiation between classes (Fig. 5b). Signals relative to quinic acid/CGAs (2.08 ppm) (Consonni et al., 2018; Wei et al., 2010); malic acid (2.54 ppm) (Wei et al., 2010) and kahweol (5.75 ppm) (Medina et al., 2017; Monakhova et al., 2015; Okaru et al., 2020) characterized C. arabica beans from plants cultivated under elevated CO$_2$ (Fig. 5b). The same metabolites indicated by PCA and ASCA also were responsible for discrimination of C. arabica beans from plants grown under current CO$_2$, i.e. the fatty acids (1.16 and 1.22 ppm) (Barison et al., 2010; Consonni et al., 2012; Knothe and Kenar, 2004; Toci et al., 2018), caffeine (3.22, 3.50 and 3.87 ppm) (Bosco et al., 1999; del Campo et al., 2010; Toci et al., 2018), glucose (5.18–5.19 ppm) (Abreu et al., 2018; D’Agostina et al., 2004; Hatzakis, 2019; Nunes et al., 2008; Wei et al., 2014) and trigonelline (4.43 and 8.00 ppm) (Barison et al., 2010; Toci et al., 2018). The chemical variable importance in projection (VIP) scores above the 1.0 threshold contributed significantly to the model (Fig. S2). Among the various regions pointed out by the VIP scores, most were also pointed out by the PLS-DA loadings.

New samples were prepared and investigated to test the prediction ability of the PLS-DA model for the two CO$_2$ levels and two soil-water availabilities. All samples were classified as grown under current CO$_2$, regardless of the extraction solvents and the vertical plant profile in which the beans were collected (Fig. 6). In addition, the model also showed the importance of the CO$_2$ factor, and that had good parameters for classification.

As known, the increase in atmospheric CO$_2$ can act to mitigate the effects of abiotic stressors, such as elevated temperature or drought (Marcheafave et al., 2020b; Ramalho et al., 2018). In C. arabica plants, the elevated atmospheric CO$_2$ influence rates of leaf-photosynthesis (Damatta et al., 2016; Rakovec et al., 2018), leaf area (Rakovec and Matsunaga, 2018), stomatal conductance (Ghini et al., 2015; Rakovec et al., 2018), water use efficiency (Rakovec et al., 2018), among others, but it is worth mentioning that different coffee genotypes can respond in a particular way concerning this effect (Ghini et al., 2015; Martins et al., 2016; Ramalho et al., 2013; Scotti-Campos et al., 2019). Specifically in raw coffee beans, there is still little data about the impact in the metabolic changes in response to CO$_2$ enrichment compared to the current level. Also, few researchers have found a CO$_2$ mitigation effect on drought in the chemical composition of C. arabica beans. Recently, the interaction between CO$_2$ atmospheric concentrations and soil-water supply factors was detected, which can be configured as mitigation of drought, through spectral metabolic fingerprints in the UV region for C. arabica beans (Marcheafave et al., 2020b). In this spectral region, bands of chromophoric groups are strongly overlapped and few metabolites can be detected or characterized based on spectral profiles of chemical species. As few chemical species were detected by the UV profile, probably, these were principally responsible for the mitigation effect. Although the interaction between the supplement of water and CO$_2$ is important in the process of mitigating drought in coffee plants, it was not evidenced by the $^1$H NMR results. The $^1$H NMR spectral analysis, has the capacity to analyze the identity, structure, concentration and behavior of numerous molecules simultaneously due to its wide window of chemical shifts. Compared to UV spectra of extracts from ethanol-dichloromethane binary mixture and ethanol solvent (Marcheafave et al., 2020b), the $^1$H NMR mixture design fingerprints provide a greater variety of information about metabolite classes. Perhaps for this reason the compounds corresponding to the NMR signals, discussed in this research, responded significantly to the CO$_2$ effect but not for water and its interaction (CO$_2 \times$ water).

Distinct chemometric methods applied to the $^1$H NMR spectra showed that atmospheric CO$_2$ had a prominent effect on the metabolic changes in raw coffee beans. The $^1$H NMR chemical shifts with pronounced loading values (Figs. 3b, 4b, 5b) provided by PCA, ASCA and PLS-DA, showed coincident spectral bands for the discrimination of coffee beans at the two CO$_2$ levels. This suggests that the corresponding metabolites, quinic acid/CGAs, malic acid and kahweol, could be used as potential markers of the effects produced by an elevated CO$_2$ level.

Characteristic chemical shifts for caffeine and trigonelline of C. arabica beans were found for plants cultivated under current CO$_2$, indicating that these beans can provide more energetic beverages compared to those from elevated CO$_2$ conditions. The amount of caffeine and trigonelline was also lower in beans from plants cultivated under elevated CO$_2$ and exposed to elevated temperatures in the last stages.
of fruit maturation (Ramalho et al., 2018). Both metabolites contribute to the overall aroma perception of both roasted coffee beans and a brewed coffee beverage, being that caffeine influences the perceived strength, body and bitterness of brewed coffee (Sunarharum et al., 2014). Although trigonelline is also a precursor of flavor/aroma compounds in coffee and undergoes significant degradation during roasting, high contents of trigonelline and its volatile derivatives, especially pyroles, have been reported for coffee of low quality (Farah et al., 2019).

The fatty acid signals also discriminated C. arabica beans collected from plants grown under current CO2, whereas fatty acids are not prevalent in beans grown in an elevated CO2 atmosphere. The increased participation of green berries and lower lipid concentrations (in which fatty acids are a part) in whole berry production is detected under elevated CO2 compared to normal levels (Rakocevic et al., 2020), suggesting delayed flowering flushes or prolonged ripening.

Finally, the glucose region was also important for discrimination with positive signals for beans from plants grown under current CO2 (Figs. 3b, 4b, 5b). Under elevated CO2, the most important effect is the increase in the concentration of total non-structural carbohydrates (starch, fructan, sucrose, etc.) in the leaves of various plant species (Poorter et al., 1997). These carbohydrates are used to produce all C-skeletons, including primary and secondary metabolites in plants, by using glucose necessary for C-skeletons and also the glucose required for NAD(P)H and ATP production to drive the various biosynthetic reactions. Free glucose appeared to be more abundant in coffee beans at current rather than elevated CO2 levels, which could be attributed to higher glucose required for NAD(P)H and ATP production in advanced berry ripening (Rakocevic et al., 2020).

The $^1$H NMR region for quinic acid, indicated by loading values, for beans grown under elevated CO2 levels may be the result of the ripening
process (Figs. 3b, 4b, 5b). The presence of this quinic acid signal can indicate the presence of compounds with similar structures, such as the chlorogenic acid class, already known and well established in raw coffee beans (Zanin et al., 2016). Coffee beans from plants grown under elevated CO2 had delayed maturation when compared to beans grown at the current CO2 level (Rakocievic et al., 2020), which would justify a greater sensitivity of these bands in samples grown under this atmospheric condition. Malic acid also discriminated against the coffee beans cultivated under elevated CO2. This organic acid is present in the stages of development of the coffee beans and during maturation its concentration decreases (Rogers et al., 1999). The fact that beans under elevated CO2 have a delay in maturation (Rogers et al., 1999), could explain the greater presence of malic acid. As the peak at 2.54 ppm is very weak in the original spectra (not shown) of beans grown under current CO2, this would reinforce the hypothesis of maturation delay. About diterpenes, quantitative studies involving their formation under elevated CO2 conditions are scarce. C. arabica bean plants grown under elevated CO2 without irrigation are discriminated by bands of cafestol and kahweol in the UV region, while, when irrigated, these have superior loadings in beans from plants grown under current CO2 (Marcheafave et al., 2020b). Our results investigating 1H NMR fingerprints demonstrated that the highest loadings for cafestol and kahweol are found in beans from plants grown in elevated CO2, independent of water availability.

The different applications of chemometric tools showed coincident spectral regions. Each signal/region corresponds to a corresponding metabolite or class, and as is already known, the applied solvents have characteristics that may or may not extract certain metabolites and their combination may have synergistic or antagonistic effects, depending on the plant matrix in question (Garcia et al., 2010; Guizzellini et al., 2018; Marcheafave et al., 2019; Moreira and Scarminio, 2013; Pauli et al., 2016). When working with spectral fingerprints, the range of metabolites analyzed can be greater, depending on the solvents used, or even certain ones can attenuate some spectral regions. This indicates that this type of extraction study can allow an analysis richer in responses. In our research, we can see that some solvents or combinations were highlighted with certain metabolites. In the case of extracts obtained by pure ethyl ether and their mixtures without the presence of ethanol, more selective spectra for fatty acids are found, as well as for extracts with hexane. Extracts obtained from pure dichloromethane showed selectivity for fatty acids and methylxanthines. Spectra from ethanol were less selective, showing a wide class of extracted metabolites and synergism of their interactions with solvents such as dichloromethane and hexane. When working with a variety of solvents one probably obtains greater spectral coverage than using standard solvents for which certain signals could pass detected.

4. Conclusions

Changes in 1H NMR-based metabolic fingerprints of C. arabica beans can be found for the exploration of the effects of elevated CO2 and soil-water availability by factorial design experiments. PCA and ASCA models provided evidence that atmospheric CO2 had a significant and prominent effect on the metabolic changes in raw coffee beans compared with the effect of water and the CO2–water interaction, which was confirmed by the PLS-DA method. These three chemometric methods showed that fatty acids, caffeine, trigonelline and glucose discriminate C. arabica beans collected from plants grown under current CO2 levels. Quinic acid/chlorogenic acids, malic acid and kahweol/ cafestol were important for coffee beans under an elevated CO2 atmosphere, probably due to the delayed ripening process, and could be used as potential markers of the effects produced by this condition. Furthermore, the PLS-DA model showed good prediction ability regardless of the extraction solvents and the vertical plant profile in which the beans were collected. Finally, this research showed that when working with the statistical solvent mixture design, the range of metabolites analyzed can be greater with some metabolites being attenuated and others detected, allowing analysis of more informative responses.

CRediT authorship contribution statement

Gustavo Galo Marcheafave: Conceptualization, Methodology, Investigation, Formal analysis, Writing - original draft. Cláudia Domiciano Tormena: Methodology, Investigation, Writing - original draft. Lavinia Eduarda Mattos: Methodology. Vanessa Rocha Liberatti: Methodology. Anna Beatriz Sabino Ferrari: Methodology. Miroslava Rakocievic: Conceptualization, Methodology, Writing - review & editing. Roy Edward Bruns: Writing - review & editing. Ieda Spacino Scarminio: Conceptualization, Resources, Supervision, Writing - review & editing, Project administration, Funding acquisition. Elis Daiane Pauli: Conceptualization, Formal analysis, Software, Writing - original draft.

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.


