UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIENCIAS EXATAS
DEPARTAMENTO DE QUÍMICA

Daphne Chiara Antônio

Emprego de ferramentas quimiométricas multidimensionais na análise de mel e café

Use of multiway chemometric tools for honey and coffee analysis

Belo Horizonte
2021

UFMG/ICEX/DQ. 1.472
T. 668

Daphne Chiara Antônio

Emprego de ferramentas quimiométricas multidimensionais na análise de mel e café

Use of multiway chemometric tools for honey and coffee analysis

Tese apresentada ao Departamento de Química do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do grau de Doutor em Ciências – Química.

Orientador: Marcelo Martins de Sena
Coorientador: Bruno Gonçalves Botelho

Belo Horizonte
2021

Ficha Catalográfica

A627e
2021
T

Antônio, Daphne Chiara
Emprego de ferramentas quimiométricas multidimensionais na análise de mel e café [manuscrito] / Daphne Chiara Antônio. 2021.
95 f. : il., gráfs., tabs.

Orientador: Marcelo Martins de Sena.
Coorientador: Bruno Gonçalves Botelho.

Tese (doutorado) – Universidade Federal de Minas Gerais – Departamento de Química.
Inclui bibliografia.

1. Química analítica – Teses. 2. Análise multivariada - Teses. 3. Alimentos – Análise – Teses. 4. Mel - Teses. 5. Café – Teses. 6. Preparação de amostra (Química) - Teses. 7. Espectroscopia de fluorescência - Teses. 8. Fenilalanina – Teses. 9. Alimentos – Avaliação sensorial – Teses. I. Sena, Marcelo Martins de, Orientador. II. Botelho, Bruno Gonçalves, Coorientador. III. Título.

CDU 043

Elaborada por Sérgio Ferreira da Silva – CRB6-2719.

UNIVERSIDADE FEDERAL DE MINAS GERAIS

UFMG
Programa de Pós-Graduação em Química
Departamento de Química - ICEx

química
pós-graduação-ufmg

"Emprego de Ferramentas Quimiométricas Multidimensionais Na Análise de Mel e Café"

Daphne Chiara Antônio

Tese aprovada pela banca examinadora constituída pelos Professores:

Prof. Marcelo Martins de Sena - Orientador
UFMG

Prof. Bruno Gonçalves Botelho - Coorientador
UFMG

Profa. Mariana Ramos de Almeida
UFMG

Prof. Jez Willian Batista Braga
UnB

Profa. Adriana Nori de Macedo
UFMG

Prof. Jose Manuel Amigo Rubio
Universidad del País Vasco

Belo Horizonte, 21 de outubro de 2021.

Documento assinado eletronicamente por **Mariana Ramos de Almeida**, **Professora do Magistério Superior**, em 21/10/2021, às 18:13, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Jez Willian Batista Braga**, **Usuário Externo**, em 21/10/2021, às 18:14, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Jose Manuel Amigo Rubio**, **Usuário Externo**, em 21/10/2021, às 18:14, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Adriana Nori de Macedo**, **Professora do Magistério Superior**, em 21/10/2021, às 18:14, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Marcelo Martins de Sena**, **Professor do Magistério Superior**, em 21/10/2021, às 18:16, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Bruno Goncalves Botelho**, **Professor do Magistério Superior**, em 21/10/2021, às 18:17, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1025163** e o código CRC **E45AED54**.

Referência: Processo nº 23072.253818/2021-18 SEI nº 1025163

AGRADECIMENTOS

Agradeço à minha família, mãe, Dê e Dô, e aos meus amigos pelo amor, empatia e suporte.

Agradeço aos professores Marcelo M. Sena e Bruno G. Botelho pela orientação e pelo conhecimento transmitido ao longo deste trabalho.

À professora Débora Sampaio pela parceira no trabalho, pelas conversas e palavras de incentivo. À Dra. Verônica Belchior pela aplicação de dados sensoriais e contribuições no trabalho. À professora Maria José Nunes de Paiva, Bulé, e Mirna Maciel D`Auriol Souza pela infraestrutura nas análises cromatográficas.

Agradeço à Ilda Batista e à Isabel Carvalho, melhor parte da engenharia, por me apoiarem incondicionalmente durante esses sete anos, por me oferecerem palavras de incentivo, conforto e compreensão e, acima de tudo, me acolherem como um membro da família. To Prof. K. Osseo-Asare for inspiring me with your greatness and kindness. Thank you, Professor, to teach to have faith above all things.

Aos meus amigos/revisores Desireé Antônio, Alexandre Guimarães, Fernando Madureira, Fernando César Silva, Denise Versiane, Glaucimar Resende, Mariana Diniz, Mariana Teodoro, Gabriele Cardoso e Cleber Ferreira Junior.

Aos colegas do LAMS: Adriano, Ana, Bruna, Denise, Hebert, Jaime, Jimmy, Júlia, Keila, Marden, Marina, Mariana, Matheus, Millena, Odilon, Tarlene, Ricardo, Victória e Yéssica, pela convivência e valiosas discussões acadêmicas. Ao apoio discente, em especial à Selma Bazan, pelo suporte, paciência, conselhos, incentivo e amizade ao longo desses cinco anos.

Aos colegas do MZUFV, Renato, Clodoaldo, Camila, Kaíque, Duda, Polly, Adrielle, Guilherme e todos que ouviram os meus lamentos e reflexões durante esse ano.

Ao CNPq pelo financiamento da minha bolsa de doutorado. Às agências de fomento, CAPES e Fapemig, pelo apoio financeiro.
A todos que direta ou indiretamente contribuíram para a realização deste trabalho.

“To be a Negro in this country and to be relatively conscious is to be in a rage almost all the time.”

James Baldwin

RESUMO

O objetivo desta tese foi a aplicação de métodos de estatística multivariada para análise de mel e café, essencialmente análise de fatores paralelos (PARAFAC), método quimiométrico de segunda ordem. O PARAFAC foi utilizado nas aplicações como método de resolução de curvas para a identificação de compostos fluorescentes presentes nas amostras e quantificação destes explorando a vantagem de segunda-ordem, ou seja, a capacidade de quantificar um analito na presença de interferências não calibradas. Três aplicações foram desenvolvidas. A primeira aplicação desta tese apresentou um tutorial para a construção e utilização de um modelo PARAFAC exemplificado por um estudo de caso de análise sensorial de cápsulas de café. A aplicação visou demonstrar o emprego de PARAFAC para profissionais não familiarizados com este método quimiométrico. São abordados os métodos de pré-processamento de dados, a imposição de restrições matemáticas e os critérios que norteiam a escolha do número de fatores do modelo, como o teste de consistência trilinear (CORCONDIA) e a propriedade de unicidade pela análise de *split-half*. Na segunda aplicação, PARAFAC e espectrofluorimetria foram empregados no desenvolvimento de um método para a determinação direta de fenilalanina em mel. Modelos PARAFAC foram construídos utilizando matrizes de excitação-emissão molecular (EEM) obtidas por uma curva analítica de adição-padrão. A concentração de fenilalanina nas amostras variou de 5,7 mg kg^{-1} a 32,5 mg kg^{-1} e os resultados foram validados por um método cromatográfico. Na última aplicação, foi desenvolvido um método para a identificação de amostras de mel de Aroeira adulteradas com xarope de milho, melaço e mel polifloral. O mel de Aroeira é um promissor c*ommodity* oriundo da região norte de Minas Gerais que tem ganhado destaque por seu potencial antimicrobiano. Nesta aplicação, o PARAFAC foi utilizado como um método exploratório visando identificar os compostos fluorescentes presentes no mel de Aroeira puro. Modelos de classificação supervisionada (PLS-DA, U-PLSDA e N-PLSDA) foram construídos com as EEMs das amostras puras e adulteradas. Os modelos UPLS-DA apresentaram os melhores resultados de classificação com taxas de classificação incorreta de 4% e 8% para os conjuntos de treinamento e de teste, respectivamente.

Palavras chaves: Quimiometria. Métodos de ordem superior. PARAFAC. Espectrofluorimetria. Análise de alimentos.

ABSTRACT

The objective of this thesis was the application of multivariate statistical methods for honey and coffee, essentially parallel factor analysis (PARAFAC), a second-order chemometric method. PARAFAC was utilized in the applications as a curve resolution method for the identification of fluorescent compounds present in the samples and their quantification exploring the second-order advantage, i.e., the ability to quantify an analyte in the presence of uncalibrated interferences. Three applications were developed. The first application of this thesis presented a tutorial for the construction and use of a PARAFAC model exemplified by a case study of sensory analysis of coffee capsules. The application aimed to demonstrate the use of PARAFAC for professionals unacquainted with this chemometric method. The methods of data preprocessing, the imposition of mathematical constraints and the criteria for guiding the choice of the number of factors in the model are addressed, such as the core consistency diagnostic (CORCONDIA) and the property of uniqueness by the split-half analysis. In the second application, PARAFAC and spectrofluorimetry were used to develop a method for the direct determination of phenylalanine in honey. PARAFAC models were constructed using molecular excitation-emission matrices (EEM) obtained by a standard addition analytical curve. The concentration of phenylalanine in the samples varied from 5.7 mg kg^{-1} to 32.5 mg kg^{-1} and the results were validated by a chromatographic method. In the last application, a method was developed to identify samples of Aroeira honey adulterated with corn syrup, sugar cane molasses, and polyfloral honey. Aroeira honey is a promising commodity from the northern region of Minas Gerais State, Brazil, which has gained prominence for its antimicrobial potential. In this application, PARAFAC was used as an exploratory method to identify the fluorescent compounds present in pure Aroeira honey. Supervised classification models (PLS-DA, U-PLSDA and N-PLSDA) were built with the EEMs of the pure and adulterated samples. The UPLS-DA models presented the best classification results with misclassification rates of 4% and 8% for the training and test sets, respectively.

Key-words: Chemometrics. Higher order methods. PARAFAC. Spectrofluorimetry. Food analysis.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABREVIATIONS

ALS: Alternating least squares

GC-MS: Gas chromatography–mass spectrometry

CORCONDIA: Core consistency diagnostic

CVCE: cross validation classification error

DTD: Direct trilinear decomposition

EEM: Excitation Emission Matrix

FNR: False negative rate

FOM: Figures of Merit

FPR: False positive rate

HPLC: High-Performance Liquid Chromatography

HPLC-MS: High-performance liquid chromatography-mass spectrometry

HPLC-UV: High-Performance Liquid Chromatography with Ultraviolet detection

KNN: K-Nearest Neighbors

LDA: Linear discriminant analysis

LOD: Limit of detection

LOQ: Limit of quantification

LV: latent variables

MCR: Multivariate Curve Resolution

MCR-ALS: Multivariate Curve Resolution with alternating least squares

NIRS: Near Infrared Spectroscopy

NMC: Number of misclassifications

N-PLS: N-way Partial Least Squares

PARAFAC: Parallel Factor Analysis

PCA: Principal Component Analysis

PDO: Protected designation of origin

PLS-DA: Partial least squares discriminant analysis

QDA: Quantitative Descriptive Analysis

RSD: relative standard deviation

RSS: residuals of the sum of squares

SCA: Specialty Coffee Association

SD: Standard Deviation

SEL: selectivity

SEN: sensitivity

SOSAM: Second-order standard addition method

UHPLC-FLD: Ultra High-Performance Liquid Chromatography with Fluorescence Detection

UPLS-DA: Unfolded-partial least-squares discriminant analysis

VIP: Variable Importance in Projection

OUTLINE

1. INTRODUCTION

During the last decades, the amount and complexity of data in analytical chemistry and related areas has grown considerably. There are two main reasons for this trend. First, due to technological advances in relation to instrumentation and secondly, the society's demand that has grown daily requesting more fast and practical analytical methods. To improve the processing and comprehension of these data, chemometrics emerged in the late 1970s.

Chemometrics is the discipline of chemistry that combines mathematical and statistical methods to design and/or optimize experimental procedures and extract as much as possible relevant chemical information from a dataset (Kowalski, 1978). Driven by the popularization of microcomputers and microprocessors in chemical laboratories, chemometrics presented great potential and a new way of visualizing multivariate data (Brereton *et al.*, 2017; Kowalski, 1978).

The main study lines in chemometrics are multivariate calibration, modeling of the structure-activity relationship of molecules, pattern recognition (supervised and unsupervised classification), modeling and monitoring of multivariate processes (Brereton *et al.*, 2018, 2017; Wold & Sjöström, 1998).

In this thesis, methods of multivariate calibration and pattern recognition were applied to process food analysis data. Multivariate calibration aims to correlate a series of physical measurements with a given parameter, e.g., concentration, color intensity, biological activity or other property (Brereton, 2003). Pattern recognition involves primarily classification methods that seek the assignment of objects into groups or classes (Brereton, 2015).

Multivariate analyses tools are suitable for analyzing large sets of correlated data and extracting their chemical information, providing a new and improved perspective. With these tools, it is also possible to achieve second-order advantage, i.e., the ability to quantify an analyte in the presence of uncalibrated interferences (Olivieri, 2008). Thus, aiming to explore the potential of multivariate analysis tools, this thesis presented three applications on that subject: Chapters 3 to 5.

This thesis is divided into six chapters. Chapter 1 provides an introduction on the overall contents of the thesis. Chapter 2 presents the theory basis and literature review

of the chemometric methods and analytical techniques used in this work. Chapters 3 to 5 concern the three scientific papers prepared in connection with this PhD project, all of them have been or will be submitted for publication. Chapter 3 provides a tutorial of how to apply PARAFAC to sensorial data analysis through a case study of coffee capsules conducted with non-professional tasters. Chapter 4 presents the development of a direct spectrofluorimetric method to determine phenylalanine in honey based on multiway calibration using parallel factor analysis (PARAFAC). This application aimed to explore the second-order advantage. Chapter 5 presents a method to detect three types of adulteration in Aroeira honeys using spectrofluorimetry and supervised classification. Aroeira honey is a valuable food product from the north of Minas Gerais state, currently in process of certification as protected designation of origin, which has been target of adulteration due to its high market price. Finally, Chapter 6 brought the main conclusions of the work and the perspectives for future works.

2. LITERATURE REVIEW

2.1 Multivariate data analysis

2.1.1 Order of calibration multiway methods

Calibration methods used in analytical chemistry can be classified according to the dimensionality of the generated data, being extremely important the choice of the method in accordance with the data structure (Olivieri, 2008, 2012; Smilde *et al.*, 1999; Booksh & Kowalski, 1994).

Univariate data, composed of one scalar per sample, are treated by zero-order methods, nomenclature based on linear algebra (a scalar is a zero-order tensor). These methods must necessarily be totally selective to the analyte of interest. Scalars can be generated by instruments such as ion-selective electrodes, pH meters and single filter photometers (Booksh & Kowalski, 1994). First order data, organized in a vector (first-order tensor), are generated by instruments such as spectrometers and chromatographs, capable of recording multiple measurements for a single sample. For this type of data, specificity and signal resolution are not necessary and calibration is possible in the presence of interferences, as long as they are present in the calibration set used to build the model. Second-order data are organized in a data matrix (second-order tensor) and are mostly generated by hyphenated techniques (Booksh & Kowalski, 1994).

Methods of an order equal to or higher than three have become more frequent in the literature although not so abundant, perhaps due to the lack of understanding of the advantages associated with them or the complexity and unfriendly interface of their processing methods. In general, data of order equal to or greater than two are called data of higher-order or multiway data, focus of this work. Figure 2.1 shows different data arrangements according to the order of data generated by the instrument. Note that multiway data are a subdivision of multivariate data (Olivieri, 2012).

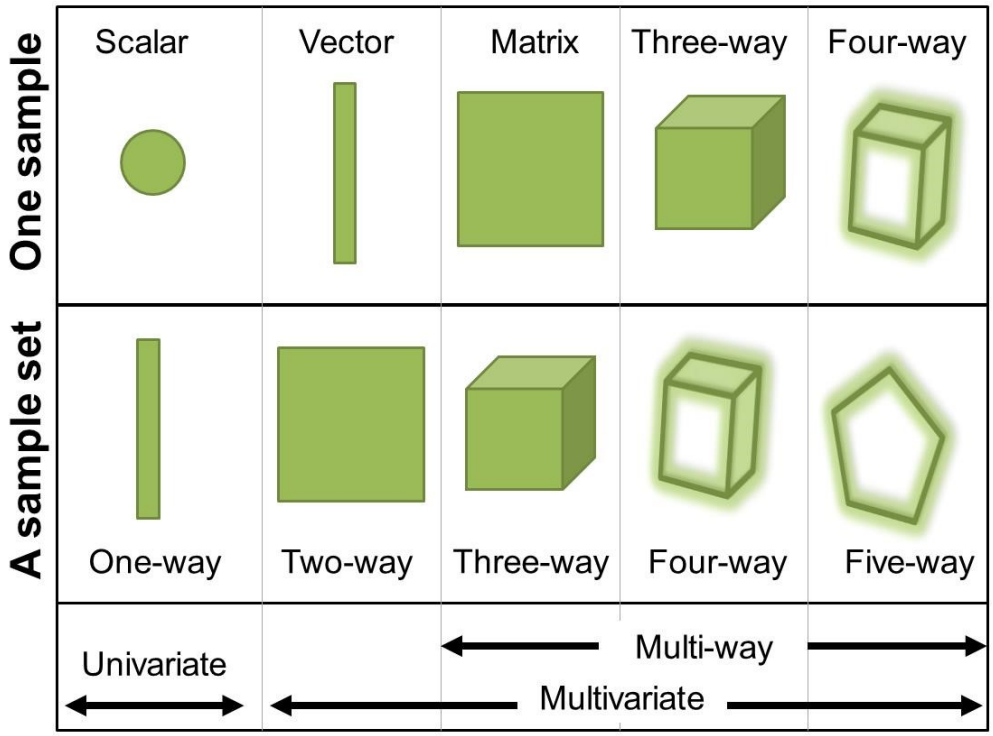

Figure 2.1. Representation of data for a single sample and for a sample dataset according to the order and number of ways (Adapted from Olivieri, 2012)

2.1.2 Organization of multiway data

For multidimensional or multiway data, the elements can be organized as a three-way array in which each element has three indices (Bro, 1998). In terms of notation for this thesis, dimension refers to the size of the data while modes, to the number of directions in the data. Therefore, Figure 2.2 presents a matrix of three dimensions *I x J x K* with three modes *I x J x K.*

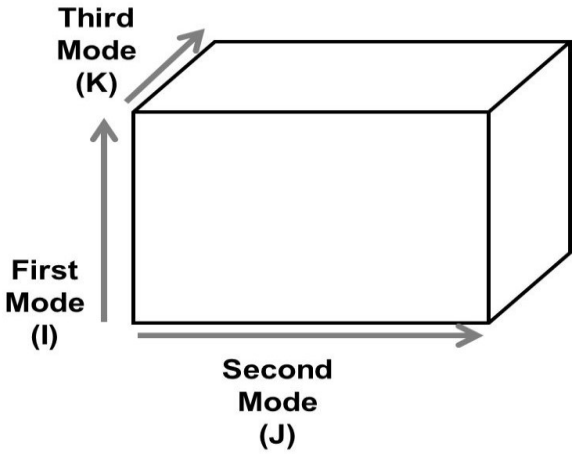

Figure 2.2. Graphical representation of a three-dimensional array, size I x J x K (Adapted from Bro, 1998)

In a typical multivariate analysis, the data are organized in a table or a matrix with two dimensions: rows and columns. Rows generally represent the samples, and columns the instrumental response, e.g., absorbance values at a given wavelength, the quantity measured by the instrument. Thus, the data can be ordered by two indices: the row-mode (*i*), e.g., samples number, and the column-mode (*j*), e.g., wavelength/variable, and the arrangement of the data depends on the instrumental technique, and the chemometric method selected for data processing.

In a three-way array, it is still possible to define submatrices by fixing one of the dimensions. In Figure 2.3, a submatrix (gray area), called a slab or slice of the array, was obtained by fixing the third mode index (*k*).

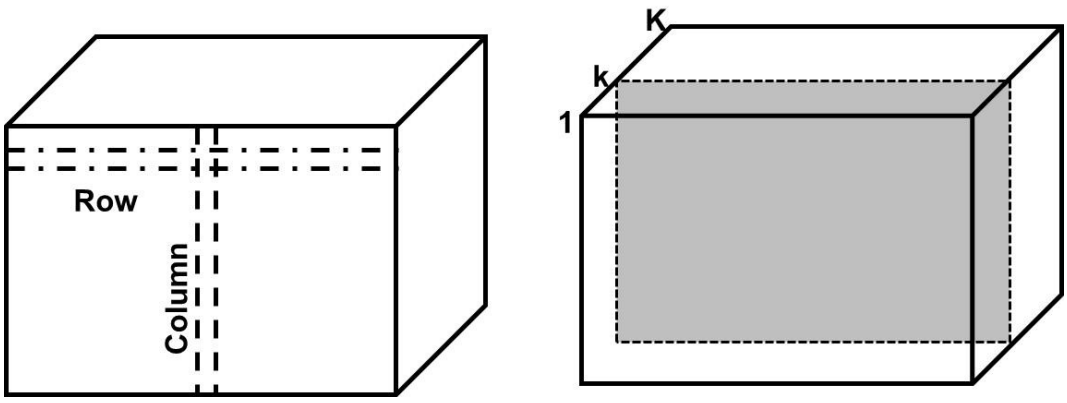

Figure 2.3. Representation of a row, column and slice of a three-way array (Adapted from Bro, 1998)

2.1.3 Multiway models

A mathematical model is a theoretical approximation of a set of real data, i.e., the matrix $\widehat{\boldsymbol{X}}$ contains the estimated values by a model of the real data contained in the matrix **X** (Bro, 1998).

The most used multiway model is PARAFAC (Carroll & Chang, 1970; Harshman, 1970). Besides PARAFAC, methods such as Multivariate Curve Resolution (MCR), Tucker3 and N-way Partial Least Squares (N-PLS) have been used to treat three-dimensional data (Christensen *et al.*, 2006). In this work, PARAFAC was used as the decomposition method and N-PLSDA and U-PLSDA as the multiway classification methods.

2.2 Parallel Factor Analysis (PARAFAC)

PARAFAC is a method to decompose three-dimensional or higher-order data into three trilinear (or more, multilinear) factors, usually based on the iterative alternating least squares (ALS) algorithm (Bro, 1997). ALS calculates the values of one of the three matrices originated from the three-way decomposition from the known loading values of the other two matrices. This operation is repeated until no significant difference is observed between the estimated values and the real data (de Juan *et al.*, 1997). The structural basis of the PARAFAC model is represented by Equation 2.1:

$$\boldsymbol{X}_{ijk} = \sum_{f=1}^{F} a_{if}\, b_{jf} c_{kf} + e_{ijk} \qquad (2.1)$$

In Equation 2.1, each element of $\mathbf{X}_{(ijk)}$ represents the instrumental response for the i^{th} sample at positions *j* and *k*. The three-way array is decomposed into a set of three matrices: $\mathbf{A}_{(if)}$, which contains the elements a_{if} corresponding to the loadings or scores of the i^{th} sample relative to the f^{th} factor (F is the total number of factors), and $\mathbf{B}_{(jf)}$ and $\mathbf{C}_{(kf)}$, containing the loadings of the two other modes (Bro, 1997). The terms loadings and scores will be interchangeably used for the elements of the first mode of the PARAFAC models. Data variance not captured by the model are represented in the elements of the residual three-way matrix $\underline{\mathbf{E}}_{(ijk)}$. Figure 2.4 shows a graphical representation of Equation 2.1 for F factors.

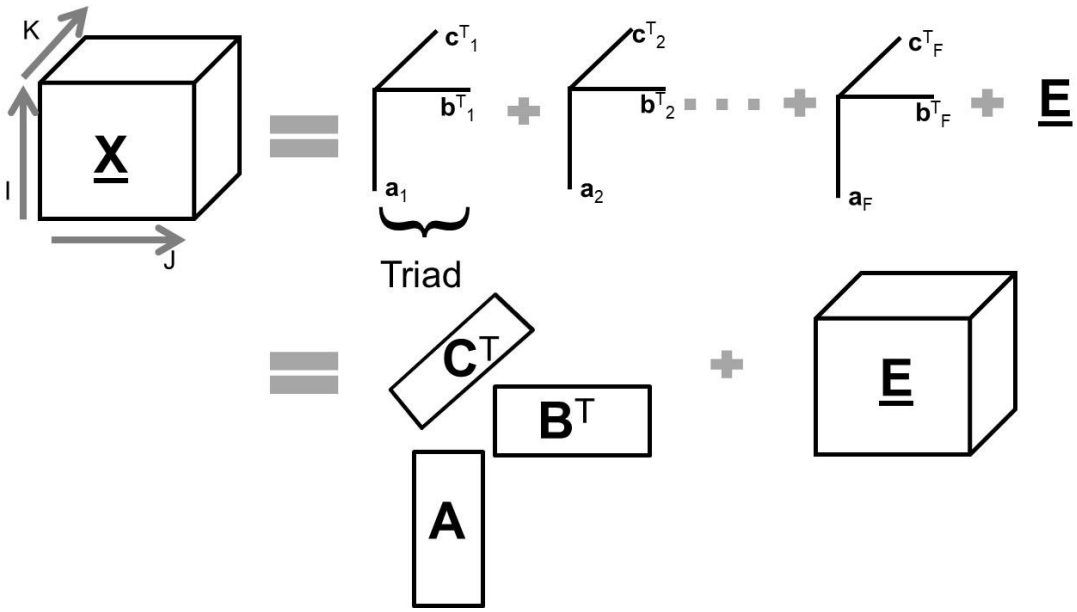

Figure 2.4. Graphical representation of PARAFAC model. Decomposition of a three-way array into three loading matrices (**A**, **B** and **C**), each containing F factors (Adapted from Sena *et al.*, 2005)

One of the most attractive features of PARAFAC model is uniqueness, i.e., the ability of providing a single mathematical solution for decomposing a given dataset. As a consequence, it is possible to determine the number of components present in the mixture in the absence of rotation ambiguity, and to recover realistic spectral profiles of the pure components. Uniqueness is a propriety specific of PARAFAC, which cannot be explored in other multivariate resolution methods as MCR, N-PLS and Tucker3. To achieve uniqueness, it is essential that the data are actually trilinear, the correct number of factors is chosen and the ratio signal-to-noise is appropriate. A trilinear component or triad is an outer product of three vectors, for PARAFAC models; each factor corresponds to one triad. Thus, it will also be possible to correlate, directly, the mathematical response to a chemical meaning, for example, the scores of the components to their relative concentrations (Bro, 1997).

The rank of a PARAFAC model provides the number of factors, F, necessary to describe all the systematic variation of the **X** data array. Ideally, the rank is equal to the number of components present in the system (Bro, 1997).

The selection of the number of factors is crucial to the model. Although there is no single criterion to guide this decision, some features of the model should be considered, e.g., the correspondence among the estimated spectral loadings and the chemical knowledge of the system, the split-half analysis strategy, the residual distribution and the relation between the trilinear consistency and the number of factors. The trilinear consistency is evaluated through the core consistency diagnostic (CORCONDIA) (Bro & Kiers, 2003).

CORCONDIA can vary from negative values up to 100%. Theoretically, values close to 100% suggest a suitable trilinear model. In general, values greater than 90% are considered to originate from trilinear data and intermediate values (~ 50%) indicate data of intermediate nature between trilinear and non-trilinear. On the other hand, CORCONDIA values negative or close to zero suggest unsuitable/non-trilinear models (Bro & Kiers, 2003). These strategies to select the number of factors will be discussed in detail in a practical tutorial of PARAFAC combined to sensorial analysis data, which will be presented in Chapter 3. In this thesis, PARAFAC was used to process data of two natures: molecular fluorescence spectroscopy and sensorial analysis data.

2.3 Molecular Fluorescence Spectroscopy

Molecular fluorescence spectroscopy, also known as spectrofluorimetry, is a relatively simple, non-destructive analytical technique, fast, of low operating cost and sensitive, currently applied in several research fields (Skoog *et al.*, 2009; Lakowicz, 2006).

Based on the phenomenon of electronic excitation-emission, spectrofluorimetry has been widely used due to its wide linear range response, high sensitivity (detection limits in the order of mg L^{-1}), selectivity and low cost in relation to other analytical techniques (Naresh, 2014; Lakowicz, 2006). Some external and internal factors can influence fluorescence intensity. Figure 2.5 presents some of them.

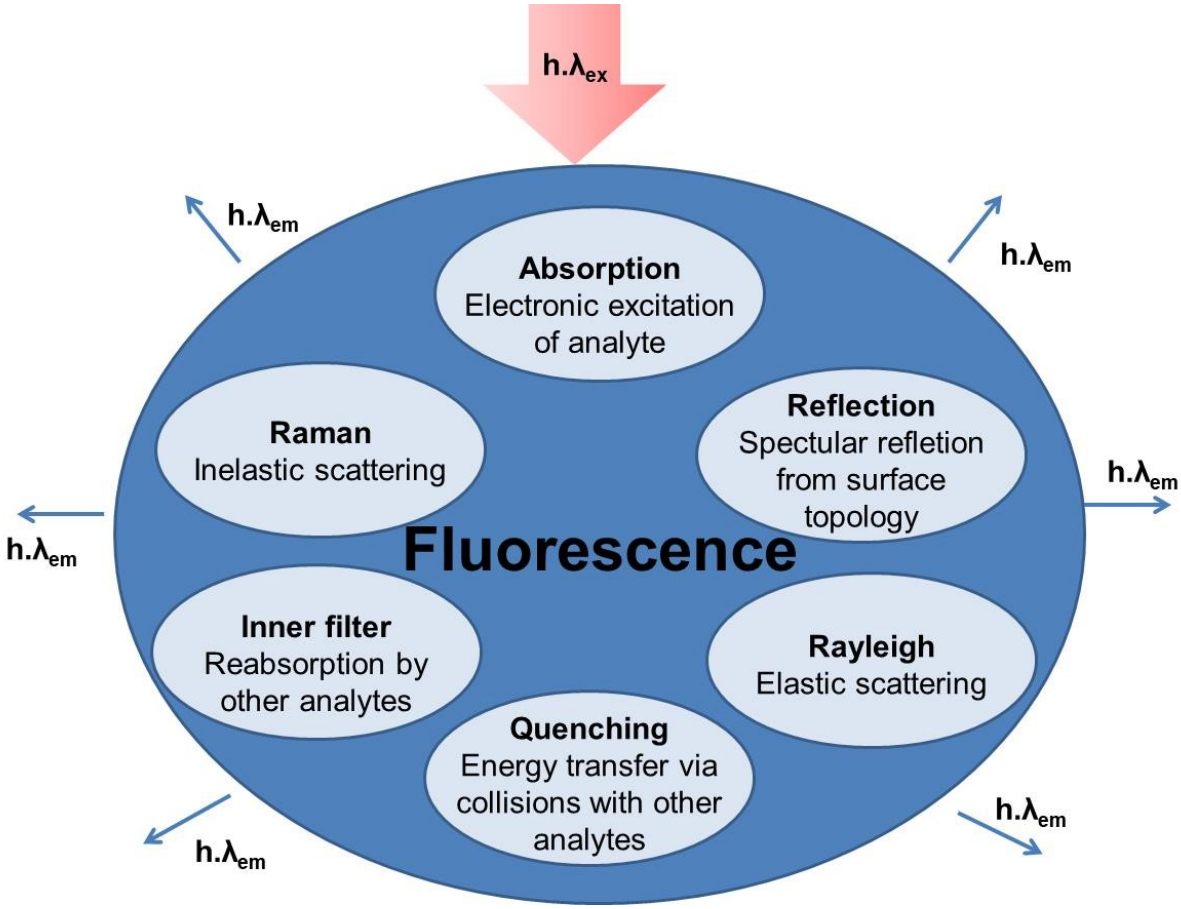

Figure 2.5. Summary of factors that affect the intensity of the fluorescent signal (Adapted from Christensen *et al.*, 2006)

Fluorescent suppression known as quenching is related to any process that reduces the intensity of fluorescence of a given species in a sample. Quenching can be either static or dynamic. Static quenching occurs when the formation of the excited state is inhibited due to the formation of a non-fluorescent complex between the fluorophore, i.e., the fluorescent compound, and a suppressor molecule still in the fundamental state. In the case of dynamic quenching, the deactivation of the fluorescent species occurs after it reaches the excited state. The increase in temperature favors molecular collisions and, in turn, increases the likelihood of intensity-suppressing effects (Christensen *et al.*, 2006; Lakowicz, 2006).

Because it is an optical phenomenon, the scattering and reflection of the incident light influences the intensity measured during fluorescence. The light scattering can be divided into elastic or inelastic, being differentiated by the loss or not of energy of the incident photon. Inelastic scattering or Raman scattering is associated with vibrational transitions of species present in solution. Raman scattering is a constant phenomenon of energy loss, but it can often be neglected, as it is weak compared to the fluorescence signal intensity (Christensen *et al.*, 2006).

Elastic scattering or Rayleigh scattering is characterized by the conservation of energy from incident radiation. At Rayleigh scattering the wavelength of the scattered light is equal to the wavelength of the incident light or integer multiples of this value, and occurs in a spectral region other than fluorescence. However, when employing large wavelength bands, Rayleigh scattering can significantly influence the emission spectrum with small displacement ranges known as Stokes' displacement (Skoog *et al.*, 2009; Christensen *et al.*, 2006).

Finally, the intensity of the fluorescent signal depends on the concentration of the fluorophore and the physicochemical properties of the solution. Under ideal conditions, the fluorescence intensity is directly proportional to the concentration of the analyte, according to Equation 2.2:

$$I_f = 2{,}3\varphi_f I_0 \varepsilon c l \qquad (2.2)$$

In Equation 2.2, I_f is the fluorescence intensity, φ_f is quantum yield, the ratio between the fluorescent molecules and the total molecules present in solution, ε is the molar absorptivity, c is the molar concentration of the fluorophore (mol L^{-1}), and l is the optical path length (cm). The fluorescent signals are independent, i.e., the total fluorescence signal is the sum of the signals from each of the fluorophores present in the system (Christensen *et al.*, 2006).

The physicochemical properties of the solution, e.g., pH, temperature and polarity, also influence the fluorescence signal. In more polar solutions, fluorophores that occupy the excited state will decay to a less energetic vibrational level, i.e., a region of higher wavelengths. Solvents of different polarities will cause shifts in the characteristic emission spectra (Skoog *et al.*, 2009; Christensen *et al.*, 2006). For highly concentrated samples, the linear relation expressed in Equation 2.2 is not obeyed since part of the incident radiation maybe by absorbed by others components of the solution and is decrease the observed fluorescence intensity. This is known as inner effect.

The pH of the solution can cause dissociation and/or molecular protonation, changing the rate of non-radioactive processes that compete with fluorescence, hence the quantum yield. The temperature interferes in the collision rate of the molecules, being directly proportional to the dynamic quenching (Christensen *et al.*, 2006; Lakowicz, 2006).

2.3.1 Excitation-Emission Fluorescence Matrices

Molecular fluorescence analysis can be performed at a fixed excitation wavelength resulting in an emission spectrum or by varying the excitation wavelengths and collecting a range of emission wavelengths obtaining an Excitation Emission Matrix (EEM). EEM represents a “fingerprint” of the sample and can be effectively used to characterize fluorophores present in solution (Lenhardt *et al.*, 2015).

EEM vary according to the chemical structure of the fluorophores and the solvent in which they are dissolved. The excitation and emission energies are specific to each fluorophore (Naresh, 2014; Lakowickz, 2006). Thus, all fluorophores will have independent and specific spectral excitation and emission profiles, which characterize their fluorescent properties.

PARAFAC has been successfully applied to process EEM fluorescence data for several types of matrices, e.g., water, food, environmental samples, and materials (Chunjian *et al.*, 2021; Vera *et al.*, 2017; Oloibiri *et al.*, 2017; Hu *et al.*, 2018; Wei *et al.*, 2020; Seopela *et al.*, 2021; Wang *et al.*, 2021). Table 2.1 presents an exhaustive literature review of the applications of PARAFAC combined to EEM fluorescence for food analysis, focus of this thesis, and this topic will be discussed in the following paragraphs.

Table 2.1. Applications of molecular fluorescence and PARAFAC in the last five years for food matrices

References	*Matrices*	*Main purpose*
Schueuermann *et al.*, 2018	Grape juice and wine	To compare profiles of samples from different vineyards
Carabajal *et al.*, 2019	Food-contact plastics	To study the kinetics of degradation and to quantify bisphenol A and nonylphenol
Montemurro *et al.*, 2019	Maize grains	To determine pirimiphos-methyl
Ríos-Reina *et al.*, 2019	Spanish wine vinegar of different PDOs[1]	To classify vinegars according to PDOs
Rubio *et al.*, 2019	Cherry	To quantify cochineal and erythrosine
Shang *et al.*, 2019	Tea	To determine the content of PAHs[2]
Wang *et al.*, 2019	Camellia oil	To identify and quantify camellia oil adulteration
Fang *et al.*, 2021	Beer	To characterize and classify Chinese pale lager beers
Palomino-Vasco *et al.*, 2021	Wine	To monitor winemaking process

[1] Protected designation of origin. [2]Polycyclic aromatic hydrocarbons

Schueuermann, Silcock & Bremer (2018) have applied front-face fluorescence spectroscopy and PARAFAC to compare profiles and obtain information on the potential quality of New Zealand Pinot Noir grape juice and wine samples. A four-factor

PARAFAC model was chosen and the signals were assigned to tryptophan, hydroxyl benzoic acid derivatives, caffeic acid, catechin and tyrosol related molecules. According to the authors, the strategy proposed proved to be a promising tool for the discrimination of grape juice and wine samples from different origins.

Aiming to determine pirimiphos-methyl (PMM), a type of insecticide, in maize grains, Montemurro and coworkers (2019) have combined molecular fluorescence data and second-order advantage through three-way calibration. Maize samples were minced and diluted in water; the extracts were analyzed in a range of excitation/emission wavelength of 240-270 and 320-450 nm. A four-component PARAFAC model was selected and the signal of PMM was assigned through equivalence with a standard solution analysis. PMM concentration was below the limit of detection and mean spiked sample recovery was 115%. The results were confirmed by UHPLC-FLD analysis.

Applying second-order advantage and EEM fluorescence; Rubio *et al.* (2019) have simultaneously quantified cochineal and erythrosine, food dyes, in cherries. For the analysis, the cherries were blended, diluted in a borate buffer solution at pH 8, filter and re-diluted in the buffer solution. A two-factor PARAFAC model was selected and the decomposed signals were assigned to the analytes. Cochineal and erythrosine concentrations were of 185.05 mg kg^{-1} and 10.76 mg kg^{-1}, respectively. Predicted results were validated by chromatography.

Several chemometrics tools and spectroscopic techniques have been combined by Ríos-Reina *et al.* (2019) in order to classify 65 Spanish wine vinegar samples of three protected designations of origin (PDO). Wine vinegar samples were analyzed directly without any sample preparation. A five-factor PARAFAC model was selected with the signals assigned to cumarins, tannins, phenols, flavonols from wine; 5-Hydroxmethylfurfural caramel; vitamin B2 in its principal forms; phenolic compounds, Maillard products, and oxidation products.

Aiming to determine three polycyclic aromatic hydrocarbons (PAHs), phenanthrene (PHE), naphthalene (NAP) and acenaphthene (ACE) in tea, Shang and co-workers (2019) have studied hexanic Da Hong Pao tea (Oolong tea, China) extract. Fluorescence measurements were performed at three temperatures levels: 15ºC, 20ºC and 25º C. Three-way PARAFAC models were built with dimensions 31 × 86 × 14, corresponding respectively to the number of excitation wavelengths x number of emission wavelengths x total of samples. The NAP results were considered poor by the

authors, thus a four-way PARAFAC model was built defining temperature as the new mode. A significant increase was observed in the NAP recovery (from 84.2 to 94.6%) and slight variation for the other analytes (ACE, from 93.5% to 96.0% and PHE, from 100.7% to 100.9%). The researchers highlighted the gain of sensitivity and selectivity with the insertion of the fourth dimension.

Three-way and four-way PARAFAC models have also been constructed by Fang *et al.* (2021). The authors applied PARAFAC as an exploratory tool to characterize 108 Chinese pale lager beer samples produced by four different manufacturers. Fluorescence measurements were acquired for beer samples undiluted and diluted to 3% in water in an excitation/emission wavelength range of 250–550 and 290–650 nm. Six components were related to PARAFAC decomposition: iso-α-acids, phenolic compounds, vitamins B2 and B6, tyrosine, tryptophan. An inversion on fluorophores intensity was observed, i.e., the three compounds with more intense fluorescence in undiluted samples (iso-α-acids, phenolic compounds and vitamin B2) presented the least intense signals in diluted samples. According to the authors, this inversion may be due to the inner filter effect.

Palomino-Vasco *et al.* (2021) have analyzed samples of wine from a winery located at the University of Extremadura, Spain. The samples were collected daily during winemaking as process control. The samples were diluted to 10% v v^{-1} in synthetic wine, in an aqueous solution at pH 3.5 with 3g L^{-1} of L-(+)-tartaric acid and 13% vv^{-1} of ethanol. A three-factor PARAFAC model was selected based on the CORCONDIA and model fit criteria. Among the decomposed signals, two components were recognized: tyrosine and tryptophan. Mean recoveries of 90.4% and 92.4% were observed for the spiked samples of tyrosine and tryptophan. The results were considered satisfactory by the authors, who pointed out fluorescence as a cheaper and faster method for wine making monitoring.

Applying PARAFAC and multiway classification methods, Wang *et al.* (2019) have proposed a method to identify and quantify adulteration of camellia oil, an important commodity to the economy of southern China. Pure samples of camellia oil and samples adulterated with soybean, peanut and sunflower oil in range of 10-80% vv^{-1} of adulterant were diluted to 1% vv^{-1} in hexane. A five-component PARAFAC model was chosen based on the split half analysis validation. The components were associated to tocopherols, and oxidation and degradation products of vegetal oils. Besides the use

for decomposition and interpretation of the fluorescence data, PARAFAC scores were also utilized to build LDA and PLS-DA models.

The review presented above shows that PARAFAC and molecular fluorescence spectroscopy offer a promising alternative to identify and quantify fluorophores in food and beverages matrices with a minimum or no sample preparation. Therefore, PARAFAC and molecular fluorescence spectroscopy were chosen for this study.

2.4 Sensory Analysis

Sensory analysis is an area that applies the five human senses, sight, smell, touch, taste, and hearing, to evaluate, analyze and interpret a sample, e.g., food, beverage, cosmetics, hygiene products (Lawless & Heymann, 2010). It has been a powerful tool for quality control, product development and to study consumer's acceptance and preference (Laczkowski *et al.*, 2021; Pereira *et al.*, 2020; Barahona *et al.*, 2020; Zhi *et al.*, 2016; Lim *et al.*, 2014; Valli *et al.*, 2014). Sensory analysis can be divided into three main types of tests according to their purpose: discrimination, descriptive and affective tests (Kilcast, 2010; Lawless & Heymann, 2010).

Discrimination tests or difference testing aim to answer whether exists any sensorial perceptible difference between two types of products. In these tests, a set of similar products is offered to the analysts and, based on the rate of correct answers, it is inferred whether there is a difference or not between the products (Lawless & Heymann, 2010). Discrimination tests usually achieve good results once the analyst's attention is focused on a specific attribute (Kilcast, 2010; Lawless & Heymann, 2010).

The most applied difference tests are the triangle, duo–trio, and paired comparison. In a triangle test, the set is composed by three samples: two equal samples and one different. Then, the analyst must choose the most different sample among the set. A similar set is offered in a duo-trio test; however, the analyst should select the sample that matches with a reference, i.e., should find the equal samples. Paired comparison is a straightforward test in which two distinct samples are offered to the analyst and one specific attribute is judged, e.g., which sample is sweeter or bitterer (Lawless & Heymann, 2010).

Descriptive analysis aims to develop a qualitative or quantitative product's profile through describing its sensory specifications, e.g., characteristic taste, smell, appearance and texture. Descriptive analysis has been widely used due to its diverse application for samples of different nature and its ability to provide more detailed information about the product (Kilcast, 2010; Lawless & Heymann, 2010).

Several approaches have been developed to perform a descriptive analysis such as flavor profile method, texture profile method, spectrum method and quantitative descriptive analysis (QDA). Although it is common to use these strategies combined, QDA has been the descriptive analysis method most applied (Gacula, 1997).

In a QDA test, a trained panel identifies and quantifies all sensory characteristics of a sample. Each sensory aspect is quantified using an unstructured line scale which matches a continuous scale. The quantitative results allow the use of several statistical tools to treat them which improves the discussion and interpretation of the data (Kilcast, 2010; Gacula, 1997).

Finally, hedonic or affective test is a type of sensory test that intends to quantify the degree of liking or disliking of a product, which can be extremely relevant to study consumer's preference (Lawless & Heymann, 2010).

A preference scale known as hedonic scale was associated with the test to estimate the degree of pleasant or unpleasant an attribute provokes in an analyst. Originally, the hedonic scale is a 9-point scale with equal intervals and the extremes representing extremely dislike and extremely like (Lawless & Heymann, 2010). Nowadays, variations of the scale length with five and seven points are commonly found (Zhi *et al.*, 2016; Lim *et al.*, 2014; Bu *et al.*, 2013; Muoki *et al.*, 2012). The hedonic scale is a powerful tool to approximate the subjective results to analytical results once they are strongly influenced by the stimuli provoked in the body and awakened memories (Lawless & Heymann, 2010; Gacula, 1997).

Regardless of the type of sensory analysis, analysts are needed, who compose a panel. They can be trained or untrained panelists according to the type and propose of the test. To develop the analytical skills, a reference product is used to illustrate the target quality. The training program duration varies with the analyst performance and the level of sensory description desired. The training minimizes possible affective

influences in the analysis. In general, a panelist requires a critical sense, objectivity and knowledge regarding the trial (Kilcast, 2010; Lawless & Heymann, 2010).

To improve the interpretation of the sensory data; statistical analysis tools have been successfully applied. Principal Component Analysis (PCA) is by far the chemometric tool most used to analyze sensory data (Eiermann *et al.*, 2020, Cortés-Diéguez *et al.*, 2020; Bressanello *et al.*, 2017; Vasilev *et al.*, 2016; Morais *et al.*, 2015; Jung *et al.*, 2014; Naes *et al.*, 2010; Bro *et al.*, 2008; Helgesen *et al.*, 1997).

Zhi *et al.* (2016), for instance, have used penalty analysis and partial least squares discriminant analysis (PLS-DA) to analyze data of Chinese consumer preferences on flavored liquid milk. Correspondence analysis and structural equation model, based on the PLS (SEM-PLS) have been used by Barahona *et al.* (2020) to analyze sensorial analysis data of Colombian coffees from different cities. PLS has also been used by Ribeiro *et al.* (2011) to correlate sensory data and NIR spectra of Brazilian Arabica coffee samples aiming to predict the scores of their sensory attributes.

Cordella *et al.* (2011) have used Tucker3, a three-way method, to interpret sensory analysis data of wheat noodles. PARAFAC has been used by Gomes *et al.* (2020); Pereira *et al.* (2020); Curi *et al.* (2017); Morais *et al.* (2015); Nunes *et al.* (2011), aiming to provide three-way preference maps to study consumer's acceptance/preference.

Despite all these applications, some of statistical methods were not explored to its full potential perhaps due to the lack of expertise in the field, since they were primary applied by non-specialists, or due to the complexity of the tools. To tackle this question, a tutorial of how to build and apply a PARAFAC model will be presented in Chapter 3.

2.5 Pattern recognition

Pattern recognition is a chemometrics area that aims to determine patterns or relationships among samples. It involves primarily classification methods, but it can be used in exploratory analysis to detect outliers and determine important variables or extract features of the data (Brereton, 2015).

Regarding the classification methods on pattern recognition, they can be supervised or unsupervised, i.e., built with or without the aid of any additional information about the origin of the samples (Brereton, 2015).

Unsupervised methods attempt to group the samples without any predefined information. They are used, in general, to evaluate the data structure, detect possible outliers and recognize natural patterns in the samples (Brereton, 2015). Principal component analysis (PCA) is the most common unsupervised pattern recognition method; it allows a graphical visualization of the data and aims to reduce its dimensionality through the establishment of new orthogonal variables, named as principal components (PCs). The method preserves the maximum of variance of the original data and the PCs are organized in descending variance order (Bro & Smilde, 2014).

Supervised methods attempt to divide the samples into groups based on their characteristics using a training set, i.e., a set of samples with predefined classes (Brereton, 2015). There are several approaches for the supervised methods, e.g., Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Partial Least Squares-Discriminant Analysis (PLS-DA), Back-Propagation and Counter-propagation Artificial Neural Networks (BP- & CP-ANN), Soft Independent Modeling of Class Analogy (SIMCA) (Marini, 2008). PLS-DA is the most used supervised classification method. In this work, PLS-DA and its multiway variants were used to discriminate authentic from adulterated honey samples. This application will be presented in Chapter 5.

2.5.1 Partial Least Squares Discriminant Analysis (PLS-DA)

PLS-DA is a discriminant bilinear model based on a PLS regression (Brereton & Lloyd, 2014). A linear relation is constructed between a **X** data matrix and a **y** vector containing the dummy variables for the assignations of each class. The **y** vector contains the information of the pre-established classes (Brereton & Lloyd, 2014). The model decomposes the matrix **X** and the vector **y** into loadings (**P** and **q**) and scores matrices (**T**), aiming simultaneously to maximize variance in **X** and correlate it with **y**. The variance not captured by the model is contained in the residuals matrix ($\mathbf{E_x}$) and the residuals vector ($\mathbf{e_y}$). Equations 2.3 and 2.4 show the decomposition of the original data contained in **X** and **y**.

$$\boldsymbol{X} = \boldsymbol{T} * \boldsymbol{P}^T + \boldsymbol{E}_x = \sum_{i=1}^{A} \boldsymbol{t}_i \, \boldsymbol{p}_i^T + \boldsymbol{E}_x \quad (2.3)$$

$$\boldsymbol{y} = \boldsymbol{T} * \boldsymbol{q}^T + \boldsymbol{e}_y = \sum_{i=1}^{A} \boldsymbol{t}_i \, \boldsymbol{q}_i^T + \boldsymbol{e}_y \quad (2.4)$$

Through successive calculations, a weight matrix **W** is obtained as presented in Equation 2.5.

$$\boldsymbol{W} = \boldsymbol{X}' * \boldsymbol{Y} \quad (2.5)$$

Matrices **X** and **Y** are correlated by applying the linear regression principle as presented in Equation 2.6, in which **b** is the regression coefficient vector of dimensions *J x 1.*

$$\hat{\boldsymbol{y}} = \boldsymbol{X} * \boldsymbol{b} \quad (2.6)$$

The regression coefficient relates the weights and the loadings as presented in Equation 2.7.

$$\boldsymbol{b} = \boldsymbol{W} * (\boldsymbol{P}^T * \boldsymbol{W})^{-1} * \boldsymbol{q}^T \quad (2.7)$$

Finally, the predicted values for the new samples ($\hat{\boldsymbol{y}}$) are obtained relating the **T** scores matrix and the **q** loadings through Equation 2.8.

$$\hat{\boldsymbol{y}} = \boldsymbol{T} * \boldsymbol{q}^T = \boldsymbol{X} * \boldsymbol{W} * \left(\boldsymbol{P}^{\boldsymbol{T}} * \boldsymbol{W}\right)^{-1} * \boldsymbol{q}^{\boldsymbol{T}} \quad (2.8)$$

Samples are assigned to each class based on their $\hat{Y}$ values. Since these predicted values are continuous, a threshold must be estimated to allow classification. Typically, thresholds are estimated based on the Bayes theorem (Gustafson *et al.*, 2009), assuming a normal distribution of $\hat{Y}$ and determining the point where the estimated probability distributions cross each other, i.e., the $\hat{Y}$ value that minimizes simultaneously the numbers of false positives and false negatives.

Corroborating for the correct assignment of the new samples, it is essential the selection of the correct number of latent variables (LVs), i.e., the number of components necessary to describe the maximum of variance with the minimum prediction error as expressed through the cross-validation classification error (CVCE). The number of LVs is usually selected through cross-validation in which a part of the samples is used for validation and the prediction error is calculated. This process is repeated with different samples of the dataset until all the samples had been predicted. The number of LVs is selected for the minimum classification error.

2.5.2 Multilinear partial least squares discriminant analysis (NPLS-DA)

Multilinear or N-way PLS (N-PLS) was proposed in 1996 as an extension of the multivariate calibration method PLS for higher-order data. For a three-away array $\underline{\mathbf{X}}_{(IxJxK)}$, composed of I samples, J excitation wavelengths and K emission wavelengths, as in the present study, a trilinear decomposition is performed providing one score vector $\mathbf{t}_{(Ix1)}$ and two weight vectors $\mathbf{w_j}_{(Jx1)}$ and $\mathbf{w_k}_{(Kx1)}$, according to the structural basis represented in Equation 2.9, in which $\mathbf{e_{ijk}}$ contains model residuals (Bro, 1996).

$$\boldsymbol{x}_{ijk} = \boldsymbol{t}_i \boldsymbol{w}_j^J \boldsymbol{w}_k^K + \boldsymbol{e}_{ijk} \qquad (2.9)$$

The decomposition shown in Equation 2.9 is represented graphically in Figure 2.6.

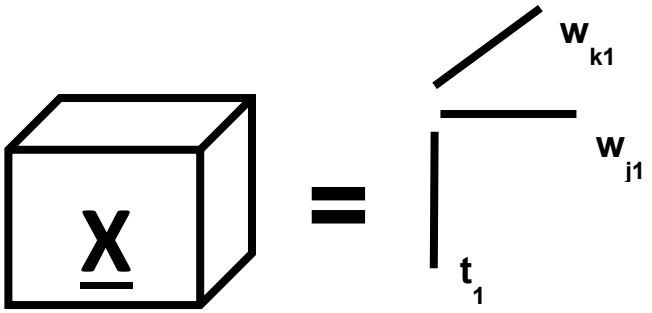

Figure 2.6. Graphical representation of a decomposed one-component trilinear model of **X** (Adapted from Bro, 1996)

NPLS-DA is the natural extension of N-PLS for discriminant analysis and is particularly appropriate to processes three-way spectrofluorimetric data in food analysis (Jimenez-Carvelo *et al.*, 2019).

2.5.3 Unfolded-partial least-squares discriminant analysis (UPLS-DA)

U-PLS is a variant of PLS proposed to process three-way data but based on a bilinear decomposition. In this sense, U-PLS is not truly a multiway method since it employs a two-way strategy to deal with three-way data. For a three-way array $\underline{\mathbf{X}}_{(I \times J \times K)}$, the third mode (emission spectra) is unfolded along the direction of the second mode (excitation spectra) originating a **X** matrix of dimensions IxJK (Wold *et al.*, 1987). This process of reshaping a three-way array into a two-way matrix is called unfolding and is represented in Figure 2.7. In UPLS-DA, after the unfolding step, the matrix is modeled similarly as in PLS-DA.

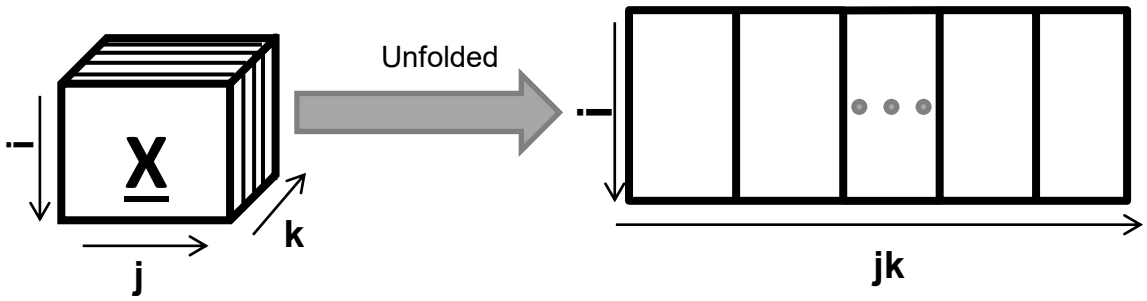

Figure 2.7. Schematic representation of unfolded strategy (Adapted from Bian *et al.*, 2016)

PLS-DA, NPLS-DA and UPLS-DA models were constructed to detect adulteration of a valuable Brazilian honey presented in Chapter 4.

3. PARAFAC APPLIED TO SENSORY ANALYSIS: A CASE OF STUDY OF COFFEE CAPSULES CUPPING

3.1 Introduction

Coffee is one of the most consumed beverages all over the world. In 2019, more than 10 billion tons were consumed in the world; Brazil is responsible for 13% of this amount (ICO, 2020). Capsules of coffee or monodoses have gained notoriety since they provide an expressive range of flavors in a short time of preparation and in general with accessible cost for consumers. In the last five years, the market of coffee capsules has grown significantly from 0.6% in 2014 to 1.3% in 2019 of the total volume consumed in the country (ABIC, 2020).

Coffee quality is assessed mainly through sensory analysis, in which the five human senses are used to evaluate the characteristics or attributes of a given product recognized by the taster. Regarding the cupping protocol for special coffees, the Specialty Coffee Association (SCA) recommends the evaluation of the following attributes: Fragrance/Aroma, Flavor, Aftertaste, Acidity, Body, Balance, Uniformity, Clean Cup, Sweetness, Defects, and Overall (SCA, 2003).

Sensory analysis results can be expressed in a preference score scale, known as hedonic scale, or in a quantitative descriptive analysis (QDA), in which the attributes are judged in individual line scales that intersect to form a polygon with the attributes occupying the position vertices. QDA aims to improve the interpretation of results through studying the interaction between attributes (Kreuml *et al.,* 2013; Ribeiro *et al.*, 2011, Gacula, 1997). To perform the analysis, the SCA recommends a 5-point scale ranging from 6 to 10 with intervals of 0.25 (SCA, 2003).

Chemometrics tools are especially useful to explore as much as possible the information and study the relation between variables. When a hedonic scale is chosen, the discreet data can be organized into a three-way array with samples, attributes and tasters as the three dimensions or modes. PARAFAC has been the multiway processing method most used to treat sensorial analyses data (Morais *et al.,* 2015; Gere *et al.,* 2014; Schmidtke *et al.,* 2013; Nunes *et al.,* 2011; Dahl *et al.*, 2008; Cocchi *et al.,* 2006). Based on the decomposition into trilinear factors, PARAFAC provides a

general visualization of the data and allows studying the relationships among samples, attributes and tasters simultaneously and link these variables to each other.

However, for who is non-familiar with this multiway method, PARAFAC can be quite complex since it requires some background knowledge and provides different results that can be difficult to interpret at first glance. Thus, this application aims to simplify and clarify the use of PARAFAC for sensory analysis through a tutorial and exploring a case study. In this application, espresso coffee capsules were assessed by six non-professional tasters and the results were expressed in a 5-point hedonic scale with respect to five attributes: aroma, aftertaste, flavor, acidity and body.

3.2. Materials and methods

3.2.1 Data

The experimental procedure was conducted in 2017 as one application described in the paper published by Belchior *et al.* (2019). In summary, twenty-two (22) espresso coffee capsules were assessed in triplicates with respect to different attributes by six non-professional tasters. These taster panelists were previously well trained by a certified professional Q-grader. Samples were coded with the letter *S*, which stands for sample and a sequential number from 1 to 22. Table 3.1 presents geographic origin, sensory characteristics, and intensity attribute provided by their manufacturers.

As established by the SCA, five attributes (aroma, flavor, aftertaste, acidity and body, i.e., texture of the beverage) were evaluated in a 5-point hedonic scale with 1 meaning the least pleasant and 5 meaning the most pleasant. The average of triplicates of each attribute was considered as the final result. The data was organized into a three-way array as *samples (lines) x attributes (columns) x tasters* (cubes), as shown in Figure 3.1.

Table 3.1. Origin, sensory characteristics, and intensity of samples according to manufacturers

Samples	Origin	Sensory characteristics	Intensity
S1	South and Central America	Cocoa, intense	9
S2	South and Central America	Cocoa, intense	9
S3	Ethiopia	Floral, intense	3
S4	South America	Cereals, balanced	5
S5	East Africa, Central America and South America	Fruity, citric	3
S6	IU	IU	2
S7	South America	Species *Arabica* and *Robusta* with cereals, cocoa roasted, intense	7
S8	IU	IU	3
S9	Latin America and Asia	Intense, bitter cocoa notes	11
S10	Brazil	Sweet, cereals, balanced	4
S11	Unspecified blend and coffees from India	Sweet cereal, roasted intense	8
S12	South of India	Species Arabica and Robusta, spicy, intense	10
S13	South America, Brazil and Guatemala	Species *Arabica* and Robusta, woody, intense, spicy	12
S14	Brazil and Colombia	Malted cereal, balanced	4
S15	South and Central America	Caramel aromas, balanced	6
S16	South America and East Africa	Species Arabica and Robusta, soft, fruity, dark roast, intense	10
S17	South and Central America	Species Arabica and Robusta, woody, intense	8
S18	Colombia	Fruity	6
S19	South America and East Africa	Floral, fruity	4
S20	South and Ethiopia	Floral, fruity	4
S21	South America	Cereal, fruit, balanced	4
S22	Brazil and Colombia	Species Arabica, cereal, fruit, balanced	4

IU: Information unavailable

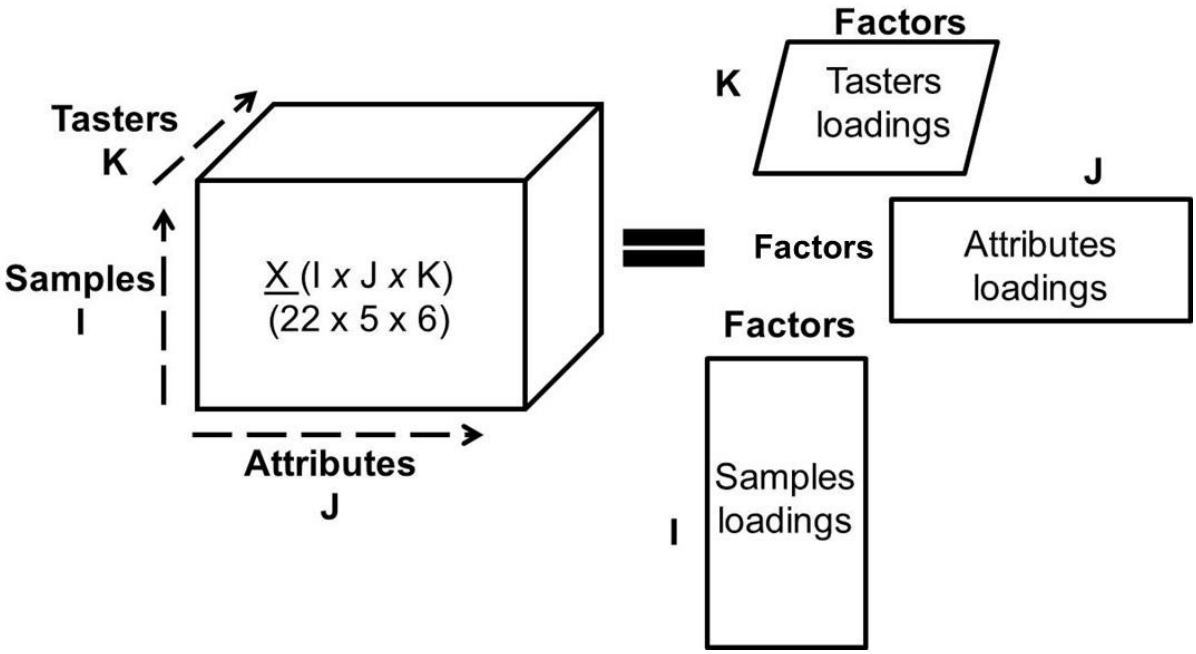

Figure 3.1. Sensory data array for PARAFAC model

3.2.2 PARAFAC theory

PARAFAC is a decomposition method in which tridimensional or higher-order data are decomposed into independent trilinear factors usually through the iterative alternating least squares (ALS) algorithm, according to Equation 3.1 (Bro, 1997):

$$\boldsymbol{X}_{ijk} = \sum_{f=1}^{F} a_{if}\, b_{jf} c_{kf} + e_{ijk} \quad (3.1)$$

Each element of the matrix $\mathbf{X}_{(ijk)}$ represents the score of the sample *i-th* at attribute *j* and taster *k*. The array is decomposed into a set of three matrices, **A**, **B** and **C**, with elements $a_{(if)}$, which contain the *i-th* sample scores of the *f-th* factors, $b_{(jf)}$ and $c_{(kf)}$, which contain loadings of the attributes and the tasters, respectively (Bro, 1997). The tree-way matrix **E** contains the data variance not captured by the model, known as residuals.

In addition to providing an overview of the relationships among sensory data, PARAFAC has a great advantage when compared to other methods, the property of producing unique solutions free of rotation ambiguity, known as uniqueness (Bro,

1997). To achieve uniqueness, however, the proper number of factors to construct the model must be chosen.

3.2.2.1 PARAFAC models

PARAFAC calculations were carried out using the PLS Toolbox, 6.71 (Eigenvector Technologies, Manson, USA) for Matlab 7.10.0 (R2010a) (MathWorks, Natick, USA). No preprocessing methods were used and nonnegativity constraint was tested in the second mode. Other parameters such as number of iterations, time, relative and absolute change of iterations for stopping criteria were used as in the default. The number of factors varied from 1 to 5 taking into account the number of attributes evaluated.

3.3. Results and discussion

3.3.1 Data Preprocessing

Data preprocessing are mathematical operations used to eliminate random and systematic sources of variation not related to the chemical composition of the system. The choice of the preprocessing method can vary according to the data nature and the aim of the analysis. Some preprocessing operations are frequently used such as scaling, i.e., to provide the same contribution weight for each variable or sample to the model, centering, i.e., to subtract a constant value from each value of a given variable, and scattering removing, i.e., to remove the non-trilinear portion of the data, which can undermine the model (Olivieri *et al.,* 2015).

In the present study, no preprocessing method was applied. The data were scored in the same units and range for each variable/attribute, and therefore scaling and centering are not needed (Bro, 2018). Preprocessing used for continuous spectroscopic measurements, such as smoothing and scattering removing, are also not suitable for this discrete data.

3.3.2 Constrains

After organizing and preprocessing the data, the PARAFAC model can be calculate. Improvement of PARAFAC solution can be achieved through applying constraints.

Constraints are imposed to the model aiming to constrain its mathematical solution based on the knowledge of the system. They provide as much known information as possible to the model and direct its mathematical solution closer to the chemical meaning of the system (de Juan *et al.*, 2014). Among the most used constraints we can mention nonnegativity, orthogonality, unimodality and known pure spectra and/or concentration profiles (Olivieri *et al.,* 2015).

Nonnegativity constraint can be used when the data cannot assume negative values, e.g., for chemical components concentration and spectra profiles. The negative values are replaced by null or positive values through softer algorithms. Unimodality can be applied when only one peak can be assumed regionally, for instance, in a chromatogram time interval only one chromatographic peak must exist in the elution profile (Olivieri *et al.,* 2015).

A constrained model can reduce the ambiguity of its solution and sometimes increases its interpretability. On the other hand, a constrained model will always present lower explained variance than an unconstrained model (Bro, 1997). Therefore, the use of constraints must be evaluated carefully aiming to obtain more “realistic” solutions at the cost of lower explained variance.

For this data set, only the nonnegativity constraint in the second mode (attributes) was adopted. A more constrained model was not tested due to the noisy nature of the data. Figure 3.2 presents the loadings of constrained and unconstrained models with two factors, as an example. These loadings were presented as continuous solid lines to provide better visualization despite the discrete nature of the data. Three-factor constrained and unconstrained models were constructed and the same observations are valid.

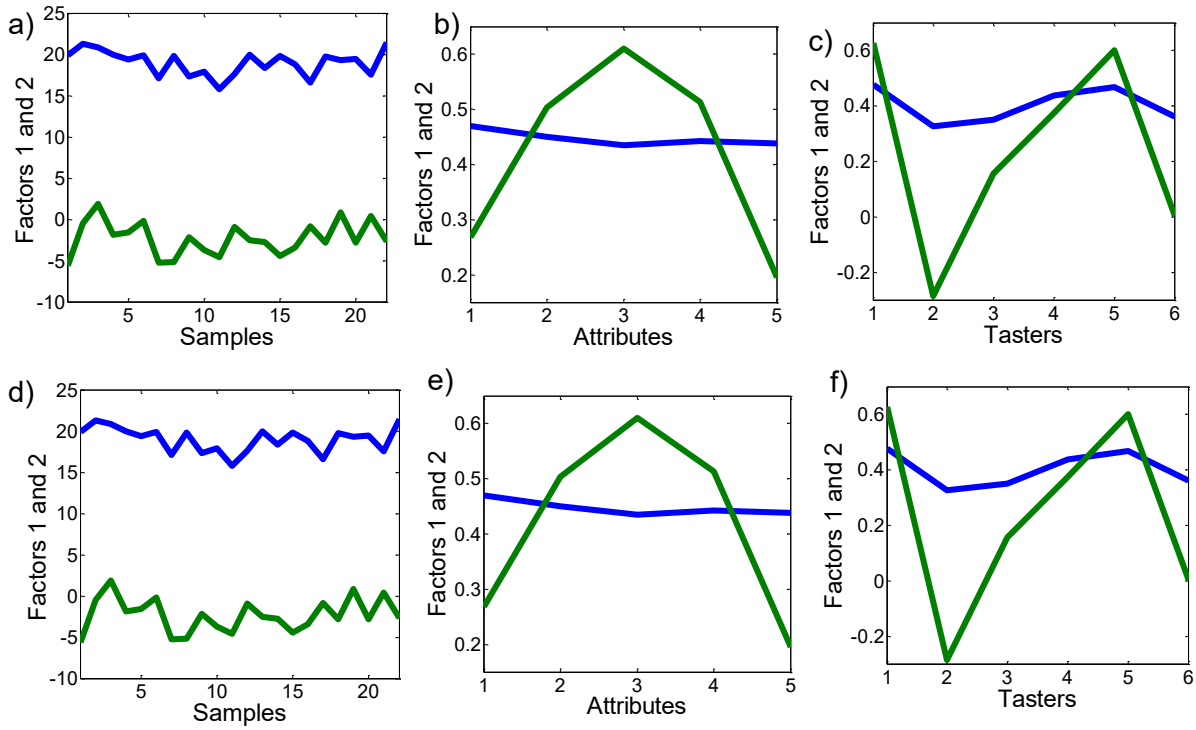

Figure 3.2. Solutions for two-factor PARAFAC models: **(a)**, **(b)** and **(c)** first, second and third mode for the constrained model; **(d)**, **(e)** and **(f)** first, second and third mode for the unconstrained model

We observed that PARAFAC model under nonnegativity constraint in the second mode provided statistically equal predictions (paired Student´s t-test, 95% of significance level) than the unconstrained model. Since no difference was observed and aiming to constraint as least as possible the model, the unconstrained model was chosen. A similar behavior was observed by comparing unconstrained and constrained models built with three factors (plots not shown).

3.3.3 Number of factors

Following the constraint step, it is necessary to choose the number of factors required to describe the data without under fitting. There is no general rule to determine the number of factors of the model, although a few strategies can be used as guidance, e.g., the relation between the number of factors and CORCONDIA, split-half analysis, visual inspection of residuals and knowledge about the system (Bro, 1997). The final decision is responsibility of the data analyst, who should be based on several

diagnostics as well as background knowledge of the data and the purpose of the study (Bro & Kiers, 2003; Smilde et al., 2004).

3.3.3.1 Visual inspection *versus* known real profiles

A careful visual inspection of the loadings of each mode was performed to compare them with the previous knowledge of the system, aiming to assess and to interpret the final model. For spectral data, for instance, the pure spectra of some components may be previously known. The same is valid for chromatographic profiles of the components. Any type of analytical signal may be previously known from the literature or experimentally obtained if standards are available (Smilde *et al.*, 2004).

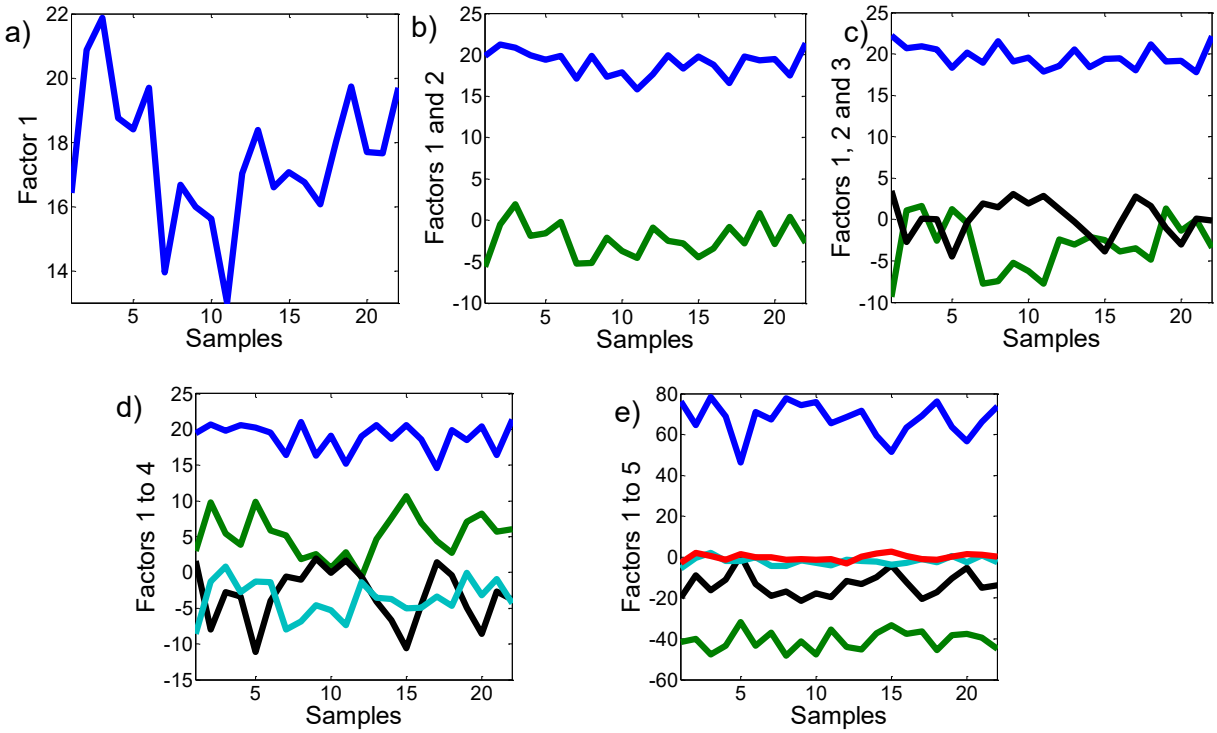

Figure 3.3. Loadings of the first mode (samples) varying the number of factors from 1 **(a)** to 5 **(e)**

In the case of sensory data, it is not possible to perform a comparison with the literature. However, by performing a visual inspection, it is clear that loading profiles of the fourth and fifth factors for the five-factor model are almost exactly the same (light

blue and red lines, Figure 3.3e), which suggests that no more than four factors should be used to describe these data.

3.3.3.2 Number of factors *vs* CORCONDIA

CORCONDIA is a mathematical tool used to evaluate the trilinear consistency of a PARAFAC solution. CORCONDIA varies from negative values to 100%, whereas values around or below 50% usually indicate lack of trilinearity and values above 90% suggest a good model fit and trilinearity consistency (Bro & Kiers, 2003).

It is important to emphasize that CORCONDIA is a mathematical tool that does not take into account the nature of the data. Therefore, noisy data can present lower values of CORCONDIA and still have trilinear consistency.

PARAFAC models were built using from 1 to 4 factors in order to choose the appropriate number of factors. Table 3.2 presents the explained variance and the CORCONDIA values according to the number of factors.

Table 3.2. Explained variance in **X** and CORCONDIA (%) for PARAFAC models built using from 1 to 4 factors

Number of factors	Explained variance (%)	Core consistence (%)
1	96.4	100
2	97.4	100
3	98.1	27
4	98.5	0

The models provided similar levels of explained variance in the **X** block, but the CORCONDIA value variation suggests that no more than three factors should be used. Taking into account the recovered profiles and the CORCONDIA values, a two- or a three-factor model should be the suitable choice.

3.3.3.3 Residuals

The residuals are other important parameters of chemometric models. Residuals are the difference between the raw data and the data used in the modeling as recovered by

the chosen number of factors in the **X** block. Thus, small residuals indicate a good fit and a good performance of the model (Smilde *et al.*, 2004; Bro & Kiers, 2003). The amount of residual variance depends on the random noise of the data. Residuals can be calculated for each mode as well as for the whole three-way matrix. Figure 3.4 shows the residual sum of squares (RSS) of each mode for a two- and a three-factor model.

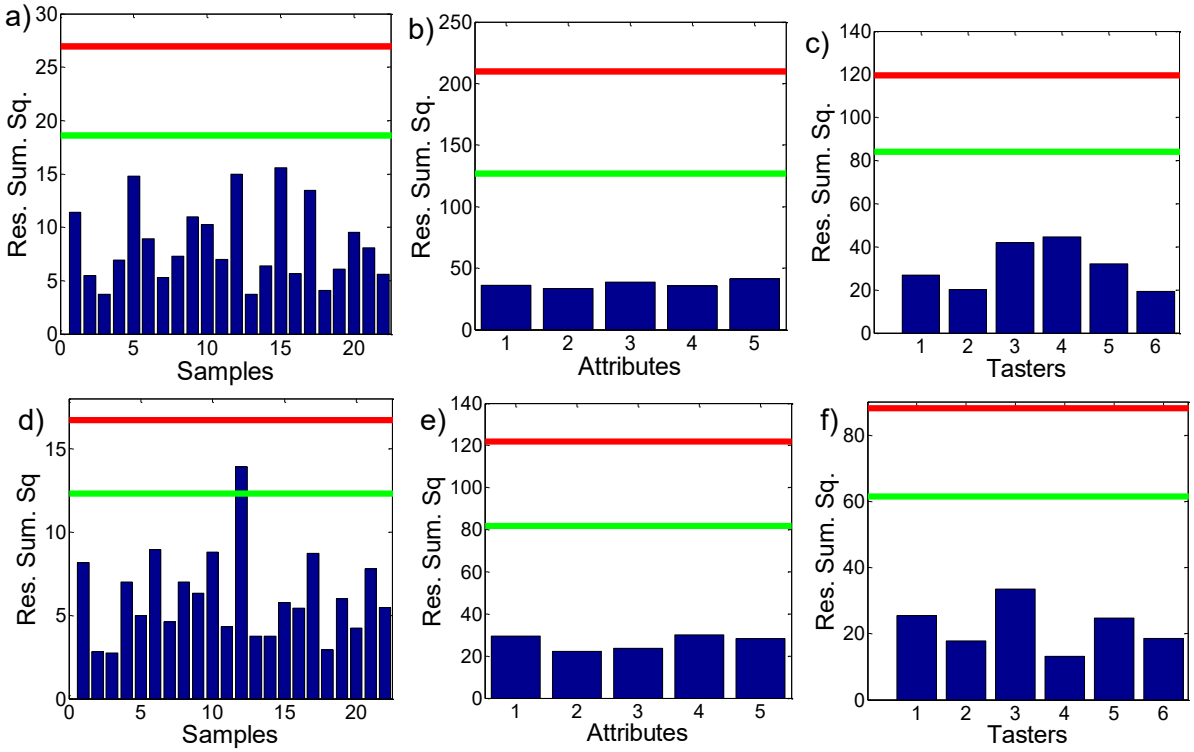

Figure 3.4. Residual sum of squares (RSS) for the **(a)** first (samples), **(b)** second (attributes) and **(c)** third (taster) modes for the two-factor PARAFAC model and **(d)** first (samples), **(e)** second (attributes) and **(f)** third (modes) for the three-factor PARAFAC model.. Green lines indicate confidence limits at 95% of significance level and red lines indicate confidence limits at 99% of significance level

Figure 3.4a and d shows a random distribution of the residuals as expected for the sample mode due to the nature of the variability of the samples. In general, no systematic trend was observed for the residuals of the three modes, indicating that this model is unbiased. For the two models, the residuals are within a 99% of confidence limit, however a better fitting is observed for the two-factor model, in which a 95% of confidence limit is observed for all the modes.

3.3.3.4 Split-half analysis

Finally, regarding split-half analysis, the data was divided into two halves and then two independent PARAFAC models are constructed. Due to the PARAFAC solution uniqueness, these models should provide the same results, i.e., the same estimated loadings should be obtained from the non-split and split models, considering any proper subset of the data, as long as the appropriate number of factors has been chosen. If split models differ, the number of factors used to build them is not appropriate (Smilde *et al.*, 2004).

In this example, the mode chosen to be divided was the first one, i.e., the sample mode. The split mode should present a number of independent variables enough to assure that each subset will provide estimates of the whole domain (Smilde *et al.*, 2004; Bro, 1997). Despite of the values from the other two modes (attributes and tasters), these variables are intrinsically qualitative, and therefore, the sample mode was the best choice for splitting.

Two and three-factor models were tested for split-half analysis with the respective subsets. Figure 3.5 presents the loadings of split-half models and of non-split unconstrained two-factor and three-factor models. It can be noted that in the upper plots, Figure 3.5 a-c, the profiles from the two data halves present a better correlation, whereas a significant difference is observed in the lower plots relative to three-factor models, Figure 3.5 d-f. Thus, the split-half analysis suggests the two-factor model as the most appropriate since loading vectors remained almost the same regardless the data subset used to build them. The deviations between the two sets of estimates may be attributed to the noise, the nature variation of the data and the small sample size (Smilde *et al.*, 2004).

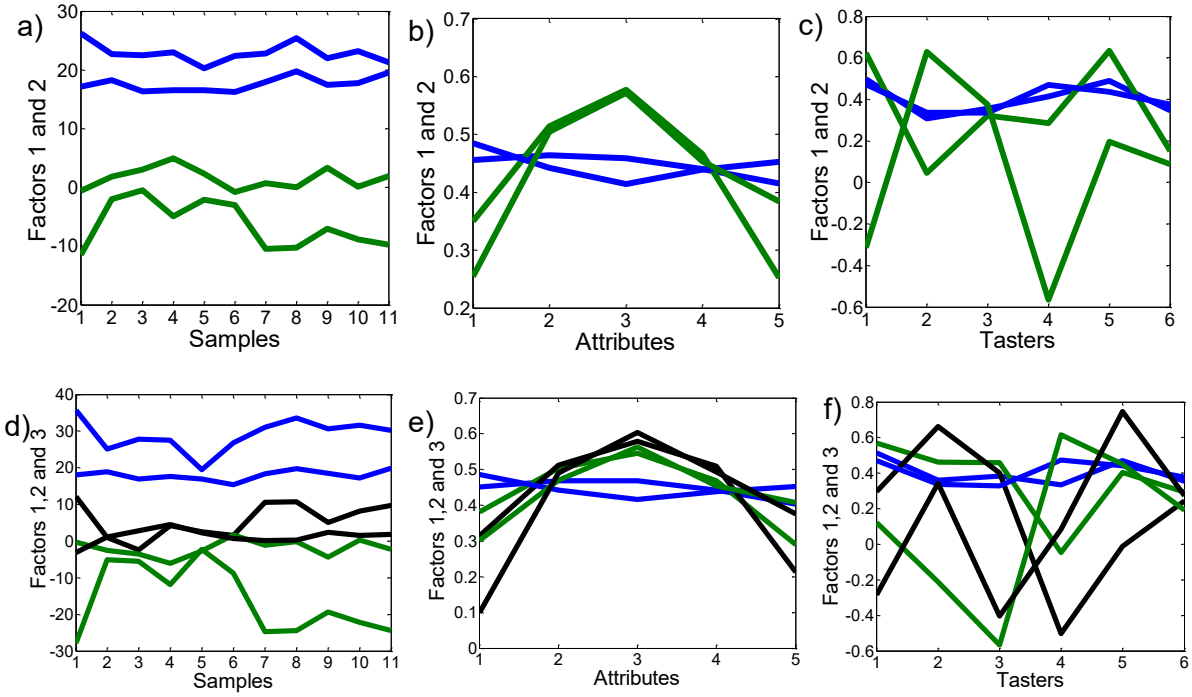

Figure 3.5. Result of split-half analysis of a two- and a three-factor PARAFAC model of sensory data **(a), (b)** and **(c)** loadings for first (samples), second (attributes) and third mode (tasters) for the two-factor PARAFAC model and **(d), (e)** and **(f)** loadings for first (samples), second (attributes) and third mode (tasters) for the three-factor PARAFAC model

3.3.4 Discrimination of coffee capsules

3.3.4.1 Data

The mean values of the 22 samples of coffee capsules are shown in Table 3.3 for the five attributes evaluated by the six tasters. A one-way analysis of variance (ANOVA) at a confidence level of $\alpha = 0.05$ pointed out a significant difference between samples. This result reflects the variability of the samples and is consistent with the data provided by the manufacturers. Table 3.4 shows the parameters for the ANOVA model of the data shown in Table 3.3.

Table 3.3. Mean and standard deviations of the samples for the five evaluated attributes

Samples	*Attributes (Mean ±SD; n=6)*				
	Aroma	Flavor	Aftertaste	Acidity	Body
S1	3.6±0.6	2.9±0.8	2.6±0.9	2.6±0.8	3.2±0.8
S2	3.8±0.5	3.9±0.6	3.9±0.5	3.8±0.6	3.5±0.7
S3	4.1±0.8	3.9±0.9	4.1±0.9	3.8±1.1	3.8±0.9
S4	3.8±0.4	3.6±0.5	3.3±0.2	3.3±0.7	3.1±0.6
S5	3.6±0.7	3.3±0.6	3.4±0.7	3.4±0.7	3.1±0.7
S6	3.8±0.4	3.7±0.7	3.3±0.6	3.4±1.0	3.6±0.8
S7	2.7±0.4	2.5±0.6	2.2±0.7	2.4±0.7	3.0±0.5
S8	3.1±0.6	3.1±0.6	2.9±0.8	3.0±0.4	3.3±0.4
S9	3.6±0.5	2.6±0.6	2.5±0.9	2.6±0.5	3.1±0.8
S10	3.2±0.7	2.6±0.5	2.4±0.4	2.8±0.5	3.2±0.5
S11	2.5±0.5	2.1±0.5	1.9±0.5	2.1±0.5	3.1±0.8
S12	3.6±0.5	2.9±0.8	2.8±0.8	3.0±0.8	3.1±0.8
S13	3.7±0.5	3.4±0.4	3.3±0.2	3.3±0.1	3.2±0.3
S14	3.2±0.5	3.0±0.4	2.7±0.6	2.9±0.4	3.2±0.4
S15	3.3±0.8	3.2±0.6	2.9±0.8	3.0±0.8	3.2±0.7
S16	3.2±0.3	2.9±0.4	3.0±0.5	3.0±0.5	3.2±0.3
S17	3.0±0.8	2.8±0.9	2.5±0.8	3.1±0.8	3.1±0.8
S18	3.6±0.6	3.2±0.3	3.2±0.3	3.0±0.4	3.5±0.5
S19	3.8±0.5	3.6±0.7	3.4±0.7	3.5±1.1	3.5±0.9
S20	3.5±0.7	3.1±0.6	2.9±0.5	3.4±0.4	3.3±0.5
S21	3.6±0.7	3.1±0.7	3.0±0.7	3.3±1.0	3.0±0.6
S22	3.8±0.6	3.8±0.4	3.5±0.3	3.4±0.6	3.4±0.7
Min.	2.5±0.5	2.1±0.5	1.9±0.5	2.1±0.5	3.0±0.5
Mean	3.4	3.2	3.0	3.1	3.3
Max.	4.1±0.8	3.9±0.6	4.1±0.9	3.8±1.1	3.8±0.9

Table 3.4. Parameters for the one-way ANOVA model of the coffee samples

Source of Variation	*Sum of squares (SS)*	*Degrees of freedom (df)*	*Mean square (MS)*	*F-value*	*P-value*	*F critical*
Between samples	2.777	4	0.694	3.925	0.005	2.458
Within samples	18.575	105	0.177			
Total	21.352	109				

Figure 3.6 shows the results for the sensory analysis of the coffee samples for the six evaluated attributes. These results showed the greatest variation for the attribute aftertaste, ranging from 1.5 to 4.1, and the smallest variation for the attribute body, ranging from 2.7 to 3.8. Outliers, i.e., values above the superior limit or below the inferior limit, were not detected. A David Test at a confidence level of α = 0.05 confirmed a normal distribution of the data (David *et al.*, 1954).

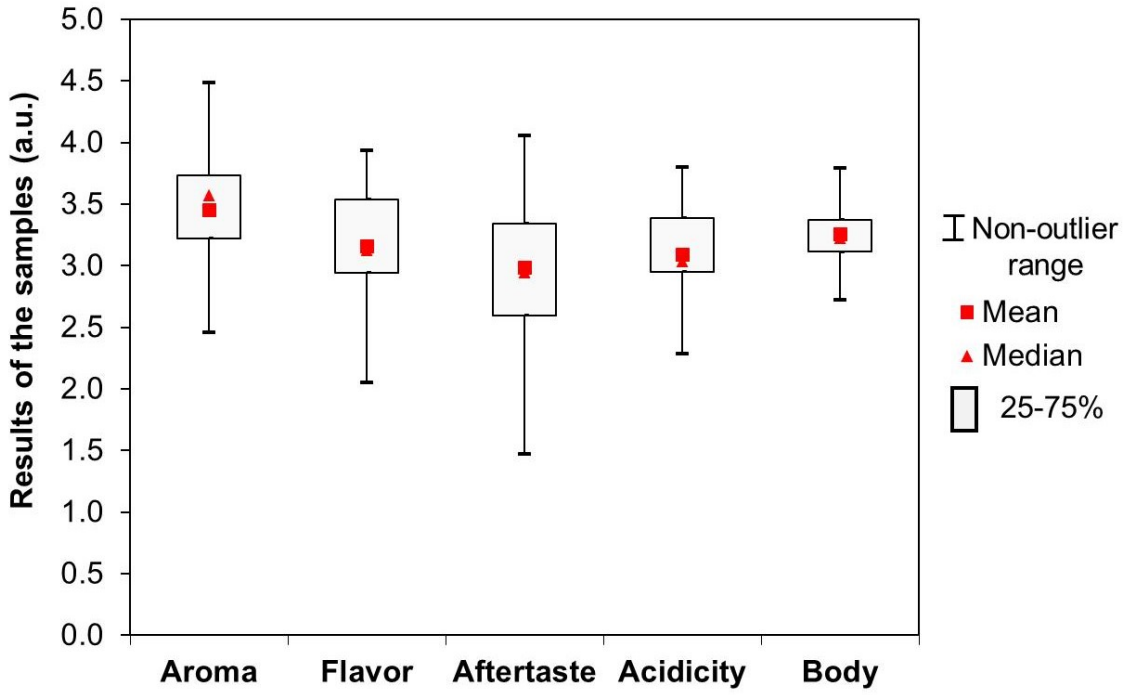

Figure 3.6. Results (a.u.) for the sensory analysis of the samples for the five evaluated attributes. Squares (in red) represent the mean, triangles (in red) represent the median; boxes indicate the 25–75% range of the distribution, and whiskers represent minima and maxima values

3.3.4.2 PCA model

As mentioned, PCA is the most used chemometric method in the world. In summary, PCA is an unsupervised pattern recognition method that seeks natural groupings through preserving the maximum of variance between groups. A PCA model was constructed with the mean of the samples for the five attributes (Table 3.4) without any pre-processing method since all attributes were judged in the same units and range.

The model with two principal components accounted for 99.87% of the total variance. Figure 3.7 presents the plots of scores and loadings for PC1 versus PC2.

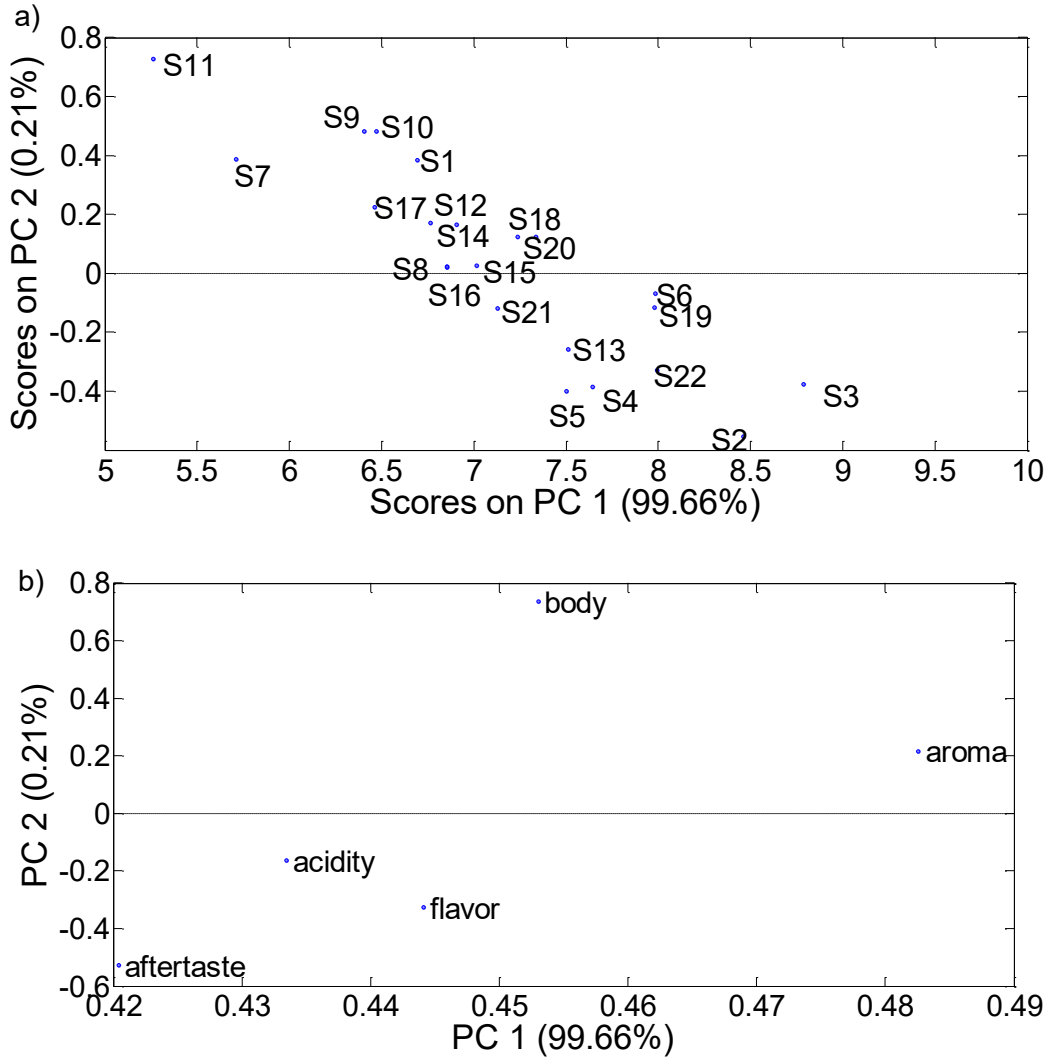

Figure 3.7. **(a)** PCA scores plot and **(b)** PCA loadings plot for PC1 versus PC2

In the scores plot, Figure 3.7a, it was not possible to observe a clear discrimination among samples. However, the PCA loadings, Figure 3.7b, provided an explicit contrast between the attributes related to taste (flavor, aftertaste, acidity), showing negative loadings, from the other ones (aroma and body) presenting positive loadings. For assessing each attribute, it is a hard task since they reflect a bulk chemical composition with unique characteristics that lead to different sensations in each taster. Aroma and body may be difficult to evaluate by the tasters since they embrace more than one human sense. Aroma, for instance, is a joint perception of taste and smell (Kilcast, 2010).

3.3.4.3 PARAFAC model

Aiming to have a better comprehension of the data, PARAFAC models were constructed and a two-factor unconstrained PARAFAC model was selected. This model explained 97.4% of sensory data variability with a CORCONDIA of 100%. Figure 3.8 presents the loadings for the three modes of this PARAFAC model.

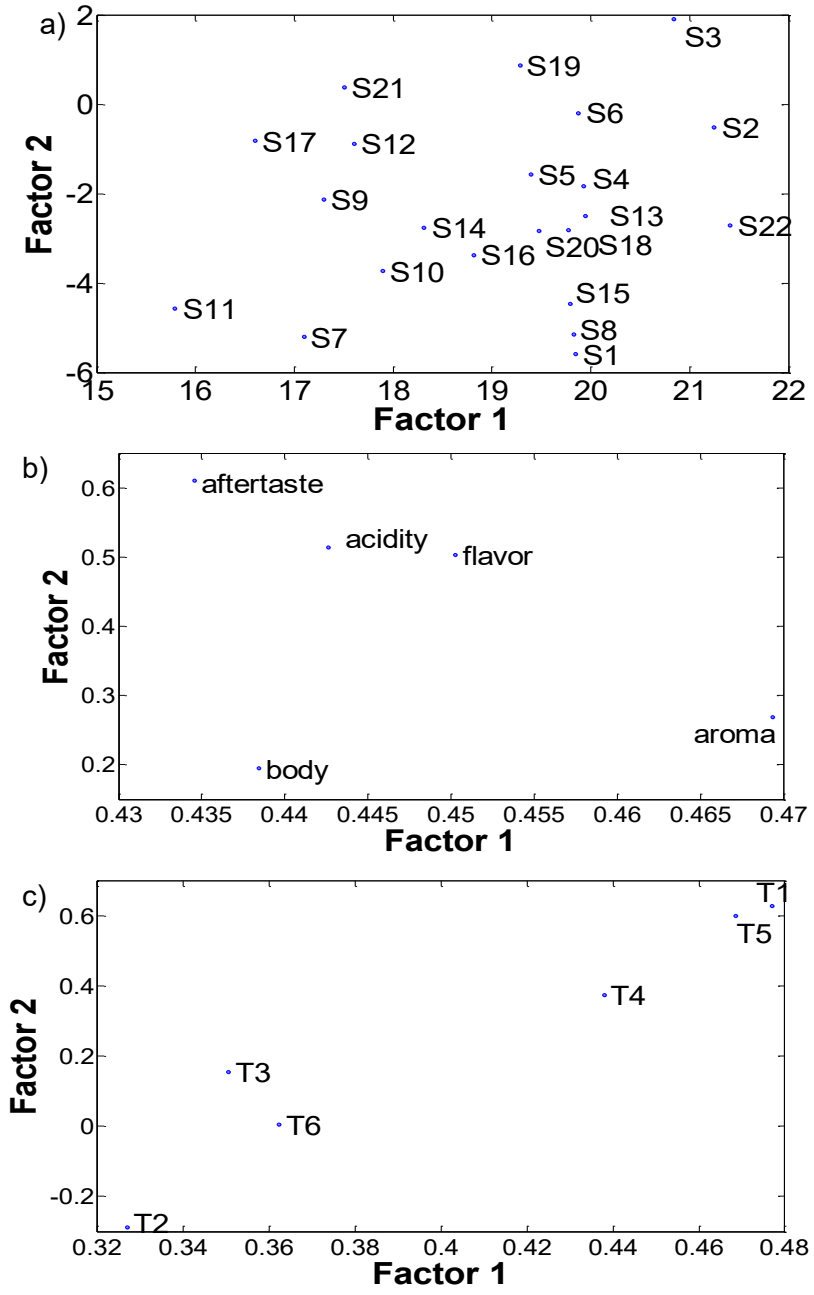

Figure 3:8. **(a)** Loadings for first (samples), **(b)** second (attributes) and **(c)** third (tasters) modes of the unconstrained two-factor PARAFAC model

Figure 3.8a shows the loadings for the first mode, i.e., samples. Similar to PCA results, it is not possible to visualize a clear separation among the samples, mainly due to the large variability existing for this dataset. However, some trends can be inferred based on the estimated loadings.

Samples S3, S19, S6, and S2, for instance, presented the highest values for all the attributes, and sample S3 particularly showed the highest ones. On the other hand, samples S7 and S11 exhibited the lowest values. Sample S11 showed the lowest values for the attributes aroma, flavor, aftertaste, and acidity, while sample S7 presented the lowest result for the attribute body. This pattern can be visualized in Figure 3.7a, which summarized the sensory behavior of the samples based on their scores in the two PARAFAC factors (more than 97% of the total variance). The group of samples with the highest attribute values is located in the upper-right corner of this plot, corresponding to the highest scores simultaneously on factors 1 and 2. On the other hand, samples S7 and S11 are located in the bottom-left corner, corresponding to the lowest scores simultaneously on factors 1 and 2. Therefore, the contrast of samples with higher and lower scores simultaneously on factors 1 and 2 is directly correlated with the attribute values of the samples.

Two other groups can be considered: the group with samples S1, S8, and S15; and the group with samples S9, S12, S17, and S21. To study these groups and to seek a common feature, Figures 3.8 and 3.10 show the mean of the attributes evaluated by the six tasters for the above mentioned samples. We can see that samples of each of these two groups presented similar behavior, whereas an ANOVA test confirmed the statistical equivalence among samples included in each of these two groups (Tables 3.5 and 3.6).

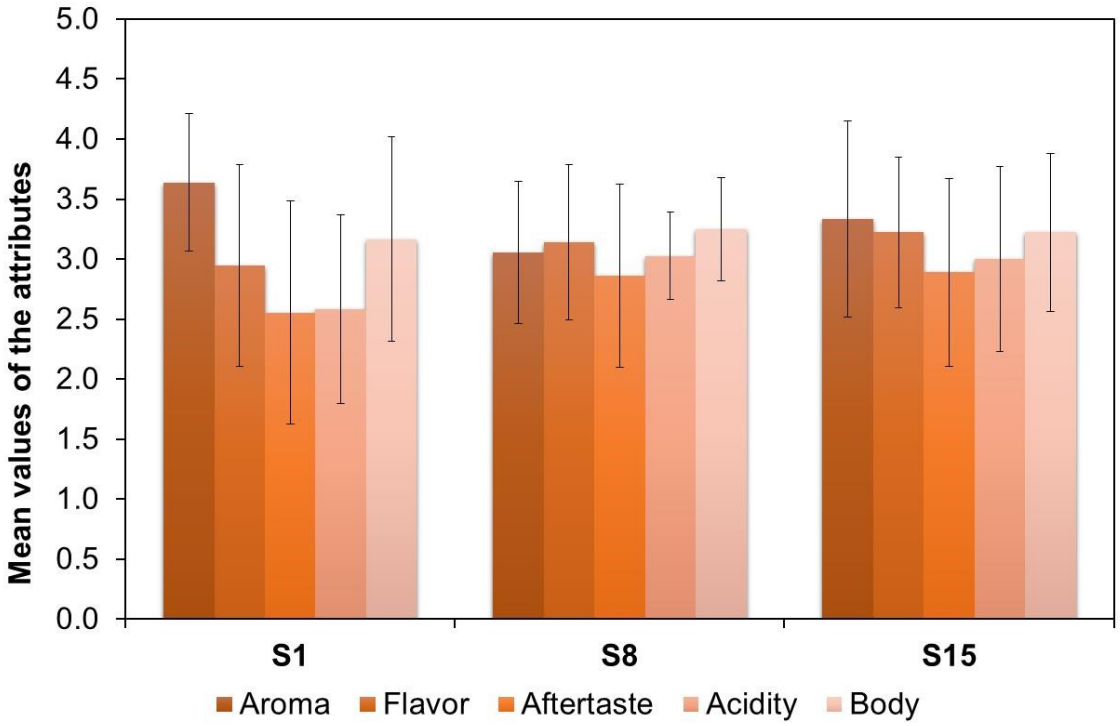

Figure 3.9. Mean values (n=6) of attributes for samples S1, S8, and S15

Table 3.5. Parameters for the one-way ANOVA model for samples S1, S8, and S15

Source of Variation	*Sum of squares (SS)*	*Degrees of freedom (df)*	*Mean square (MS)*	*F-value*	*P-value*	*F critical*
Between Groups	0.061	2	0.030	0.357	0.707	3.885
Within Groups	1.024	12	0.085			
Total	1.085	14				

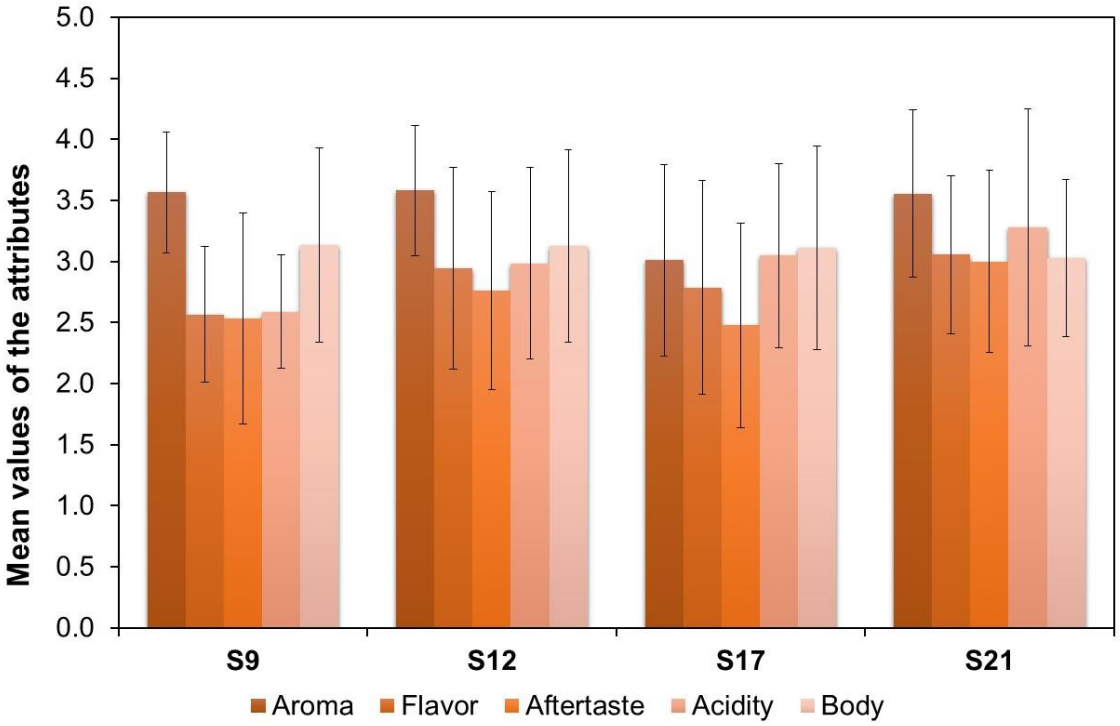

Figure 3.10. Mean values (n=6) of attributes for samples S9, S12, S17 and S21

Table 3.6. Parameters for the one-way ANOVA model for samples S9, S12, S17 and S21

Source of Variation	*Sum of squares (SS)*	*Degrees of freedom (df)*	*Mean square (MS)*	*F-value*	*P-value*	*F critical*
Between samples S9, S12, S17 and S12	0.336	3	0.112	1.044	0.400	3.239
Within samples S9, S12, S17 and S12	1.715	16	0.107			
Total	2.051	19				

In the second mode, i.e., attributes, the same separation observed by PCA was obtained by the PARAFAC model, Figure 3.8b. Finally, Figure 3.8c shows the loadings related to the tasters, third mode. Given that this is a group of tasters that received the same training evaluating the same samples for the same attributes, it would be expected similar behavior of the individuals acting as replicates; however, a clear separation is observed in this mode suggesting that the training was not sufficient to provide the same jugging of the coffee samples. Along with the first factor (and also the second factor), the tasters are divided into two groups: those with the highest sensory evaluation mean values (T1, T4, and T5) and those with the lowest sensory evaluation

mean values (T2, T3, and T6). Among these groups, tasters T1 and T5 presented statistically equal means (p value> 0.05 and F_{value}<$F_{critical}$). Figure 3.11 shows the mean of sample results per attribute for these tasters. One can see that the results are visually very similar as attested by ANOVA (Table 3.7). In conclusion, it was possible to discriminate two groups of tasters, the first one tending to assign higher attribute values to the samples, and another one tending to assign lower values to them. This last group tended to utilize body as the main attribute for discrimination. This observation can be interpreted by comparing attribute and taster loadings (Figs. 3.8b-c), and realizing that body in mode 2 and tasters T2, T3 and T6 in mode 3 presented simultaneously lower loadings on both factors 1 and 2.

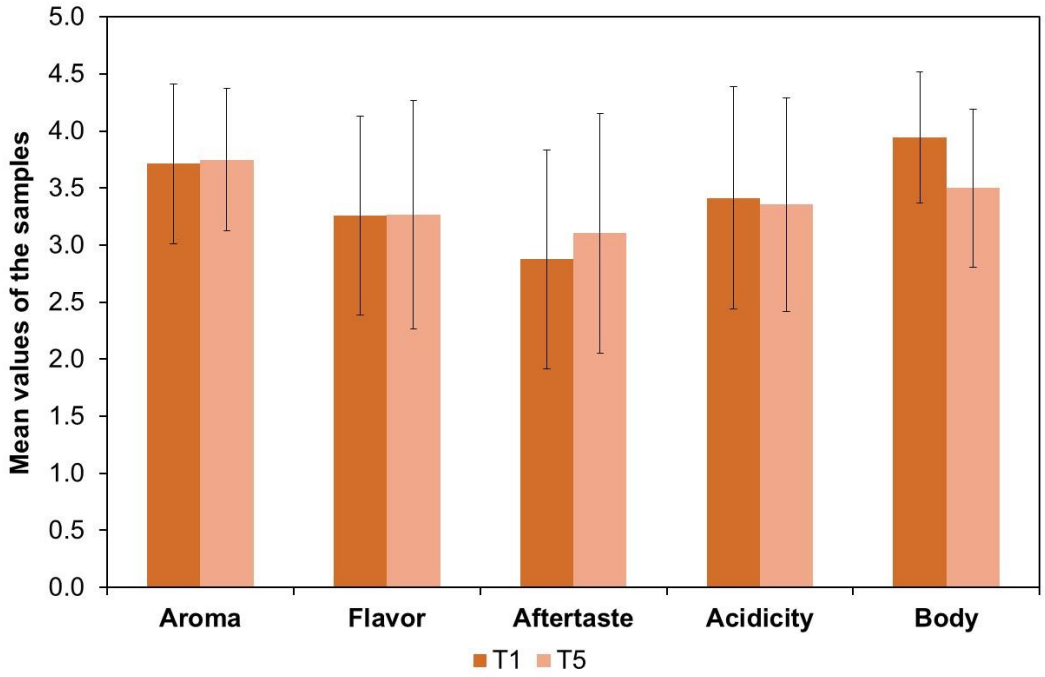

Figure 3.11. Mean values (n=22) of the samples per attribute for tasters T1 and T5

Table 3.7. Parameters for the one-way ANOVA model for tasters T1 and T5

Source of Variation	*Sum of squares (SS)*	*Degrees of freedom (df)*	*Mean square (MS)*	*F-value*	*P-value*	*F critical*
Between T1 and T5	0.005	1	0.005	0.046	0.835	5.318
Within T1 and T5	0.915	8	0.114			
Total	0.921	9				

Similarly, Figure 3.12 shows the mean of samples results per attribute for tasters T3 and T6, who also presented no significant difference (Table 3.8).

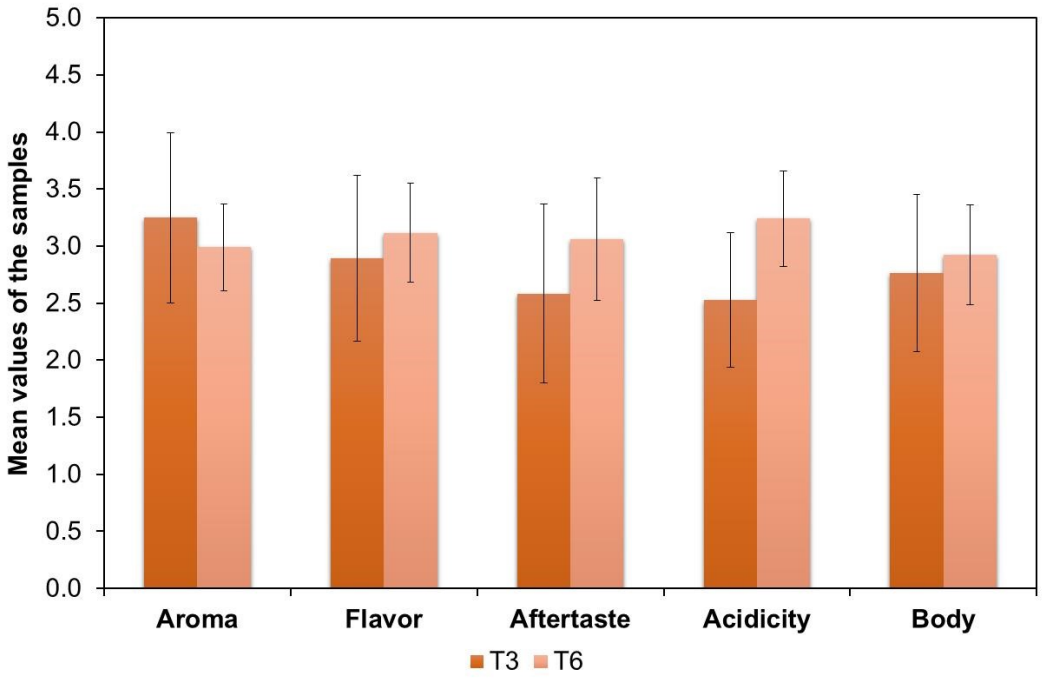

Figure 3.12. Mean values (n=22) of the samples per attribute for tasters T3 and T6

Table 3.8. Parameters for the one-way ANOVA model for tasters T3 and T6

Source of Variation	*Sum of squares (SS)*	*Degrees of freedom (df)*	*Mean square (MS)*	*F-value*	*P-value*	*F critical*
Between T3 and T6	0.173	1	0.173	3.550	0.096	5.318
Within T3 and T6	0.390	8	0.049			
Total	0.563	9				

3.4. Conclusions

A PARAFAC model built with two factors showed to be a powerful tool to interpret this sensory dataset. In this application, the essential features to construct a PARAFAC model were described as a tutorial aiming to help and spread its application among professionals unfamiliar with multiway chemometric tools. Frequently, chemometric

tools are used without an in-depth previous knowledge of their bases, hindering the obtention of accurate and chemically meaningful interpretation of the results. The satisfactory results obtained here highlighted PARAFAC as a useful tool for relating sensory attributes to taster preferences, and simultaneously considering the nature of the samples since the model provided a general visualization of the three-way data.

4. SPECTROFLUORIMETRIC DETERMINATION OF PHENYLALANINE IN HONEY BY THE COMBINATION OF STANDARD ADDITION METHOD AND SECOND-ORDER ADVANTAGE

4.1 Introduction

Honey is a complementary food, beneficial to human health, and its composition varies according to botanical origin, agricultural practices, climatic and environmental conditions (Kavanagh *et al.*, 2019; Noviyanto & Abdulla 2020). It is composed mainly by carbohydrates (approx. 95 to 99% of the dry matter), water, and other minor compounds, such as organic acids, amino acids, proteins, minerals, vitamins, and lipids. In average terms, the nectar honey is composed by fructose (38.2%), glucose (31.3%), sucrose (0.7%), water (17.2%), minerals (0.1-0.5%), amino acids and proteins (0.2-0.4%) and organic acids (0.2-0.8%). The main amino acids in honey are phenylalanine, proline, and glutamic acid (Roldán *et al.,* 2011; Rebane & Herodes 2010; Iglesias *et al*., 2006).

Honey presents several pharmacological properties, such as antioxidant, antimicrobial, antidiabetic, anti-inflammatory activity, and wound healing (Leyva-Jimenez *et al.*, 2019). Pharmacological properties are frequently attributed to honey minor compounds, which are usually present at concentration levels below 0.001 g L^{-1}. Therefore, sensitive techniques able to detect and quantify these compounds are required. Analytical techniques most applied to the determination of organic compounds in honey are gas and liquid chromatographies (Leyva-Jimenez *et al.*, 2019; Shen *et al.,* 2019; Allenspach *et al.*, 2018; Chen *et al.*, 2016; Otter, 2012; Kelly *et al.*, 2010). Despite their high sensitivity and selectivity, chromatographic techniques are expensive, consume solvents, require laborious sample preparation and well-trained professionals, resulting in low throughput analysis. On the other hand, spectroscopic techniques have been exploited as simple, rapid, and environmentally friendly analytical alternatives, since they allow simultaneous multicomponent analysis with minimum sample preparation. Specifically, molecular fluorescence spectroscopy is a simple, fast, and sensitive technique that can be applied to qualitative and quantitative analysis of fluorophores, molecules that show fluorescence, through the relationship between absorbed and emitted energy (Lakowicz 2006).

Over the last decades, fluorescence spectroscopy has become popular for the analysis of food, environmental and biological samples, mainly due to its wide linear response

range, high sensitivity, selectivity, and low cost in comparison to chromatography techniques (Christensen *et al.*, 2006; Lakowicz 2006). Its potential has been improved when combined with chemometric methods. The combination of spectrofluorimetry and chemometrics has been applied to direct determinations in honey and other food matrices (Lenhardt *et al.,* 2015, Lenhardt *et al.*, 2017; Sádecká & Tóthová, 2007; Ruoff *et al.,* 2005).

Multiway calibration methods can be applied to process higher-order spectrofluorimetric data arrays, allowing to determine the number of fluorescent components, to estimate their pure spectra, and to determine their relative concentrations in each sample. PARAFAC has been the most used method for this task (Bro, 1997). Other methods, such as multivariate curve resolution-alternating least squares (MCR-ALS) (de Juan *et al.*, 2014), and N-way partial least squares (N-PLS) (Bro, 1996), have also been employed.

The determination of amino acid profile using spectrofluorimetry has been explored for the study of the botanical and geographical origin of food and for the detection of food fraud (Strelec *et al.*, 2018; Siddiqui *et al.*, 2017; Lenhardt *et al.*, 2015; Sergiel *et al.*, 2014; Ruoff *et al.*, 2005). Honey contains several intrinsic fluorophores, such as proteins, peptides, and free amino acids, which can be used as markers for characterization studies (Lenhardt *et al.*, 2015; Ruoff *et al.,* 2005). Several articles have determined botanical origin and checked authenticity of honey samples using different analytical techniques (Shen *et al.*, 2019; Kavanagh *et al.,* 2019; Lenhardt *et al.*, 2015; Karoui *et al.,* 2007), however most of them focused on discriminant analysis by comparing honey profiles instead of quantifying specific analytes.

Among the main amino acids present in honey, phenylalanine is the only intrinsic fluorophore. Thus, it can be directly determined by fluorescence spectroscopy, which represents a promissory alternative for honey characterization. Phenylalanine is an essential aromatic amino acid, i.e., it must be ingested by diet and cannot be synthesized by the human body. It can be found in two isomer forms, D–phenylalanine and L-phenylalanine, Figure 4.1, the latter being the most common in food, such as meat, fish, egg, honey, peanut, and milk derivatives (Perkowski & Warpeha, 2019). In addition, phenylalanine has been considered by some authors a discriminant marker to trace the floral origin of rosemary, eucalyptus, lavender, thyme, and orange blossom honeys, among others (Biluca *et al.*, 2019; Chen *et al.,* 2016; Rebane & Herodes, 2008).

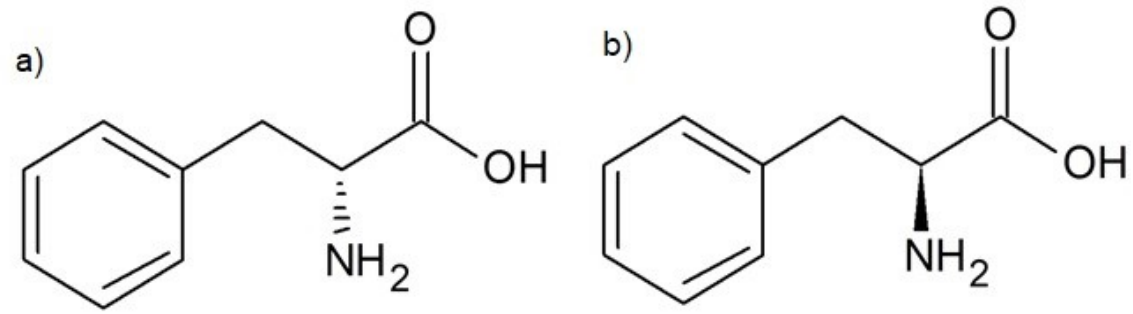

Figure 4.1. Structural formula of phenylalanine optical isomers. **(a)** D-phenylalanine **(b)** L-phenylalanine

Thereby, this chapter proposed a new direct spectrofluorimetric method for the quantification of phenylalanine in polyfloral honey using PARAFAC, a second-order calibration method. The content of this natural component is an important parameter that contributes to the characterization of honeys in relation to their pharmacological properties, their botanical and geographical authentication, and the possible detection of frauds. Due to the presence of a matrix effect that changes the fluorescence signal of phenylalanine depending on the honey composition, a strategy of second-order standard addition was adopted. Considering the variability in honey composition, an important advantage of this strategy is the possibility to quantify the analyte in the presence of uncalibrated interferences. This chemometric strategy, which has been applied to clinical analysis (Sena *et al.*, 2006; Bernardes *et al.*, 2010; Privitera & Lozano 2017), is applied here for the first time concerning to the authors knowledge to food analysis. The developed method is a clean, low cost, and rapid alternative to the traditional protocols.

4.2 Materials and Methods

4.2.1 Sample Preparation

Honey samples were purchased from local markets and stored at room temperature (25ºC) in the original packages until analysis. Nine samples from different origins and brands were analyzed. Table 4.1 presents botanical and geographical origins of the honey samples certified by the respective producers. Aliquots of 2.00 g of each sample

were placed into a 50-mL polypropylene tubes with 40.00 mL of deionized water (Mili-Q, Merck, Darmstadt, Germany). Sample solutions were homogenized and stored at 4±2°C until analysis by molecular fluorescence spectroscopy. A standard solution of L-phenylalanine (Synth, São Paulo, Brazil) 20 mg L^{-1} was also prepared in deionized water.

Table 4.1. Botanical and geographical origins of the honey samples analyzed in this study, certified by the producers

Samples	*Botanical origin*	*Geographical origin*
S1	Polyfloral honey	National Beekeeping Cooperative, Minas Gerais, Brazil
S2	Polyfloral honey	Southeast of Minas Gerais, Minas Gerais, Brazil
S3	Polyfloral honey	National Beekeeping Cooperative, Minas Gerais, Brazil
S4	Polyfloral honey	National Beekeeping Cooperative, Minas Gerais, Brazil
S5	Wildflower honey - Honey produced with nectar from different flowers	National Beekeeping Cooperative, Minas Gerais, Brazil
S6	Wildflower honey	Midwest of Minas Gerais, Brazil
S7	Polyfloral honey	National Beekeeping Cooperative, Minas Gerais, Brazil
S8	Polyfloral honey	National Beekeeping Cooperative, Minas Gerais, Brazil
S9	Predominantly wild flowering honey	Santa Barbara, Minas Gerais, Brazil

4.2.2 Fluorescence Analysis

Fluorescence measurements were performed using a Varian Cary Eclipse fluorescence spectrophotometer (Walnut Creek, California, USA) equipped with two Czerny-Turner monochromators, and a Xenon discharge lamp pulsed at 80 Hz with a half peak height of 2 ls (peak power equivalent to 75 kW). A high-performance R298

photomultiplier tube detector was used for collection of the fluorescence spectra. The temperature was hold at 25.00 ± 0.05 ºC using a Peltier thermostat. The spectrometer was interfaced with a computer using Cary-Eclipse 1.1 (132) scan software for spectral acquisition and exportation. EEM were obtained by varying the excitation wavelength (λ_{ex}) between 280 and 350 nm (step 10 nm) and recording the emission spectra (λ_{em}) from 360 to 490 (step 1 nm). Excitation and emission slits were both set at 5 nm, and the scan rate was fixed at 600 nm min^{-1}.

Rayleigh scatter was removed from all PARAFAC models through the selection of a proper working spectral region in EEM. The characteristic region of elastic scattering provides a non-trilinear analytical signal unrelated to the sample composition. Therefore, this region should be removed before building multilinear PARAFAC models (Bro, 1997).

4.2.3 Phenylalanine Quantification

Quantification of phenylalanine was carried out by the standard addition method, in a situation of variable total volume with continuous variation of standard (Bader 1980). Four constant volumes of the standard solution were directly added in the cuvette, thus simplifying the experimental procedure, and eliminating the need to prepare one solution for each addition. In this case, the standard addition curve is based on Equation 4.1 (Bader, 1980):

$$R = k\left[\frac{V_x C_x}{V_x + NV_s} + \frac{NV_s C_s}{V_x + NV_s}\right] \quad (4.1)$$

where R is the instrumental response, here the sample scores of phenylalanine obtained by PARAFAC, V_x and C_x are the initial volume and concentration of the sample, C_s is the concentration of the standard solution, V_s is the volume of each standard addition, N is the number of the addition (from 0 to 4) and k is a constant of proportionality.

Each honey sample was analyzed in triplicate. For each measurement, 2.00 mL of a diluted honey sample was placed into a 10.00 mm quartz cuvette containing a micro magnetic bar, and the spectrum was recorded. In the sequence, four successive additions of 10 µL of a 20 mg L^{-1} L-phenylalanine solution were performed, and the

respective spectra recorded. After each addition, samples were homogenized under magnetic stirring, and further analyzed. Figure 4.2 shows a scheme summarizing the experimental procedure adopted for the developed method.

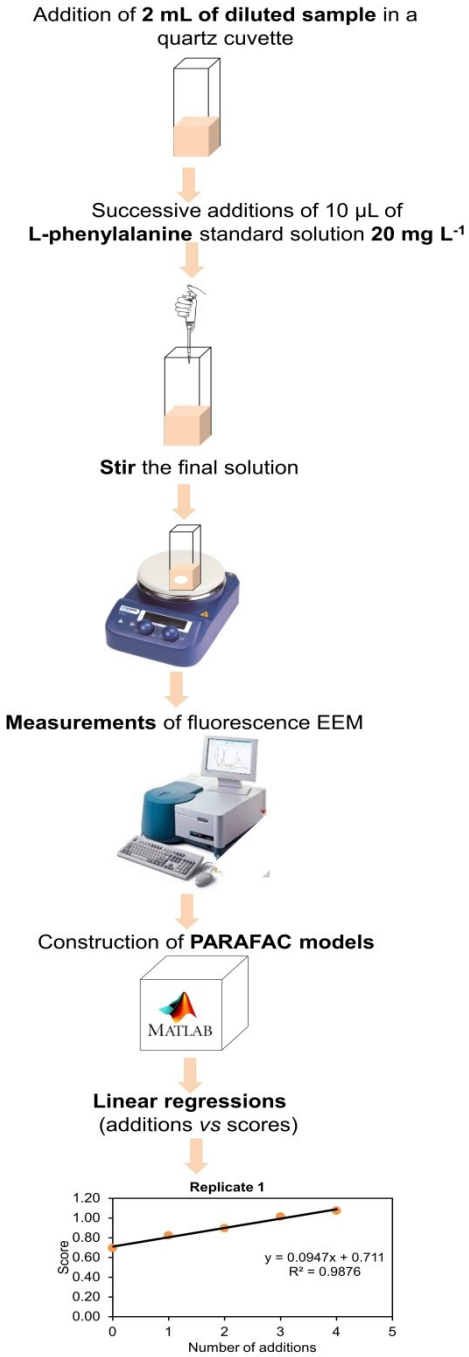

Figure 4.2. Scheme summarizing the experimental procedure for the developed spectrofluorimetric method

4.2.4 HPLC Analysis

Chromatographic measurements were carried out on a liquid chromatograph aiming at amino acid quantification by a reference method. A liquid chromatograph Alliance HT Waters e2795 (Milford, Massachusetts, USA) in-line degasser coupled to a fluorescence detector Waters 2475 was utilized. A reversed-phase Nova-Pak C18 column with an average particle size of 4 µm and dimensions of 150 mm x 4 mm, and a fluorescence detector set at λ_{exc} 260 nm and λ_{em} 300 nm were used. A 200 µL aliquot of each sample was dissolved with 0.1% (v/v) TFA (Sigma-Aldrich, Missouri, USA) in HPLC-grade water: 0.1% (v/v) TFA in acetonitrile (30:70 v/v). The chromatographic method was based on the procedure published by Allenspach et al. (2018). The elution program was an isocratic elution with 85% of 0.1% (v/v) TFA in water–15% 0.1% (v/v) TFA in acetonitrile for 25 minutes with a flow rate of 0.3 mL min^{-1}. The injection volume was 4 µL at 30±5°C.

4.2.5 Data Analysis and Software

Data were processed using Matlab software, version 7.10.0 (R2010a) (MathWorks, Natick, USA), along with the PLS_Toolbox package, version 6.71 (Eigenvector Technologies, Manson, USA). The graphical interface Multivariate Calibration 2 (MVC2), written by Olivieri *et al.* (2009), was employed for estimating analytical figures of merit and concentration standard errors. This freely available toolbox for MATLAB allows data processing using several second-order multivariate calibration methods, such as PARAFAC, MCR-ALS and N-PLS.

4.2.6 Combination of Second-Order Advantage and Standard Addition Method

The use of first-order calibration methods, such as in traditional multivariate calibration with PLS, allows the quantification of a given analyte without the need of physical separation. However, the interferences must be present in the calibration set. The use of some second-order calibration methods, such as PARAFAC and MCR-ALS, provides the second-order advantage, which permits the mathematical separation of the analyte signal and its determination in the presence of uncalibrated sample components. This ability of determining analytes in the presence of unknown interferences is attractive from the point of view of green chemistry (Olivieri & Escandar, 2019).

The standard addition method is traditionally used in univariate calibration aiming to cope with matrix effects that change the analyte signal from sample to sample. The primitive version of the second-order standard addition method (SOSAM) was proposed in 1995 by applying direct trilinear decomposition (DTD) to kinetic-spectrophotometric data (Booksh *et al.*, 1995). More recently, the combination of PARAFAC and standard addition method has been successfully applied to the determination of drugs in complex biological matrices (Sena *et al.*, 2006, Bernardes *et al.*, 2010).

4.3 Results and Discussion

4.3.1 Preliminary Exploratory Analysis

Aiming to assess the most appropriate excitation-emission wavelength range for this analysis, a preliminary scanning was performed based on the parameters found in the literature (Stanković *et al.,* 2019; Bong *et al.*, 2016; Lenhardt *et al.*, 2015; Chen *et al,.* 2014; Karoui *et al.*, 2007; Sádecká & Tóthová, 2007). Figure 3.3 shows the EEM of sample S2 as a contour map and as emission spectra at different excitation wavelengths (excitation range of 250-350 nm, and emission range of 300-800 nm).

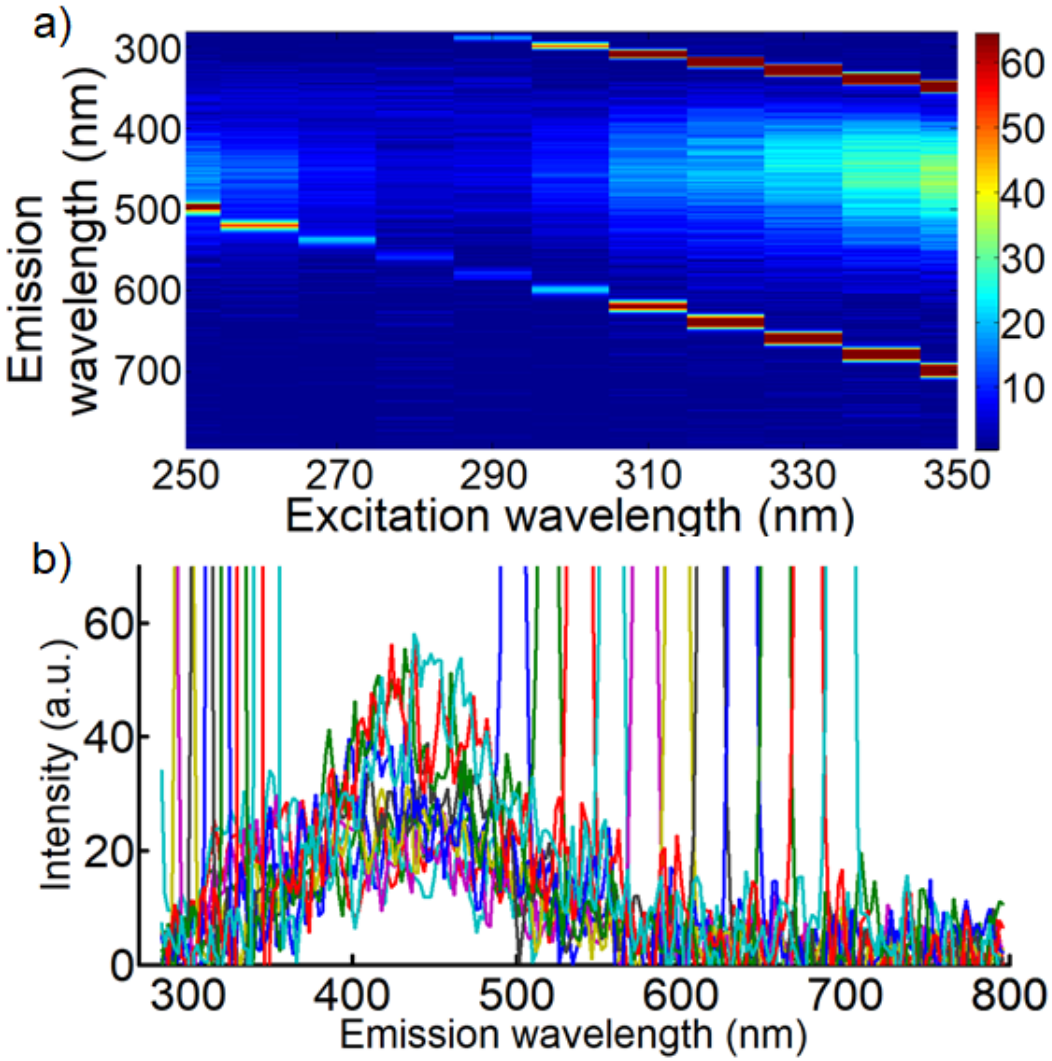

Figure 4.3. Excitation-emission matrix (EEM) of sample S2 as **(a)** a contour map and **(b)** as emission spectra at different excitation wavelengths. Excitation and emission wavelength ranges between 250-350 nm and 280-800 nm, respectively

It can be noted that the honey fluorescence band is mainly contained in the emission range between 300-550 nm. Thus, the emission wavelength range between 360 and 490 nm was chosen for honey analysis aiming to provide rapid measurements, and to avoid the scattering region between 300-350 nm and the second-order diffraction above 490 nm. The excitation wavelength range was set between 280-350 nm with a step of 10 nm, aiming to avoid as much as possible the scattering region and providing a short time of analysis (Lakowicz, 2006; Christensen *et al.,* 2006; Ruoff *et al.,* 2005).

4.3.2 Fluorescence Fingerprints

Figure 4.4 shows the EEM recorded for honey samples as contour maps after removing Rayleigh scattering. As can be observed, the emission patterns and intensities varied among samples due to their different botanical origin as previously reported in the literature (Lenhardt *et al.,* 2015; Chen *et al.*, 2014; Karoui *et al.*, 2007; Sádecká & Tóthová, 2007).

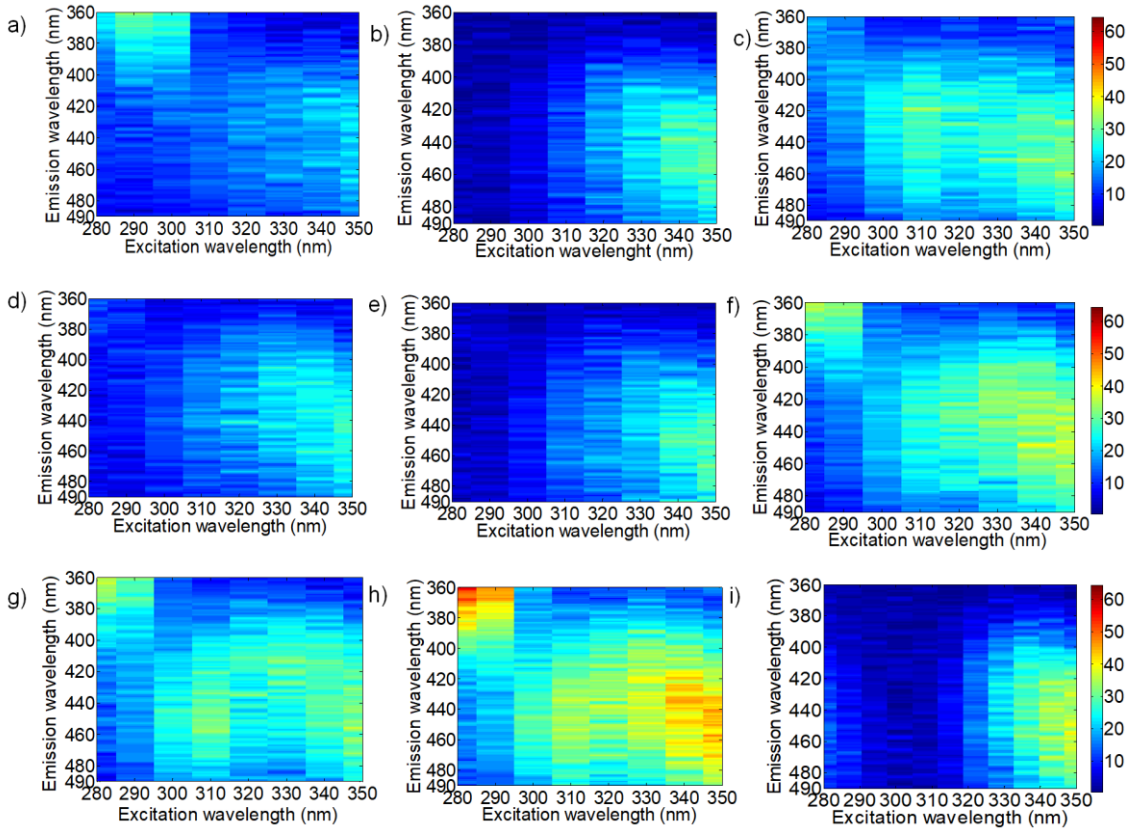

Figure 4.4. Fluorescence contour maps obtained for honey samples after removing Rayleigh scattering. **(a)** S1, **(b)** S2, **(c)** S3, **(d)** S4, **(e)** S5, **(f)** S6, **(g)** S7, **(h)** S8 and **(i)** S9

A spectral band is present in all the honey samples in the excitation range between 300-350 nm and in the emission range between 400-470 nm. This region may be assigned to phenolic compounds that have often been associated to the botanical origin of honeys (Stanković *et al.,* 2019; Lenhardt *et al.*, 2015; Karoui *et al.*, 2007). Another spectral band present in some samples is in the emission range between 360–

400 nm, at excitation wavelengths from 280 nm to 300 nm. This analytical signal may be assigned to aromatic amino acids (Stanković *et al.*, 2019; Lenhardt *et al.*, 2015). In the sequence, PARAFAC models were built to determine the number of fluorescent patterns and to estimate their resolved spectra.

4.3.3 PARAFAC Models

Three individual PARAFAC models were built for each sample, one for each replicate. For each model, three-way data arrays of dimensions 5 x 130 x 8 were analyzed. In these arrays, the first mode stands for number of measurements, i.e., the original sample plus four standard additions, the second and third modes represent the number of emission and excitation wavelengths, respectively. Non-negativity constraint was applied to all the three modes. The number of factors for each model was determined based on several criteria, such as the relation between the number of factors and the core consistency diagnostic (CORCONDIA) values (Bro & Kiers, 2003), observation of estimated loadings/resolved spectra, split-half analysis, and model residuals (Bro, 1997).

Since each sample can show specific spectral features (Figure 4.4) and due to the presence of matrix effect, the number of factors may change from sample to sample. For all the samples, the best PARAFAC models were selected with two or three factors, with the corresponding CORCONDIA values varying from 62% to 99%, and with mean of 73%. All these models accounted for at least 98% of the spectral data variance. Figure 4.5 shows excitation and emission loadings estimated by PARAFAC for one replicate of one of the analyzed honey samples (S4), as well as its fluorescence surface.

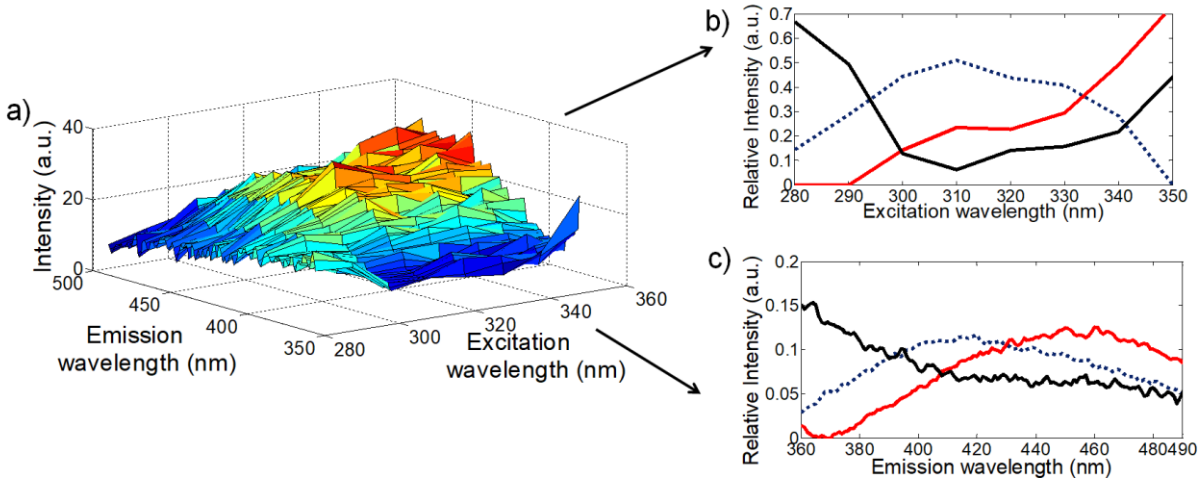

Figure 4.5. **(a)** Fluorescence surface, and **(b)** excitation and **(c)** emission loadings estimated by PARAFAC for honey sample S4

Two factors are common in all these models. The first factor (black solid lines in Figure 3.5b-c) showed two local excitation maxima at 280 nm and 350 nm, and an emission maximum around 380 nm. This factor was assigned to phenylalanine based on spectral comparison with the literature (Stanković *et al.,* 2019; Chen *et al.*, 2016; Lenhardt *et al.*, 2015; Karoui *et al.,* 2007; Pérez *et al.*, 2007; Hermosín *et al.*, 2003). This factor was also the only one associated with a significant increase in sample loadings (scores) as a function of standard additions. Excitation maxima at 258 nm (Christensen *et al.*, 2006), 260 and 282 nm (Lakowicz, 2006), and emission maxima at 284 nm (Christensen *et al.,* 2006) and 280 nm (Lakowicz, 2006) for phenylalanine in water have also been reported. Interactions of phenylalanine with other components of the matrix such as honey and honeydew can result in shifts of the spectral bands to neighboring wavelengths. This results in a matrix effect that prevents the use of external calibration curves to quantify phenylalanine in those matrices.

The second factor (red solid lines in Figure 4.5b-c) shows excitation maximum around 350 nm and emission maximum at 443 nm. This factor may be assigned to phenolic compounds that constitute the fluorescence background of honey samples. A variety of phenolic compounds are present in honey, such as hydroxycinnamic acids, chlorogenic acid, caffeic acid, coumarins, and stilbenes. Phenolic compounds have been considered as honey markers and their concentrations varies widely according to the floral origins (Kavanagh *et al.,* 2019; Stanković *et al.,* 2019; Bong *et al.*, 2016; Sergiel *et al.*, 2014). Thus, spectral profiles can vary depending on the honey samples. For

most samples, a third factor (blue dotted lines in Figure 4.5b-c) was also estimated, with excitation maximum at 310 nm, and emission maximum around 420 nm. This factor may be related to Maillard reaction products, such as furosine and hydroxymethylfurfural (HMF) (Christensen *et al.,* 2006).

4.3.4 Quantification of Phenylalanine by Second-order Standard Addition

A previous attempt to use an external calibration model with different honey samples demonstrated its infeasibility since the presence of matrix effect was observed. Then, the strategy of standard addition was employed. PARAFAC measurement loadings (first mode) of the first factor were used to construct univariate linear regressions/standard addition curves by applying Eq. 4.1, thus estimating phenylalanine concentration in each sample. Good correlations (correlation coefficients between 0.96 and 0.99) were observed between the instrumental response and the addition of phenylalanine standard solution for all the samples. Figure 4.6 shows the standard addition curve built with PARAFAC loadings for one replicate of honey sample S4. In this figure, it can also be observed the increase of phenylalanine fluorescence intensity with the phenylalanine standard solution addition (20 mg L^{-1}) through the spectral contour maps presented below the curve. Estimated phenylalanine concentration for each sample (n=3, mean ± standard deviations, SD) obtained from PARAFAC models are presented in Table 4.2.

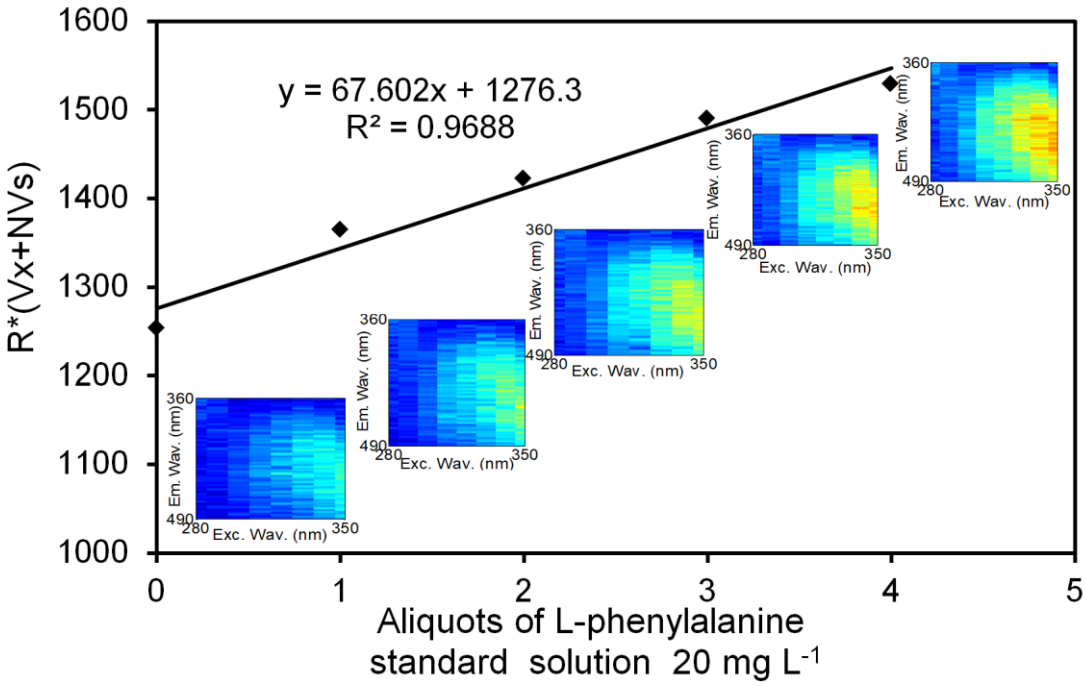

Figure 4.6. Standard addition curve built with PARAFAC loadings of the first factor from the first mode for honey sample S4. Insets show excitation-emission contour maps relative to each point of the curve.

Table 4.2. Phenylalanine concentration determined for each honey sample by spectrofluorimetric data and second-order standard addition, and verified by HPLC-UV (n=3) (Mean ± SD)

Samples	[Phenylalanine] (mg kg^{-1})	
	PARAFAC	HPLC-UV
S1	6.1±0.4	6.3±0.4
S2	9.3±1.0	9.6±0.7
S3	5.7±0.2	5.8±0.5
S4	8.2±0.6	8.9±0.2
S5	13.9±1.7	14.5±0.1
S6	32.5±3.4	32.1±0.6
S7	30.6±1.0	31.8±1.1
S8	8.5±0.2	8.2±0.3
S9	9.9±0.9	9.5±0.7
Mean	13.9	14.1
Median	9.3	9.5
Min.	5.7	5.8
Max.	32.5	32.1

Phenylalanine concentration varied significantly among samples, from 5.7 mg kg^{-1} to 32.5 mg kg^{-1}. This variation can be explained due to the different botanical and floral origins of the analyzed honeys. This analytical range agree with the literature for the phenylalanine concentration in other types of honey, such as polyfloral (7.92-140.65 mg kg^{-1}, Rebane & Herodes, 2010; 28.27±13.37 mg kg^{-1}, Iglesias *et al.*, 2006), honeydew (13.13-31.39 mg kg^{-1}, Rebane & Herodes, 2010), phacelia (20.41±11.99 mg kg^{-1}, Kuś *et al.*, 2018), heather (11.01-66.26 mg kg^{-1}, Rebane & Herodes, 2010), and dandelion (15.34-61.93 mg kg^{-1}, Rebane & Herodes, 2010).

The results of the developed spectrofluorimetric method were verified by HPLC-UV. Reference values provided by HPLC for each honey sample are also presented in Table 4.2. A paired t test at 95% significance level and with 8 degrees of freedom corroborated that there are no significant differences between the results of the two methods (estimated t of 1.20 versus tabulated t of 2.306).

4.3.5 Figures of Merit

Aiming to evaluate the performance of the developed spectrofluorimetric method, suitable figures of merit (FOM) were calculated (Olivieri, 2014). Table 4.3 shows the estimated values for precision, sensitivity (SEN), selectivity (SEL), the inverse of analytical sensitivity (γ^{-1}), limit of detection (LOD) and limit of quantification (LOQ).

Table 4.3. Figures of merit for the quantification of phenylalanine in honey by the developed spectrofluorimetric method

	Figures of merit								
	Samples								
	S1	S2	S3	S4	S5	S6	S7	S8	S9
Precision (%)[a]	6.6	10.7	3.5	7.3	12.2	10.5	3.3	2.4	9.1
Sensitivity	1333	1967	1600	2333	607	430	527	1900	1239
Selectivity	0.76	0.83	0.88	0.83	0.83	0.94	0.82	0.82	0.63
γ^{-1} (mg kg^{-1})[b]	0.02	0.02	0.02	0.02	0.06	0.07	0.07	0.01	0.02
LOD (mg kg^{-1})	0.3	0.1	0.1	0.2	0.3	0.5	0.4	0.2	0.01
LOQ (mg kg^{-1})	0.9	0.4	0.3	0.6	0.9	1.4	1.1	0.7	0.04

[a] relative standard deviation of triplicates. [b] inverse of analytical sensitivity

These FOM were specifically calculated for each sample. This is typical for second-order calibration models, since the composition of interferences can vary from sample to sample. Thus, mean values of FOM are not representative of the model. Precision was estimated as relative standard deviation (RSD) of triplicates for each sample. RSD varied between 2.4% and 12.2%. SEN is the variation of the analytical signal caused by a change in the analyte concentration (Olivieri, 2014). The large range of SEN values observed for honey samples (430 – 2333 a.u.) corroborated the presence of matrix effect and the need to employ the standard addition method. Analytical sensitivity (γ) is the relation between SEN and the instrumental noise, and allows comparing different methodologies, since it is independent on the food matrix and the analytical technique employed. The inverse of γ, γ^{-1}, is an estimate of the minimum concentration difference that the method can discriminate in the absence of experimental error. This inverse value also defines the number of decimal places that should be used to express the prediction results. Although the estimated inverses of analytical sensitivities varied at the second decimal place (0.01-0.07 mg kg^{-1}), we decided to express the results with only one decimal place as a more realistic representation. Finally, LOD/LOQ also varied to a great extent (0.1-0.5 mg kg^{-1}/0.3-1.4 mg kg^{-1}) and these values were considered satisfactory for practical purposes.

4.4 Conclusion

Determination of phenylalanine in honey is important because this analyte is one of the predominant free amino acids in this food matrix and has been considered a biomarker of its origin. This study developed a direct and environmentally friendly spectrofluorimetric method to determine phenylalanine in honey based on multiway calibration using PARAFAC. Due to the presence of a strong matrix effect, this determination cannot be performed using external validation. Thus, an analytical strategy combining PARAFAC and the standard addition method (second-order standard addition) was employed. This strategy utilizes the second-order advantage, which allows to quantify an analyte in the presence of uncalibrated interferences that can be present in different honeys. Honey samples from different botanical and geographical origins were analyzed and phenylalanine concentration ranged from 5.7 to 32.5 mg kg^{-1}, with a mean of 13.9 mg kg^{-1}. These results were verified by HPLC and the proposed method was validated by estimating proper figures of merit, such as precision, sensitivity, analytical sensitivity, limits of detection and quantification.

The developed method presented several advantages over the methods currently used for this determination, which are mainly based on chromatographic techniques. Among these advantages, it can be cited rapidity, relative low cost, simple sample pretreatment, no need of organic solvents or reagents, and no production of chemical waste. The analytical and chemometric strategy employed in this study can be suggested to be applied to quantify phenylalanine or other fluorescent analytes in different food matrices that present matrix effect.

5. DETECTION OF ADULTERATIONS IN A VALUABLE BRAZILIAN HONEY BY USING SPECTROFLUORIMETRY AND MULTIWAY CLASSIFICATION

5.1 Introduction

Honey is a natural sweet substance produced by honey bees *Appis mellifera* from the nectar of plants or from secretions of their living parts. The secretions are collected by bees and deposited with their own specific substances in the honeycomb for maturation. Honey is mainly constituted by carbohydrates (70−80%, ww^{-1}) and water (10−20%, ww^{-1}) (*Codex Alimentarius* Commission, 2001). Minor components, such as organic acids, minerals, amino acids and phenolic compounds, correspond to less than 10% ww^{-1} and they vary with the floral origin, bee species, environmental and storage conditions (Patrignani *et al.,* 2018; Kaškoniene & Venskutonis, 2010). These substances are directly related to improving overall health and preventing some diseases due its antioxidant potential, antibacterial properties, and anti-inflammatory activities (Jibril *et al.,* 2019; Kavanagh *et al.*, 2019).

Honey can be classified according to origin as blossom or honeydew honey. Blossom honey is obtained from the nectar of flowers of blossoming plants, from one type of flower (monofloral) or by combing nectars from several types of flowers (polyfloral). Monofloral honeys have higher market prices than polyfloral honeys, due to their intrinsic physicochemical and sensory characteristics. Honeydew honey is a particular type of honey produced by bees from the secretions of plants or excretions of aphids and occasionally insects (*Codex Alimentarius* Commission, 2001). This type of honey most commonly originates from plants of the genera *Pinus, Abies, Castanea,* and *Quercus*, which are typically found in semi-arid climatic conditions. In Brazil, one of the most valuable honeydew honeys is the Bracatinga honey, produced from the plant species *Mimosa scabrella Bentham* and commonly found in the state of Santa Catarina, in the southern region of the country. Almost all the Brazilian high-priced honey production is exported (Bergamo *et al.,* 2019). Honeydew honey production is favored by climates with high temperatures and low humidity. In these environments, floral resources are limited and hence bees feed from the secretion instead of nectar as an adaptation mechanism. As a result, honeydew honey presents higher antibacterial and antioxidant activities when compared to blossom honeys. These characteristics have made honeydew and monofloral honeys potential targets for adulteration and fraud (Vasić *et al.,* 2020; Pita-Calvo & Vázquez, 2018; Azevedo *et al.,* 2017).

Taking this into account, the development of reliable analytical methods able to provide rapid and cost-effective results aiming to detect adulterations and frauds in honeydew and monofloral honeys become extremely relevant. Several types of adulteration are possible, such as direct incorporation of foreign substances into honey, addition of inexpensive honey types or indirect adulteration of honey through bee-feeding with industrial sugars (Geana & Ciucure, 2020; Oroian *et al.,* 2018; Pita-Calvo & Vázquez, 2018). Analytical methods employed in food adulteration/food authentication studies are often based on the identification and quantification of characteristic markers, such as some classes of compounds (phenolic acids, carbohydrates, volatile constituents) (Vasić *et al.,* 2020; Oroian *et al.,* 2018; Pita-Calvo & Vázquez, 2018). These compounds are typically present in minor concentrations (mg L^{-1}); hence more selective and sensitive quantification techniques are demanded. Chromatographic techniques (HPLC-MS, HPLC-UV, HPLC-fluorescence, GC-MS) involving sample preparation steps are the most applied methodologies (Geana & Ciucure, 2020; Pita-Calvo & Vázquez, 2018). As disadvantages, these techniques require complex sample pretreatments, laborious and time-consuming procedures, expensive analytical instrumentation, and highly trained professionals. In addition to these univariate methodologies, melissopalynological analysis is considered a standard method to check honey authentication regarding its geographical and botanical origin. This method consists of a microscopic analysis that requires the identification and counting of pollen grains and other particles in honey. Origin assignment is based on reference to the literature (Louveaux *et al.*, 1978). Despite being a simple method, it spends a lot of time and requires very specific skilled analysts. Other alternative that is laborious and requires skilled analysts for honey authentication is isotope ratio mass spectrometry (IRMS) (AOAC International, 1999).

Molecular spectroscopic techniques coupled with chemometric tools have been interesting alternatives to those traditional univariate procedures (Brereton *et al.,* 2018). Non-target spectroscopic fingerprinting approaches have dramatically reduced or even eliminated sample preparation steps, allowing the development of simpler, low-cost, green (not consuming solvents and generating minimum chemical waste), and rapid methods. On the other hand, chemometric tools are needed to extract maximum information and provide better interpretation of the large amounts of data that are usually acquired (Geana & Ciucure, 2020; Pita-Calvo & Vázquez, 2018; Strelec et al., 2018; Lenhardt et al., 2015; Christensen et al., 2006). Chemometric tools have also been combined with multi-element analysis by applying inductively coupled plasma

optical emission spectrometry (ICP- OES) to detect honey adulteration (Liu et al., 2021).

Multivariate supervised classification methods have been widely applied to the detection of food fraud and adulteration. In addition to the most common partial least squares discriminant analysis (PLS-DA) and soft independent modeling of class analogy (SIMCA), more sophisticated chemometric classification tools have also been applied, such as artificial neural networks (ANN), support vector machines (SVM) and unfold SIMCA (U-SIMCA) (Zeng et al., 2021; Ríos-Reina, Elcoroaristizabal et al., 2017; Azcarate et al., 2015). The most used tool has been PLS-DA (Brereton & Lloyd, 2014), which allows to classify a given sample as belonging or not to predefined classes, such as authentic or adulterated honey. Within this perspective, this chapter aimed to develop a classification model to discriminate Aroeira honey samples collected at different cities of the north of Minas Gerais State, Brazil, from samples adulterated with sugar cane molasses, corn syrup, and polyfloral honey by applying spectrofluorimetry and chemometrics. These have been the most common types of adulteration in honeys (Geana & Ciucure, 2020). The proposed method can also be used to investigate the botanical origin of Aroeira honey samples as a simple alternative to the exhaustive microscopic analysis. To the best of our knowledge, only a few papers have been published about Aroeira honey (Viana *et al.*, 2018; Bastos *et al.*, 2016) and none of them addressing the detection of possible adulteration.

Aroeira honey is a characteristic monofloral honey originated from a specific region in the north of Minas Gerais State, Brazil (Figure 5.1). It is produced from the secretions of *Astronium urundeuva* (*Anacardiaceae* - Aroeira) and is considered one of the most valuable honeys from Brazil. Aroeira honey has gained notoriety in the last years due its unique features, such as higher phenolic compounds concentration and antibacterial activity when compared to other types of honey (Viana *et al.*, 2018; Bastos *et al.*, 2016). Viana *et al.* (2018) have studied the antibacterial activity for ten Aroeira honey samples collected in the northern of Minas Gerais State. These samples were tested against pathogenic microorganisms *Staphylococcus aureus* TSST (clinical isolated) and enterohaemorrhagic *Escherichia coli* ATCC 43895 for the determination of their minimum inhibitory concentrations (MIC). The results evidenced the inhibition of the bacterial growth, pointing out to the potential of Aroeira honey.

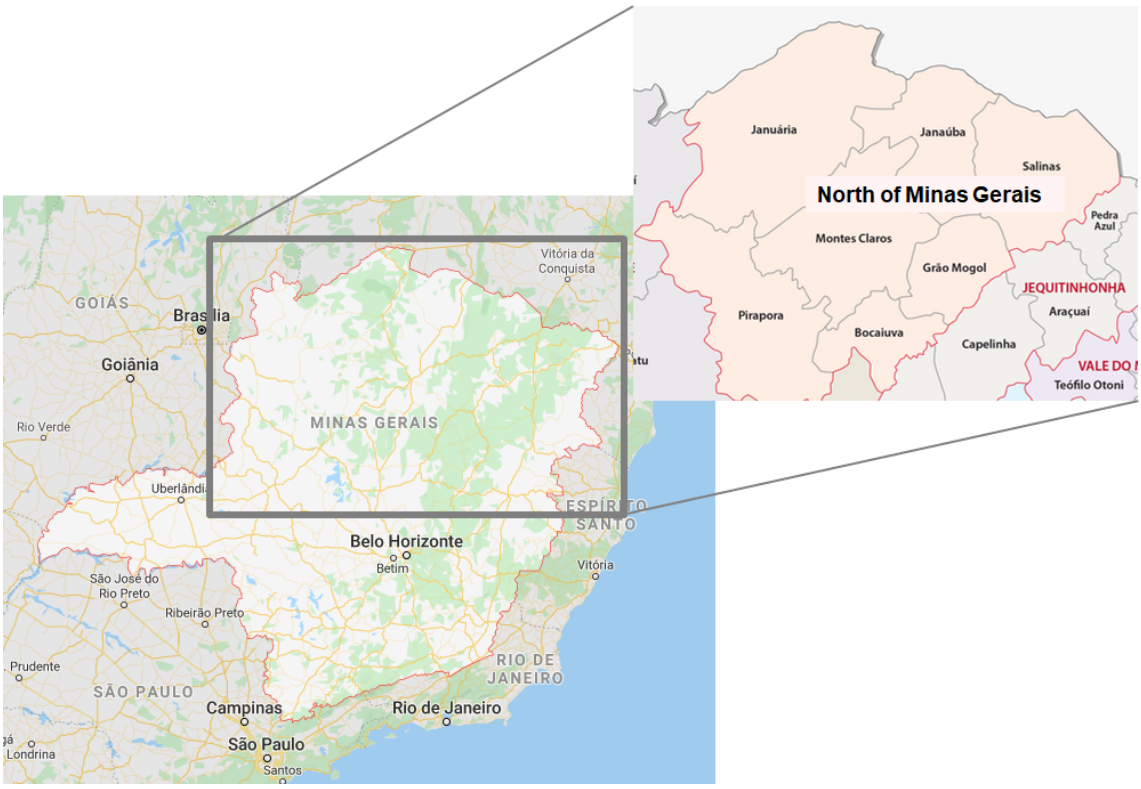

Figure 5.1. The map of Minas Gerais State, Brazil, with the (northern) region of production of Aroeira honey highlighted (Adapted from Jornal Montes Claros, 2019)

The north of Minas Gerais State was recognized in 2019 as official Aroeira honey production region. Due its particular characteristics, Aroeira honey presents about 100% higher market prices than polyfloral honeys and is currently in process to become a Protected Designation of Origin (PDO) product (IMA, 2019; Agência Minas, 2020). Consequently, Aroeira honey is a possible target for food fraud and adulteration. Since this PDO recognition is very recent, a few papers have been published about this food product so far, thus increasing the importance of new contributions.

5.2. Materials and methods

5.2.1. Sample preparation

Aroeira honey samples were collected from ten different selected official beekeepers aiming to ensure their representativeness and authenticity. Samples were placed into plastic bottles, sealed, and labeled with the city of origin and collection date. The flasks

were stored under refrigeration (4.0±0.5 °C) and kept in the dark until analysis. The select adulterants were sugar cane molasses, corn syrup and polyfloral honey, all from different origins. Adulterated samples were prepared gravimetrically at 5, 10, 20 and 30% w w^{-1} ratios of adulterant in aliquots of 2.0 g. They were prepared in 2 mL Eppendorf flasks, in which each adulterant was weighted at the planned ratio and the sample mass completed to 2.0 g with authentic Aroeira honey. Blends were homogenized with the help of a spatula and diluted with 40.00 mL of deionized water (Mili-Q, Merck, Darmstadt, Germany). The solutions were transferred into 50 mL polypropylene tubes, manually homogenized, and analyzed immediately by spectrofluorimetry. All samples were analyzed in triplicate. The averages of these triplicates were used to construct the dataset (models built with individual replicates showed similar performance, but the use of averages was considered a more representative scenario).

A total of 232 samples, 78 authentic and 154 adulterated, were prepared. Among the 154 non-authentic samples, 50 were adulterated with sugar cane molasses, 52 with corn syrup and 52 were mixed with polyfloral honey. Melissopalynological analyses (Louveaux et al., 1978) were performed as the reference method to check the authenticity of Aroeira honey samples. In summary, 10g of Aroeira honey were placed into a 50-mL polypropylene tube with 20.00 mL of deionized water (Mili-Q, Merck, Darmstadt, Germany), manually homogenized, and centrifuged at 3000 rpm per 7 minutes. The supernatant liquid was discarded and the decanted sediment was resuspended with 10.00 mL of deionized water. The solutions were homogenized and centrifuged following the initial conditions. The entire sediment was placed on a histological slide and the material was fixed at 35ºC. After drying, the slides were mounted with glycerin jelly at room temperature (25ºC) for 24 hours. Finally, the pollen grains contained in the honey samples were microscopically examined regarding number and morphology.

5.2.2. Apparatus and software

Spectrofluorimetric measurements were performed using a Varian Cary Eclipse fluorescence spectrophotometer (Walnut Creek, California, USA) equipped with two Czerny-Turner monochromators, and a Xenon discharge lamp pulsed at 80 Hz with a half peak height of 2 ls (peak power equivalent to 75 kW). A high-performance R298 photomultiplier tube detector was used for collection of the fluorescence spectra. The

temperature was hold at 25.00 ± 0.05 ºC using a Peltier thermostat. The spectrophotometer was interfaced with a computer containing th Cary-Eclipse 1.1 (132) software for spectral acquisition and exportation. The EEMs were obtained by varying the excitation wavelength (λ_{ex}) ranging between 250 and 510 nm (step 20 nm) and recording the emission spectra (λ_{em}) from 270 to 640 (step 1 nm). Excitation and emission slits were both set at 5 nm, and the scan rate was 600 nm min^{-1}.

Rayleigh scatter emission was removed from all the spectra by selecting working sensor regions in the EEMs before building the models. The characteristic region of elastic scattering provides non-bilinear analytical signals that are unrelated to the chemical composition of the samples. Therefore, this region should be removed when constructing multiway models, otherwise it would introduce nonlinearities causing lack of trilinearity (Bro, 1997). All the models were built using Matlab software, version 7.10.0 (R2010a) (MathWorks, Natick, USA), jointly with the PLS Toolbox package, version 6.71 (Eigenvector Technologies, Manson, USA).

5.3. Results and discussion

5.3.1. Fluorescence spectra

Aiming to evaluate the optimal instrumental conditions, three increments/steps of excitation wavelength were tested: 10, 15 and 20 nm (Figure 5.2). As it was expected, a trend of lower resolution in the contour plots was observed directly proportional to the size of the excitation increments. Nevertheless, the run time was about twice longer for 10 nm step than for 20 nm step. In addition, spectral regions of higher intensity seemed not to be significantly changed. Thus, excitation increments of 20 nm were chosen, resulting in a run time of about 9 minutes per measurement.

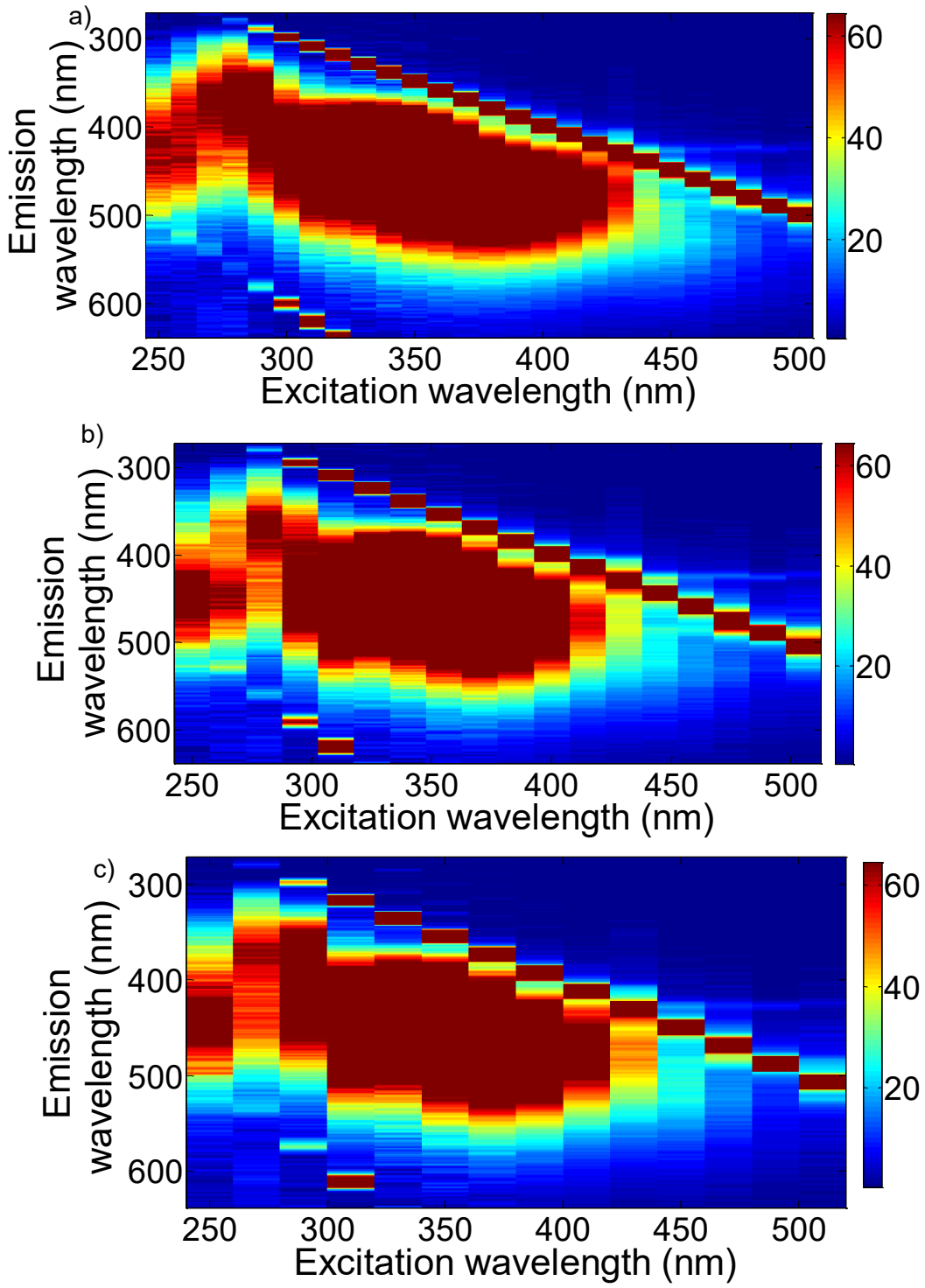

Figure 5.2. Fluorescence contour maps of one pure Aroeira honey sample for three different tested excitation wavelength increments: **(a)** 10 nm, **(b)** 15 nm, and **(c)** 20 nm

Figure 5.3 shows contour maps for one pure Aroeira honey samples before (a) and after (b) the scattering removal. Spectral contour maps of the honey samples are visually similar in general, as can be seen in Figure 5.4.

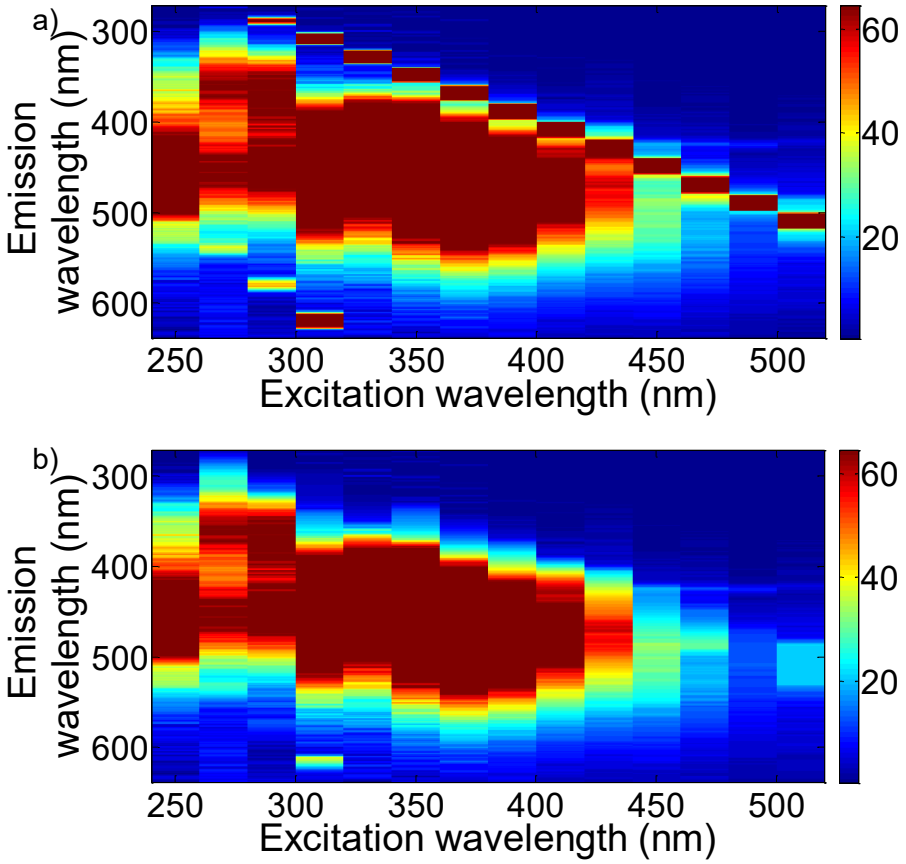

Figure 5.3. Fluorescence contour maps of one pure Aroeira honey sample **(a)** before and **(b)** after scattering removal

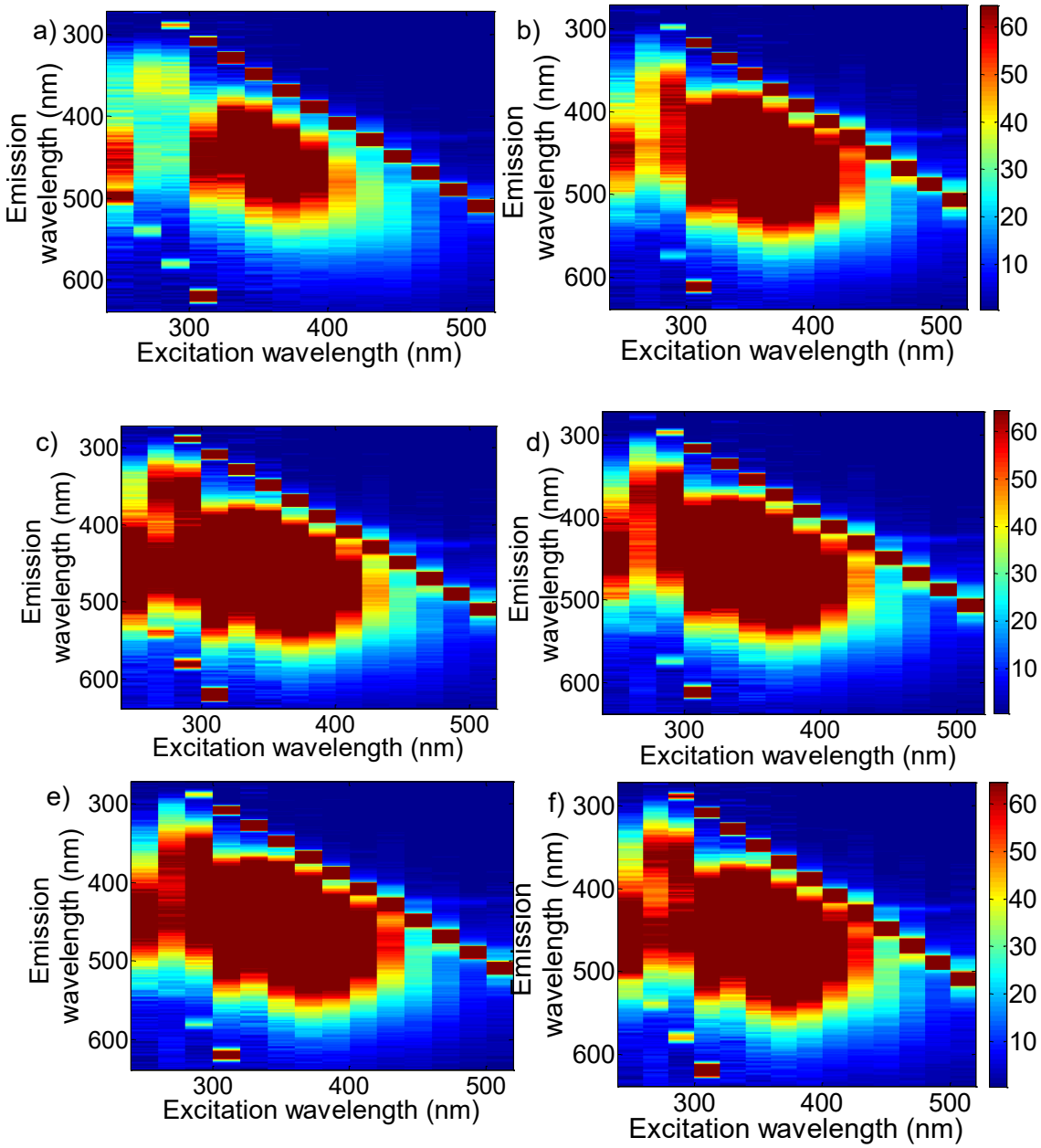

Figure 5.4. Fluorescence contour maps of six different Aroeira honey samples

Figures 5.5-4.7 show the fluorescence contour maps of some of the adulterated honey samples. Nonetheless, fluorescence contour maps for samples adulterated at 20% and 30% w w^{-1} with corn syrup and sugar cane molasses (Figures 5.6c-d and 5.7c-d) can be visually discriminated from those containing 5% and 10% w w^{-1} of these same adulterants.

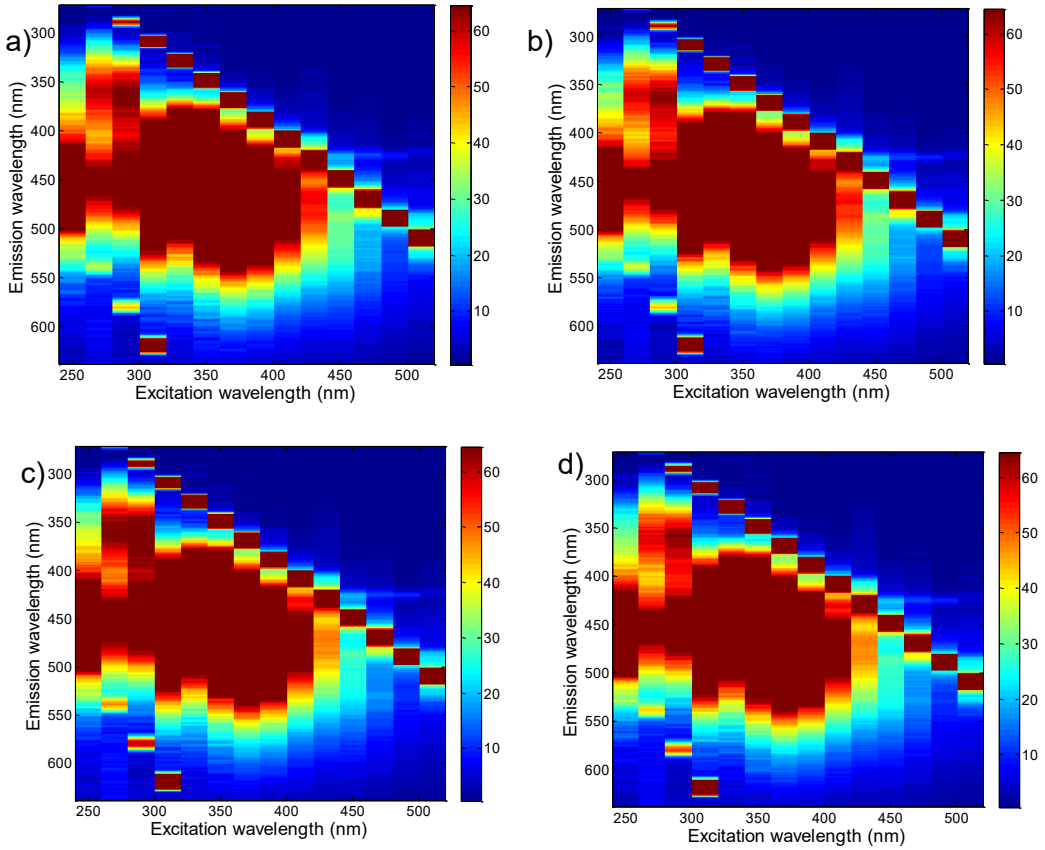

Figure 5.5. Fluorescence contour maps for an Aroeira honey sample adulterated with **(a)** 5% ww^{-1}, **(b)** 10% ww^{-1}, **(c)** 20% ww^{-1}, and **(d)** 30% ww^{-1} of polyfloral honey

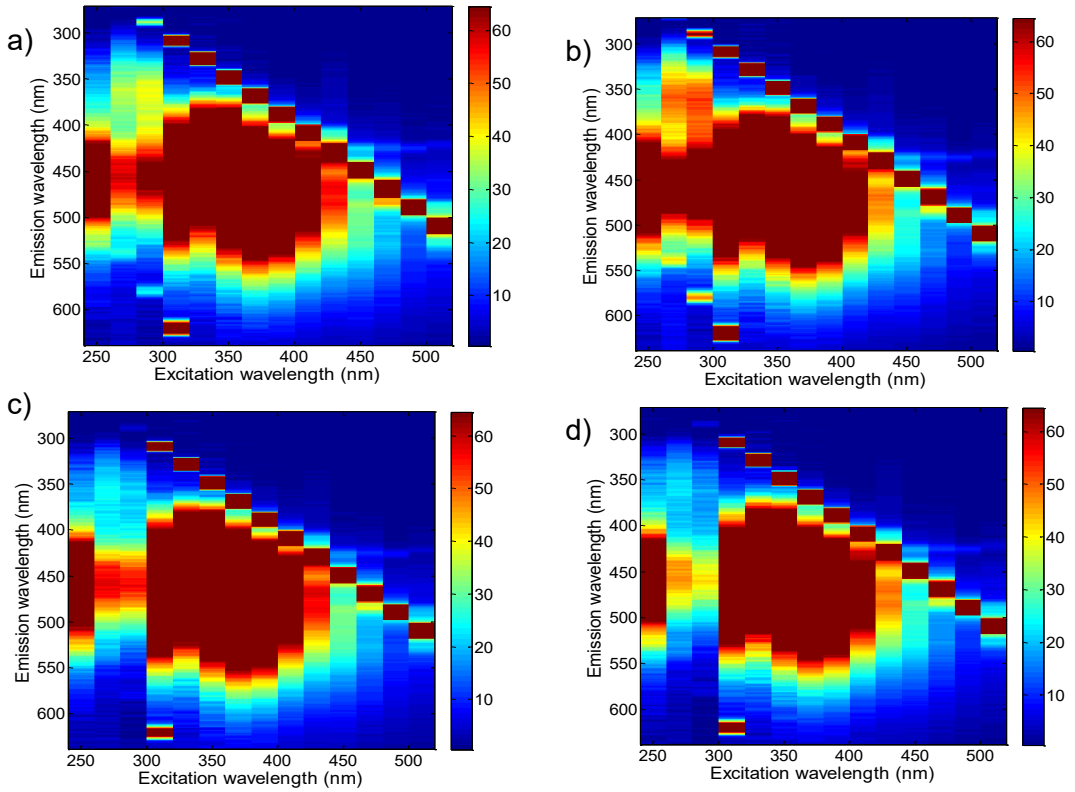

Figure 5.6. Fluorescence contour maps for an Aroeira honey sample adulterated with **(a)** 5% ww^{-1}, **(b)** 10% ww^{-1}, **(c)** 20% ww^{-1}, and **(d)** 30% ww^{-1} of corn syrup

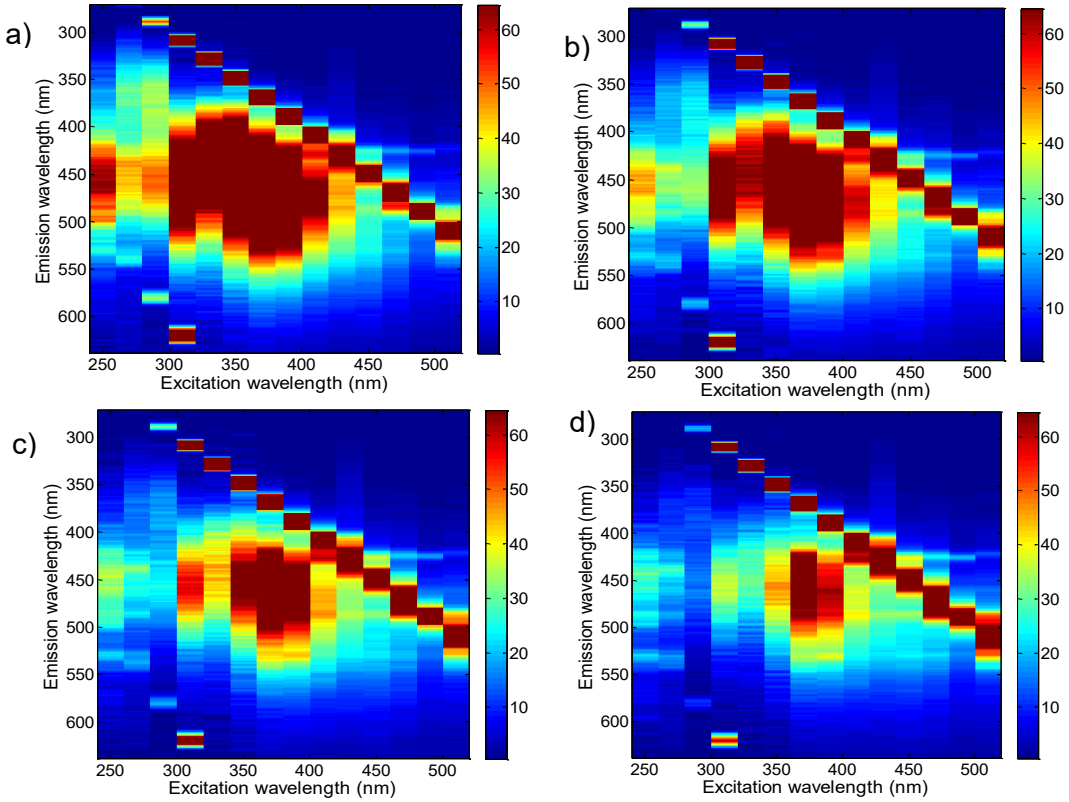

Figure 5.7. Fluorescence contour maps for an Aroeira honey sample adulterated with **(a)** 5% ww^{-1}, **(b)** 10% ww^{-1}, **(c)** 20% ww^{-1}, and **(d)** 30% ww^{-1} of sugar cane molasses

Therefore, only the two lower levels for these two adulterants were included in chemometric models, while for adulteration with polyfloral honey mixtures all the four levels were maintained since no visual discrimination was evident (Figure 5.5). Thus, these low adulterant contents were chosen to reproduce spectra that cannot be visually discriminated. It is interesting to note that some recent articles have detected the presence of only one individual adulterant at a time in honey, typically sugars, at higher levels (above 10% w w^{-1}) by employing vibrational analytical techniques, such as NIR, Fourier transform infrared (FTIR) and terahertz spectroscopies (Liu *et al.*, 2020; Shiddiq *et al.*, 2019; Ferreiro-González *et al.*, 2018; Guelpa *et al.*, 2017). It should be noted that spectrofluorimetry provides higher sensitivity than vibrational spectroscopic techniques.

5.3.2. PARAFAC exploratory modeling

A PARAFAC model was built only with the authentic samples (78*x*357*x*14) as an exploratory tool aiming to investigate the fluorescence composition of the Aroeira honey samples. The number of factors was estimated based on model residuals and the core consistency diagnostic (CORCONDIA). This last tool evaluates the trilinear consistency of the data. In general, CORCONDIA values above 50% indicate a good trilinear structure (Bro & Kiers, 2003). The first exploratory model was optimized with four factors that accounted for 99.43% of the total variance. This model was built imposing non-negativity constraint to the three modes and presented a CORCONDIA of 62%. The loadings of this model are shown in Figure 5.8. Loadings of the first mode (Figure 5.8a) correspond to the relative concentration patterns present in Aroeira honeys, whereas loadings of the second and third modes (Figure 5.8b-c) correspond to the excitation and emission spectral profiles recovered by the model, respectively.

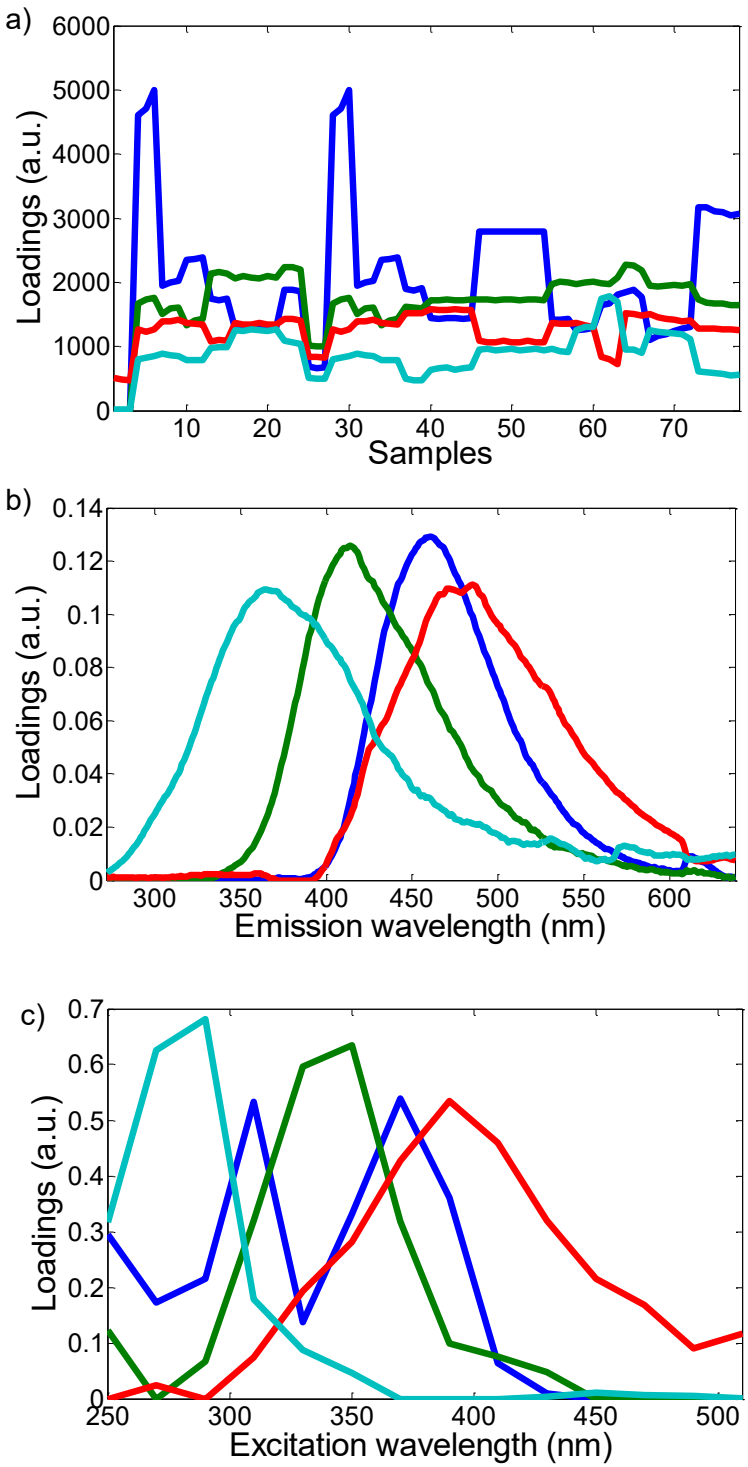

Figure 5.8. Loadings of an exploratory three-way PARAFAC model built for pure Aroeira honey samples with four components. Loadings of **(a)** first/sample mode, **(b)** second/excitation mode, and **(c)** third/emission mode

Four excitation and emission spectral profiles were modeled. They presented excitation/emission wavelengths maxima at 290/366 nm (light blue loadings, Figure 4.8), 350/414 nm (green), 370/460 nm (blue), and 390/480 nm (red). The presence of these fluorescence spectral profiles in honeys is in accordance with other studies found

in the literature (Stankovic *et al.*, 2019; Lenhardt *et al.,* 2015; Christensen *et al.*, 2006). In general, these spectral patterns have been assigned to proteins, amino acids, and phenolic compounds. Phenolic compounds show maximum excitation at 265-335 nm, and maximum emission at 358-426 nm (Rodríguez-Delgado *et al.*, 2001). These compounds can be found in different concentrations, according to the type of honey. Aroeira honey has a high content of phenolic compounds, which are supplied by flowers to bees for honey production (Bastos et al., 2016; Viana et al., 2018). Therefore, the first factor could be associated to phenolic compounds present in honey. Since several phenolic compounds are found in honey, it is difficult to assign this PARAFAC factor to specific compounds (Kavanagh *et al.*, 2019; Lenhardt *et al.,* 2015).

The second (green) and third (blue) factors could be ascribed to Maillard reaction products such as furosine and hydroxymethylfurfural (HMF), which present maxima excitation at 360 nm, and emission at 425 and 440 nm (Christensen *et al.,* 2006). There are no data in the scientific literature about the Maillard reaction products content in Aroeira honey. However, HMF content is considered an important parameter for the evaluation of the quality of honey in general. This compound can be formed when honey is submitted to heat treatment or during its storage period, due to dehydration of sugars in acidic environment (Castro-Vázquez et al., 2007). The fourth PARAFAC factor (red), presenting maximum emission at 480 nm, may be assigned to vitamin A (retinol), which presents an excitation peak around 350 nm (Christensen *et al.*, 2006). Honey is a complex food matrix that contains small amounts of proteins, peptides, and free amino acids, including tryptophan (Ruoff et al., 2005). The shape and the maxima excitation/emission of the fluorescence spectra from each component may vary according to the type of honey and its physico-chemical properties, such as pH, moisture content, viscosity, etc. (Lenhardt *et al.*, 2015; Karoui *et al.*, 2007; Christensen *et al.*, 2006).

5.3.3. PLS-DA and discriminant PARAFAC models

Linear discriminant analysis (LDA) models were built from the scores of a PARAFAC model constructed with the whole dataset. This strategy is known as PARAFAC-LDA, coupling the multiway decomposition method with discriminant analysis (Silva *et al.*, 2016). The performance of the discriminant PARAFAC models was considered unsatisfactory, with misclassification rates above 50%.

Binary PLS-DA models were built to discriminate between authentic and adulterated honey samples, coded as 0 and 1, respectively. Four different two-way PLS-DA models were built with the emission spectra related to excitation wavelengths at 290, 350, 370 and 390 nm. These emission spectra correspond to the slices of the three-way dataset related to the most intense excitation fluorescence signals as observed for each one of the excitation loadings of the four PARAFAC factors (Figure 5.9b). Binary PLS-DA models were constructed without distinguishing the three types of adulterants. For these models, about two-thirds of the samples, 152 (52 pure Aroeira honey samples and 100 adulterated honey samples), were selected for the training set by employing the Kennard–Stone algorithm (Kennard & Stone, 1969). The remaining 80 samples (26 pure Aroeira honey samples and 54 adulterated honey samples) were used for the test set, aiming to independently validate the model. Dummy (y) variables 1.0 and 0.0 were assigned to adulterated and authentic honey samples, respectively.

Data were mean centered, and the number of LV was chosen by venetian blinds cross validation (10 splits) based on the lowest CVCE. In general, the results were poor with a high rate of misclassifications (Table 5.1). The best model was obtained with emission spectra corresponding to excitation at 370 nm. This PLS-DA model was built with four LV that accounted for 99.8% of variance in the **X** block, and 21.4% in the **Y** block. The plot of individual y predicted values is shown in Figure 5.9a.

Table 5.1. Classification results of PLS-DA models for training and test sets according to the select excitation wavelength

	Excitation wavelength							
	290 nm		350 nm		370 nm		390 nm	
	Training	Test	Training	Test	Training	Test	Training	Test
FPR[a]	46%	58%	69%	54%	48%	58%	19%	19%
FNR[b]	21%	6%	16%	11%	8%	4%	45%	44%
Sensitivity	79%	94%	84%	89%	92%	96%	55%	56%
Specificity	54%	42%	31%	46%	52%	42%	81%	81%
Precision	57%	79%	50%	67%	77%	85%	48%	47%
NMC[c]	67%	63%	85%	65%	56%	61%	64%	64%

[a] False positive rate. [b] False negative rate. [c] Number of misclassifications.

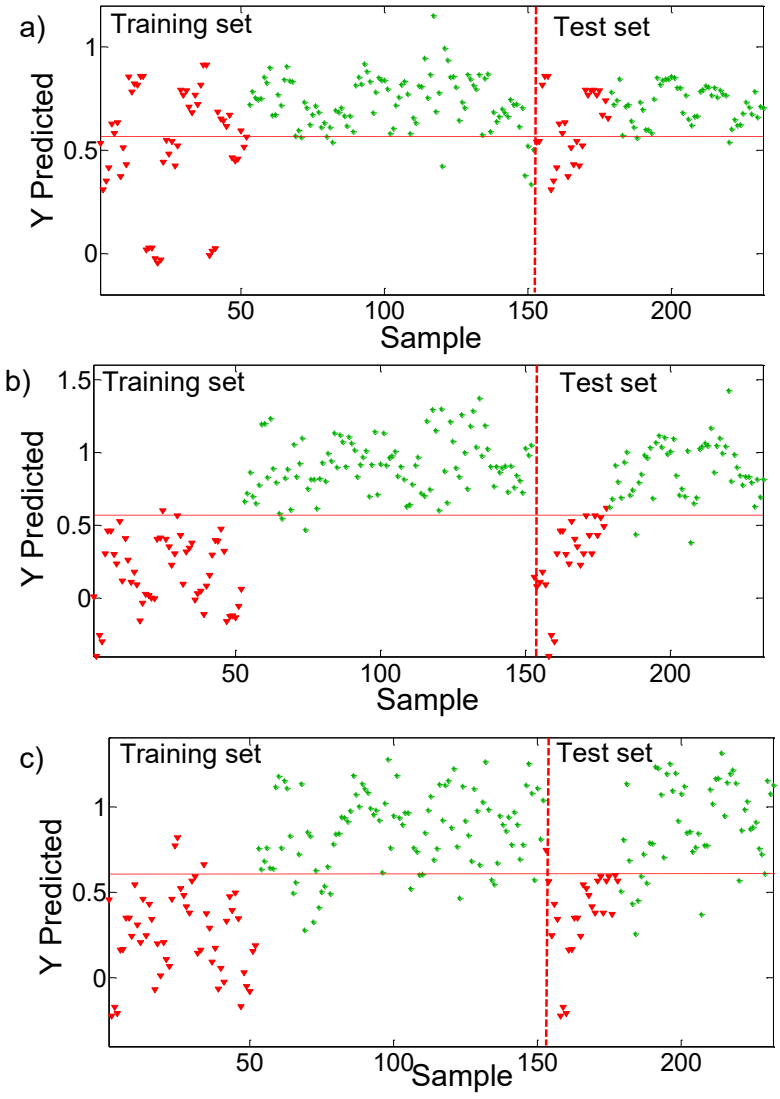

Figure 5.9. Y predicted values for **(a)** PLS-DA (built with emission spectra obtained at 370 nm), **(b)** UPLS-DA, and **(c)** NPLS-DA models. The stars (in green) represent the adulterated samples and the down triangles (in red) represent the authentic samples. Horizontal dashed lines represent Bayesian thresholds and vertical dashed lines indicate the separation between training and test samples

Even the best PLS-DA model presented poor performance, with a false positive rate (FPR) of 58% for the test set, i. e., more than half of authentic honey samples were predicted as adulterated. This poor performance of the PLS-DA models may be attributed to a lack of modeled information in relation to the whole three-way dataset. It is then assumed that the all the information about fluorescence molecular composition of the samples is distributed through the whole EEM, and thus multiway classification

models are needed to obtain good prediction rates. Models for multiway data will be discussed in the next two sections.

5.3.4. Unfold PLS-DA model

For constructing UPLS-DA models, original three-way data arrays should be unfolded into two-way matrices by combining the spectral modes. The original training and test datasets had dimensions 154*x*367*x*14 and 78*x*367*x*14, and were respectively unfolded into 154*x*5138 and 78*x*5138 matrices. These matrices were mean centered, and the best model was obtained with 11 LV, accounting for 99.7% of the variation in the **X** block and 74.4% in the **Y** block.

Figure 5.9b shows the y predicted values for training and test samples. As can be seen, this model had a much better performance than the best PLS-DA model, what can be explained by the use of the whole spectral information in the model. Misclassifications were smaller than 4% in general, and FPR and false negative rate (FNR) were both equal to 2% for the training set and 4% for the test set (Table 5.2). These results mean that only one sample of pure Aroeira honey and two adulterated samples were misclassified in each training and test sets. Therefore, only six out of 232 samples were not correctly predicted.

Table 5.2. Classification results of UPLS-DA and NPLS-DA models for training and test sets

	UPLS-DA		**NPLS-DA**	
	Training	Test	Training	Test
FPR[a]	2%	4%	8%	8%
FNR[b]	2%	4%	13%	9%
Sensitivity	98%	96%	87%	91%
Specificity	98%	96%	92%	92%
Precision	96%	93%	79%	83%
NMC[c]	4%	8%	21%	17%

[a] False positive rate. [b] False negative rate. [c] Number of misclassifications.

The spectral interpretation is an important aspect of any chemometric model. For this aim, variable importance in projection (VIP) scores should be evaluated aiming to detect the most discriminant variables. VIP scores are an important informative vector in classification and calibration models, but in the case of multiway models we have a matrix of VIP scores. Figure 5.10 shows the refolded VIP scores for the UPLS-DA model. It is possible to note the most intense four peaks at excitation/emission wavelengths 310/440 nm, 310/610 nm, 350/421 nm, and 370/481 nm. These last two spectral bands correspond to the second and third PARAFAC factors, and as already discussed in section 4.3.2, they may be related to Maillard reaction products (Christensen et al., 2006). The peak at 310/440 nm may be assigned to compounds derived from caramel dye, a substance commonly found in corn syrup that could be considered a marker of honey fraud in this case (Ma *et al.*, 2019).

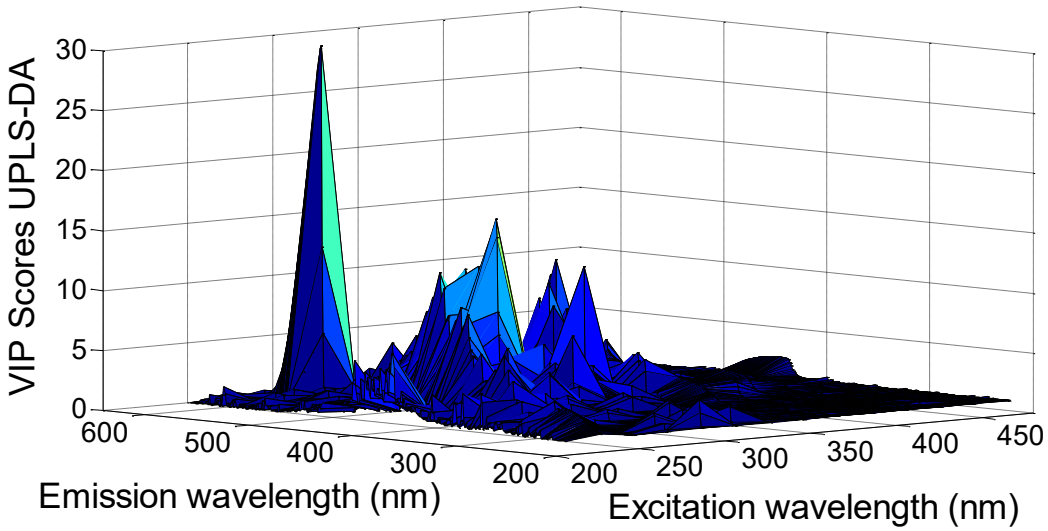

Figure 5.10. Refolded variable importance in projection (VIP) scores for the best discriminant model, UPLS-DA

5.3.5. NPLS-DA

The selected NPLS-DA model was built using 10 LV, which accounted 99.68% and 60.93% of variance in the **X** and **Y** blocks, respectively. Figure 5.9c shows the y predicted values for this model. The performance of this model was much better than PLS-DA models, but worse than UPLS-DA (Table 5.2). FPR and FNR for the test set were 8% and 9%, respectively. This represents approximately twice the rate of

misclassification obtained for the UPLS-DA model. The more feasible results observed for the UPLS-DA model may be explained by its higher flexibility in comparison with NPLS-DA for three-way data modeling (Bro, 1996). The higher flexibility of UPLS-DA is due the larger number of degrees of freedom used in this model, a consequence of the unfolding strategy, which makes it flexible enough not to discard relevant spectral patterns, though also makes it more sensitive to noise than NPLS-DA (Bro, 1996, 1998).

5.3.6. Figures of merit

The analytical validation of the models was performed based on six FOM: FPR, FNR, SEN, SPE, precision and number of misclassifications (NMC). FPR and FNR are the basic FOM for any classification model, while SEN and SPE are their complementary parameters. SEN is the rate of true positive (TP) predictions, thus complementary to FNR, while SPE is the rate of true negative (TN) predictions, complementary to FPR (Botelho *et al.*, 2015).

Precision is a FOM related to FPR, which expresses the percentage of correctly classified positive/adulterated samples among all samples classified as positive.

$$Precision = \frac{TPR}{TPR + FPR} \quad (5.2)$$

Thus, this is a more robust FOM to evaluate the detection of adulterated samples. The NMC is a FOM more appropriate to globally evaluate the models and is calculated as the sum of FP and FN predictions (Botelho *et al.*, 2017).

$$NMC = FPR + FNR \quad (5.3)$$

Tables 5.1-5.2 allow the comparison between models through these six FOM. By observing these tables, it can be clearly noted the best FOM for the UPLS-DA model. This model presented precision of 96% and 93% for the training and test set, respectively. For an overall evaluation of this model, independent of positive or negative misclassifications, the most representative FOM is NMC, which was equal to 4% and 8% for the training and test set, respectively.

5.4. Conclusions

This study developed a simple, direct, environmentally friendly, and low-cost analytical method to detect simultaneously three adulterants in Aroeira honeys, based on chemometrics and spectrofluorimetry. The poor performance of bilinear classification models (PLS-DA) built with emission spectra obtained at fixed excitation wavelengths demonstrated the limitations of this strategy and the need to use all the spectral information obtained from fluorescence EEM. Thus, more sophisticated multiway tools (NPLS-DA and UPLS-DA) were necessary to model all the spectral information provided by this analytical technique. The highest classification ability was obtained with UPLS-DA, presenting NMC between 4-8%, which represents only six out of 232 misclassifications. This developed analytical protocol was the first study to apply the above-mentioned multilinear discriminant models to detect adulterations in honey.

An exploratory PARAFAC model and the spectral interpretation of the best (UPLS-DA) model through its VIP scores matrix allowed to suggest possible discriminant substances, such as phenolic compounds, Maillard reaction products, vitamins, and caramel dye derivatives. Thus, the developed analytical methodology can be considered a feasible alternative to the traditional methods in the detection of frauds in honeys.

6. GENERAL CONCLUSIONS

The proposed applications attested the high potentiality of chemometric multiway tools in food analysis. Despite being proposed in the last century, high-order chemometric methods are, in general, still not popularized and not explored to their full potential. In this context, PARAFAC was applied as a data processing method for sensory analysis of coffee capsules and for the quality control of honey samples. High-order classification methods, N-PLSDA and U-PLSDA, were also used in this last application.

One of the main features of PARAFAC, second-order advantage, was explored to quantify phenylalanine in polyfloral honey samples. The combination of EEM data and this advantage allowed to successfully replace an analytical separation method by an analytical strategy based on “mathematical chromatography”.

N-PLSDA and U-PLSDA were used to discriminate adulterated honey samples. This application highlighted the importance of the suitability of the nature and structure of the dataset in relation to the choice of the chemometric method. This suitability is extremely important to avoid erroneous conclusions.

Therefore, the chemometric tools utilized in this thesis represented a feasible, comprehensive and simple alternative to processing multilinear data. Hereafter, more work should be performed to the popularization and validation of multiway chemometric approaches as alternatives to those traditional methods. For instance, employing these methods to explore the second-order advantage for various analytes in matrices of different natures, e.g., pharmaceutical, environmental, forensic and food samples, showing the versatility of these chemometric tools.

REFERENCES

Associação Brasileira da Indústria de Café (ABIC) (2020) Available in: <https://www.abic.com.br/wp-content/uploads/2020/01/Euromonitor_Coffe-Market-Trends-in-Brazil_-Encafe.pdf>. Rio de Janeiro, RJ, Brazil. Accessed in January 2021.

Agência Minas (2020) Available in: <https://www.diariodoaco.com.br/noticia/0075471-mel-de-aroeira-produzido-no-norte-de-minas-a-alvo-de-pesquisadores-e-do-mercado>.Ipatinga, MG, Brazil. Accessed in April 2020.

Allenspach, M.D., Fuchs, J.A., Doriot, N., Hiss, J.A., Schneider, G. & Steuer, C. (2018) **Quantification of hydrolyzed peptides and proteins by amino acid fluorescence.** Journal of Peptide Science, vol. 24, e3113.

AOAC International (1999) **C-4 Plant Sugars in Honey. Official Methods of Analysis,** 16th ed., 5th rev., Gaithersburg, Method 998.12.

Azcarate, S. M., Gomes, A. A., Alcaraz, M. R., Araújo, M. C. U., Camiña, J. M. & Goicoechea, H. C. (2015) **Modeling excitation-emission fluorescence matrices with pattern recognition algorithms for classification of Argentine white wines according grape variety.** Food Chemistry, vol. 184, pp. 214-219.

Azevedo, M.S., Seraglio, S.K.T., Rocha, G., Balderas, C.B., Piovezan, M., Gonzaga, L.V., Falkenberg, D.D.B., Fett, R., de Oliveira, M.A.L. & Costa, A.C.O. (2017) **Free amino acid determination by GC-MS combined with a chemometric approach for geographical classification of bracatinga honeydew honey (*Mimosa scabrella Bentham*).** Food Control, vol. 78, pp. 383-392.

Bader, M. (1980) **A systematic approach to standard addition methods in instrumental analysis.** Journal of Chemical Education, vol. 57, pp. 703-706.

Barahona, I., Jaimes, E.M.S. & Yang, J.-B. (2020) **Sensory attributes of coffee beverages and their relation to price and package information: A case study of Colombian customers' preferences**. Food Science and Nutrition, vol. 8, pp. 1173-1186.

Bastos, E.M.A.F., Calaça, P.S.S.T., Simeão, C.M.G. & da Cunha, M. R. (2016) **Characterization of the honey from Myracrodruon urundeuva (*Anacardiceae* - Aroeira) in the Dry Forest of northern of Minas Gerais/Brazil.** Advances in Agricultural Science, vol. 4, pp. 64-71.

Belchior, V., Botelho, B.G., Oliveira, L.S. & Franca, A.S. (2019) **Attenuated Total Reflectance Fourier Transform Spectroscopy (ATR-FTIR) and**

chemometrics for discrimination of espresso coffees with different sensory characteristics. Food Chemistry, vol. 273, pp. 178-185.

Bergamo, G., Seraglio, S.K.T., Gonzaga, L.V., Fett, R. & Costa, A.C.O. (2019) **Physicochemical characteristics of bracatinga honeydew honey and blossom honey produced in the state of Santa Catarina: An approach to honey differentiation.** Food Research International, vol. 116, pp. 745-754.

Bernardes, C.D., Poppi, R.J. & Sena, M.M. (2010) **Direct determination of trans-resveratrol in human plasma by spectrofluorimetry and second-order standard addition.** Talanta, vol. 82, pp.640-645.

Bian, X., Li, S., Lin, L., Tan, X., Fan, Q. & Li, M. (2016) **High and low frequency unfolded partial least squares regression based on empirical mode decomposition for quantitative analysis of fuel oil samples.** Analytica Chimica Acta, vol. 925, pp. 16-22.

Biluca, F.C., Bernal, J., Valverde, S., Ares, A.M., Gonzaga, L.V., Costa, A.C. & Fett, R. (2019) **Determination of free amino acids in stingless bee (*Meliponinae*) honey.** Food Analytical Methods, vol. 12, pp.902–907.

Bong J., Loomes, K.M., Schlothauer, R.C. & Stephens, J.M. (2016) **Fluorescence markers in some New Zealand honeys.** Food Chemistry, vol. 192, pp. 1006-1014.

Booksh, K., Henshaw, J.M., Burgess, L.W. & Kowalski, B.R. (1995) **A second-order standard addition method with application to calibration of a kinetics–spectroscopic sensor for quantitation of trichloroethylene.** Journal of Chemometrics, vol. 9, pp. 263-282.

Booksh, K.S. & Kowalski, B.R. (1994) **Theory of Analytical Chemistry.** Analytical Chemistry, vol. 66, pp. 782-791.

Botelho, B.G., Oliveira, L.S. & Franca, A.S. (2017) **Fluorescence spectroscopy as tool for the geographical discrimination of coffees produced in different regions of Minas Gerais State in Brazil.** Food Control, vol. 77, pp. 25-31.

Botelho, B.G., Reis, N., Oliveira, L.S. & Sena, M.M. (2015) **Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA.** Food Chemistry, vol. 181, pp. 31-37.

Brereton, R.G. (2003) **Chemometrics: Data Analysis for the Laboratory and Chemical Plant.** University of Bristol, Wiley, Chichester, 497 p.

Brereton, R.G. (2015) **Pattern recognition in chemometrics.** Chemometrics and Intelligent Laboratory Systems, vol. 149, pp. 90-96.

Brereton, R.G. & Lloyd, G.R. (2014) **Partial least squares discriminant analysis: Taking the magic away.** Journal of Chemometrics, vol. 28, pp. 213-225.

Brereton, R.G., Jansen, J., Lopes, J., Marini, F., Pomerantsev, A., Rodionova, O., Roger, J.M., Walczak, B. & Tauler, R. (2017) **Chemometrics in analytical chemistry—part I: history, experimental design and data analysis tools.** Analytical and Bioanalytical Chemistry, vol. 409, pp. 5891-5899.

Brereton, R.G., Jansen, J., Lopes, J., Marini, F., Pomerantsev, A., Rodionova, O., Roger, J.M., Walczak, B. & Tauler, R. (2018) **Chemometrics in analytical chemistry—part II: modeling, validation, and applications.** Analytical and Bioanalytical Chemistry. vol. 410, pp. 6691-6704.

Bressanello, D., Liberto, E., Cordero, C., Rubiolo, P., Pellegrino, G., Ruosi, M.R. & Bicchi, C. (2017) **Coffee aroma: Chemometric comparison of the chemical information provided by three different samplings combined with GC–MS to describe the sensory properties in cup.** Food Chemistry, vol. 214, pp. 218-226.

Bro, R. & Smilde, A.K. (2014) **Principal component analysis.** Analytical Methods, vol. 6, pp. 2812-2831.

Bro, R., Qannari, E.M., Kiers, H.A.L., Næsd, T. & Frøsta, M.B. (2008) **Multi-way models for sensory profiling data.** Journal of Chemometrics, vol. 22, pp. 36-45.

Bro, R. & Kiers, H.A.L. (2003) **A new efficient method for determining the number of components in PARAFAC models.** Journal of Chemometrics, vol. 17, pp. 274-286.

Bro, R. (1996) **Multiway calibration. Multilinear PLS.** Journal of Chemometrics, vol. 10, pp. 47-61.

Bro, R. (1997) **PARAFAC. Tutorial and applications.** Chemometrics and Intelligent Laboratory Systems, vol. 38, pp.149-171.

Bro, R. (1998) **Multi-way analysis in the food industry: models, algorithms, and applications.** PhD Thesis, Universiteit van Amsterdam, Amsterdam.

Bu, J., Dai, Z., Zhou, T., Lu, Y. & Jiang, Q. (2013) **Chemical composition and flavour characteristics of a liquid extract occurring as waste in crab (*Ovalipes punctatus*) processing.** Journal of the Science of Food and Agriculture, vol. 93, pp. 2267-2275.

Carabajal, M.D., Arancibia, J.A. & Escandar, G.M. (2019) **Excitation-emission fluorescence-kinetic third-order/four-way data: Determination of bisphenol A and nonylphenol in food-contact plastics.** Talanta, vol. 197, pp. 348-355.

Carroll, J.D. & Chang, J.-J. (1970) **Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition.** Psychometrika, vol. 35, pp. 283-319.

Chen, H., Jin, L., Chang, Q., Peng, T., Hu, X., Fan, C., Pang, G., Lu, M. & Wang, W. (2016) **Discrimination of botanical origins for Chinese honey according to free amino acids content by high-performance liquid chromatography with fluorescence detection with chemometric approaches.** Journal of the Science of Food and Agriculture, vol. 97, pp. 2042-2049.

Chen, Q., Qi, S., Li, H., Han, X., Ouyang, Q. & Zhao, J. (2014) **Determination of rice syrup adulterant concentration in honey using three-dimensional fluorescence spectra and multivariate calibrations.** Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 131, pp. 177-182.

Christensen, J., Nørgaard, L., Bro, R. & Engelsen, S.B. (2006) **Multivariate autofluorescence of intact food systems.** Chemical Reviews, vol. 106, pp. 1979-1994.

Chunjian, Y., Liu, R., Li, X., Song, Y. & Gao, H. (2021) **Degradation of dissolved organic matter in effluent of municipal wastewater plant by a combined tidal and subsurface flow constructed wetland.** Journal of Environmental Sciences, vol. 106, pp. 171-181.

Cocchi, M., Bro, R., Durante, C., Manzini, D., Marchetti, A., Saccani, F., Sighinolfi, S. & Ulrici, A. (2006) **Analysis of sensory data of Aceto Balsamico Tradizionale di Modena (ABTM) of different ageing by application of PARAFAC models.** Food Quality and Preference, vol. 17, pp. 419-428.

Codex Alimentarius Commission (2001) Revised codex standard for honey. Codex Committee Meeting Report (Alinorm), pp. 19–26.

Cordella, C.B.Y., Leardi, R. & Rutledge, D.N. (2011) **Three-way principal component analysis applied to noodles sensory data analysis.** Chemometrics and Intelligent Laboratory Systems, vol. 106, pp. 125-130.

Cortés-Diéguez, S., Otero-Cerviño, C., Rodeiro-Mougán, H. & Feijóo-Mateo, J.A. (2020) **Quantitative descriptive analysis of traditional herbal and coffee liqueurs made with grape marc spirit (Orujo).** Foods, vol. 9, 753.

Curi, P.N., de Almeida, A.B., Tavares, B.S., Nunes, C.A., Pio, R., Pasqual, M. & de Souza, V.R. (2017) **Optimization of tropical fruit juice based on sensory and nutritional characteristics.** Food Science and Technology, vol. 37, pp. 308-314.

Dahl, T., Tomic, O., Wold, J.P. & Næs, T. (2008) **Some new tools for visualising multi-way sensory data.** Food Quality and Preference, vol. 19, pp. 103-113.

David, H. A., Hartley, H. O. & Pearson, E. S. (1954) **The Distribution of the Ratio, in a Single Normal Sample, of Range to Standard Deviation.** Biometrika, vol. 41, pp. 482-493.

de Juan, A., Jaumot, J. & Tauler, R. (2014) **Multivariate Curve Resolution (MCR). Solving the mixture analysis problem.** Analytical Methods, vol. 6, pp. 4964-4976.

de Juan, A., Vander Heyden, Y., Tauler, R. & Massart, D.L. (1997) **Assessment of new constraints applied to the alternating least squares method.** Analytica Chimica Acta, vol. 346, pp. 307-318.

Eiermann, A., Smrke, S., Guélat, L.-M., Wellinger, M., Rahn, A. & Yeretzian, C. (2020) **Extraction of single serve coffee capsules: linking properties of ground coffee to extraction dynamics and cup quality.** Scientific Reports, vol. 10, 17079.

Fang, H., Wu, H.-L., Wang, T., Long, W.-J., Chen, A.-Q., Ding, Y.-J. & Yu, R.-Q. (2021) **Excitation-emission matrix fluorescence spectroscopy coupled with multi-way chemometric techniques for characterization and classification of Chinese lager beers.** Food Chemistry, vol. 342, 128235.

Ferreiro-González, M., Espada-Bellido, E., Guillén-Cueto, L., Palma, M., Barroso, C.G. & Barbero, G.F. (2018) **Rapid quantification of honey adulteration by visible-near infrared spectroscopy combined with chemometrics.** Talanta, vol. 188, pp. 288-292.

Gacula, M.C. (1997) **Descriptive sensory analysis in practice. Food & Nutrition Press**, Trumbull, Connecticut, USA, 698 p.

Geana, E.-. & Ciucure, C.T. (2020) **Establishing authenticity of honey via comprehensive Romanian honey analysis.** Food Chemistry, vol. 306, 125595.

Gere, A., Losó, V., Györey, A., Kovács, S., Huzsvai, L., Nábrádi, A., Kókai, Z. & Sipos, L. (2014) **Applying parallel factor analysis and Tucker-3 methods on sensory and instrumental data to establish preference maps: Case study on sweet corn varieties.** Journal of the Science of Food and Agriculture, vol. 94, pp. 3213-3225.

Gomes, H.B., Rodrigues, L.M., Massingue, A.A., Lima, I.A., Ramos, A.L.S. & Ramos, E.M. (2020) **Sensory profile and technological characterization of boneless dry-cured ham with lactulose added as a prebiotic.** Asian-Australasian Journal of Animal Sciences, vol. 33, pp. 339-348.

Guelpa, A., Marini, F., du Plessis, A., Slabbert, R. & Manley, M. (2017) **Verification of authenticity and fraud detection in South African honey using NIR spectroscopy.** Food Control, vol. 73, pp. 1388-1396.

Gustafson, S.C., Costello, C.S., Like, E.C., Pierce, S. J. & Shenoy, K.N. (2009) **Bayesian Threshold Estimation.** IEEE Transactions on Education, vol. 52, pp. 400-403.

Harshman, R. A. (1970) **Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis.** UCLA Working Papers in Phonetics, vol. 16, pp.1-84. (University Microfilms, Ann Arbor, Michigan, No. 10,085).

Helgesen, H., Solheim, R. & Næs, T. (1997) **Consumer preference mapping of dry fermented lamb sausages.** Food Quality and Preference, vol. 8, pp. 97-109.

Hermosín, I., Chicón, R.M., Cabezudo, M.D. (2003) **Free amino acid composition and botanical origin of honey.** Food Chemistry, vol. 83, pp. 263-268.

Hu, L., Yin, C., Ma, S. & Liu, Z. (2018) **Tracing the geographical origin of burdock root based on fluorescent components using multi-way chemometrics techniques.** Microchemical Journal, vol. 137, pp. 456-463.

International Coffee Organization (ICO) (2020) Available at <http://www.ico.org/prices/new-consumption-table.pdf> London, United Kingdom. Accesses in October 2020.

Iglesias, M.T., Martín-Álvarez, P.J., Polo, M.C., de Lorenzo, C., González, M. & Pueyo, E. (2006) **Changes in the free amino acid contents of honeys during storage at ambient temperature.** Journal of Agricultural and Food Chemistry, vol. 54, pp. 9099-9104.

Jibril, F.I., Hilmi, A.B. M. & Manivannan, L. (2019) **Isolation and characterization of polyphenols in natural honey for the treatment of human diseases.** Bulletin of the National Research Centre, vol. 43, 4.

Jimenez-Carvelo, A. M., Lozano, V. A., & Olivieri, A. C. (2019) **Comparative chemometric analysis of fluorescence and near infrared spectroscopies for authenticity confirmation and geographical origin of Argentinean extra virgin olive oils.** Food Control, vol. 96, pp. 22-28.

Jornal Montes Claros (2019) Mapa norte de Minas Gerais. Available in: <https://jornalmontesclaros.com.br/2019/11/12/norte-de-minas-microrregioes-do-norte-de-minas-se-reunem-em-grao-mogol-para-discutir-integracao/>. Montes Claros, MG, Brazil. Accessed in April 2020.

Jung, H., Lee, S.-., Lim, J.H., Kim, B.K. & Park, K.J. (2014) **Chemical and sensory profiles of makgeolli, Korean commercial rice wine, from descriptive,**

chemical, and volatile compound analyses. Food Chemistry, vol. 152, pp. 624-632.

Karoui, R., Dufour, E., Bosset, J. & de Baerdemaeker, J. (2007) **The use of front face fluorescence spectroscopy to classify the botanical origin of honey samples produced in Switzerland.** Food Chemistry, vol. 101, pp. 314-323.

Kaškoniene, V. & Venskutonis, P.R. (2010) **Floral Markers in Honey of Various Botanical and Geographic Origins: A Review.** Comprehensive Reviews in Food Science and Food Safety, vol. 9, pp. 620-634.

Kavanagh, S., Gunnoo, J., Marques Passos, T., Stout, J.C. & White, B. (2019) **Physicochemical properties and phenolic content of honey from different floral origins and from rural versus urban landscapes.** Food Chemistry, vol. 272, pp. 66-75.

Kelly, M.T., Blaise, A. & Larroque, M. (2010) **Rapid automated high performance liquid chromatography method for simultaneous determination of amino acids and biogenic amines in wine, fruit and honey.** Journal of Chromatography A, vol. 1217, pp. 7385-7392.

Kennard, R.W. & Stone, L.A. (1969) **Computer aided design of experiments.** Technometrics, vol. 11, pp. 137-148.

Kilcast, D. (2010) **Sensory analysis for food and beverage quality control.** Woodhead Publishing Limited, UK, 392 p.

Kowalski, B.R. (1978) **Analytical chemistry: The journal and the science, the 1970's and beyond.** Analytical Chemistry, vol. 50, pp. 1309A-1313A.

Kreuml, M.T.L., Majchrzak, D., Ploederl, B. & Koenig, J. (2013) **Changes in sensory quality characteristics of coffee during storage.** Food Science and Nutrition, vol. 1, pp. 267-272.

Kuś, P.M., Włodarczyk, M. & Tuberoso, C.I.G. (2018) **Nitrogen compounds in Phacelia tanacetifolia Benth. honey: First time report on occurrence of (−)-5-epi-lithospermoside, uridine, adenine and xanthine in honey.** Food Chemistry, vol. 255, pp. 332-339.

Laczkowski, M.S., Baqueta, M.R., de Oliveira, V.M.A.T., Gonçalves, T.R., Gomes, S.T.M., Março, P.H., Matsushita, M. & Valderrama, P. (2021) **Application of chemometric tools in the development and sensory evaluation of gluten-free cracknel biscuits with the addition of chia seeds and turmeric powder**. Journal of Food Science and Technology, 9p.

Lakowicz, J.R. (2006) **Principles of Fluorescence Spectroscopy.** Springer, New York, 960 p.

Lawless, H.T. & Heymann, H. (2010) **Sensory Evaluation of Food: Principles and Practices.** Springer, New York, 603 p.

Lenhardt, L., Zeković, I., Dramićanin, T., Milićević, B., Burojević, J. & Dramićanin, M.D. (2017) **Characterization of cereal flours by fluorescence spectroscopy coupled with PARAFAC.** Food Chemistry, vol. 229, pp.165-171.

Lenhardt, L., Bro, R., Zeković, I., Dramićanin, T. & Dramićanin, M.D. (2015) **Fluorescence spectroscopy coupled with PARAFAC and PLS DA for characterization and classification of honey.** Food Chemistry, vol. 175, pp. 284-291.

Leyva-Jimenez, F.J., Lozano-Sanchez, J., Borras-Linares, I., Cadiz-Gurrea, M.D.L.L. & Mahmoodi-Khaledi, E. (2019) **Potential antimicrobial activity of honey phenolic compounds against Gram positive and Gram negative bacteria.** LWT – Food Science and Technology, vol.101, pp.236-245.

Lim, W., Miller, R., Park, J. & Park, S. (2014) **Consumer Sensory Analysis of High Flavonoid Transgenic Tomatoes.** Journal of Food Science, vol. 79, pp. S1212-S1217.

Liu, W., Zhang, Y., Li, M., Han, D. & Liu, W. (2020) **Determination of invert syrup adulterated in acacia honey by terahertz spectroscopy with different spectral features.** Journal of the Science of Food and Agriculture, vol.100, pp. 1913-1921.

Liu, T., Ming, K., Wang, W., Qiao, N., Qiu, S., Yi, S., Huang, X. & Luo, L. (2021) **Discrimination of honey and syrup-based adulteration by mineral element chemometrics profiling.** Food Chemistry, vol. 343, 128455.

Lobo-Prieto, A., Tena, N., Aparicio-Ruiz, R., García-González, D.L. & Sikorska E. (2020) **Monitoring virgin olive oil shelf-life by fluorescence spectroscopy and sensory characteristics: A multidimensional study carried out under simulated market conditions.** Foods, vol. 9, 1846.

Louveaux, J., Maurizio, A. & Vorwohl, G., (1978) **Methods of melissopalynology.** Bee World, vol. 59, pp. 139–157.

Marini, F. (2008) **Classification Methods in Chemometrics.** Multivariate Analysis and Chemometry Applied to Environment and Cultural Heritage, pp. 1-4.

Ma, J., Bian, L., Zhao, L., Feng, X., Zhao, L., Wang, Z. & Pu, Q. (2019) **Dialysed caramel as an effective fluorophore for the simultaneous detection of three nitrophenols.** Talanta, vol. 197, pp. 159-167.

Montemurro, M., Brasca, R., Culzoni, M.J. & Goicoechea, H.C. (2019) **High-performance organized media-enhanced spectrofluorimetric**

determination of pirimiphos-methyl in maize. Food Chemistry, vol. 278, pp. 711-719.

Morais, E.C., Pinheiro, A.C.M., Nunes, C.A. & Bolini, H.M.A. (2015) **Influence of Functional and Diet/Light Claims on Chocolate Dairy Dessert Consumers' Evaluations: Bilinear and Multilinear Decomposition Methods.** Journal of Sensory Studies, vol. 30, pp. 349-359.

Muoki, P.N., De Kock, H.L. & Emmambux, M.N. (2012) **Effect of soy flour addition and heat-processing method on nutritional quality and consumer acceptability of cassava complementary porridges.** Journal of the Science of Food and Agriculture, vol. 92, pp. 1771-1779.

Næs, T., Lengard, V., Bølling Johansen, S. & Hersleth, M. (2010) **Alternative methods for combining design variables and consumer preference with information about attitudes and demographics in conjoint analysis.** Food Quality and Preference, vol. 21, pp. 368-378.

Naresh, K. (2014) **Applications of Fluorescence Spectroscopy. National Level Workshop on Spectroscopic Techniques in Structural Elucidation.** Journal of Chemical and Pharmaceutical Sciences, pp. 18-21.

Noviyanto, A. & Abdulla, W.H. (2020) **Honey botanical origin classification using hyperspectral imaging and machine learning.** Journal of Food Engineering, vol. 265, 09684.

Nunes, C.A., Pinheiro, A.C.M. & Bastos, S.C. (2011) **Evaluating Consumer Acceptance Tests By Three-Way Internal Preference Mapping Obtained By Parallel Factor Analysis (PARAFAC).** Journal of Sensory Studies, vol. 26, pp. 167-174.

Olivieri, A.C. (2008) **Analytical Advantages of Multivariate Data Processing. One, Two, Three, Infinity?** Analytical Chemistry, vol. 80, pp. 5713-5720.

Olivieri, A.C., Wu, H.-L. & Yu, R.-Q. (2009) **MVC2: A MATLAB graphical interface toolbox for second-order multivariate calibration.** Chemometrics and Intelligent Laboratory Systems, vol. 96, pp. 246-251.

Olivieri, A.C. (2012) **Recent advances in analytical calibration with multi-way data.** Analytical Methods, vol. 4, pp. 1876-1886.

Olivieri, A.C. (2014) **Analytical figures of merit: From univariate to multiway calibration.** Chemical Reviews, vol., 114, pp. 5358-5378.

Olivieri, A.C., Escandar,G., Goicoechea, H.C. & de la Peña, A.M. (2015) **Fundamentals and Analytical Applications of Multiway Calibration**. Volume 29, 1st ed., Elsevier, Amsterdam, 618 p.

Olivieri, A.C. & Escandar, G.M. (2019) **Analytical chemistry assisted by multi-way calibration: A contribution to green chemistry.** Talanta, vol. 204, pp. 700-712.

Oloibiri, V., De Coninck, S., Chys, M., Demeestere, K. & Van Hulle, S.W.H. (2017) **Characterisation of landfill leachate by EEM-PARAFAC-SOM during physical-chemical treatment by coagulation-flocculation, activated carbon adsorption and ion exchange.** Chemosphere, vol. 186, pp. 873-883.

Oroian, M., Ropciuc, S. & Paduret, S. (2018) **Honey Adulteration Detection Using Raman Spectroscopy.** Food Analytical Methods, vol. 11, pp. 959-968.

Otter, D.E. (2012) **Standardised methods for amino acid analysis of food.** British Journal of Nutrition, vol. 108, pp. S230-S237.

Palomino-Vasco, M., Acedo-Valenzuela, M.I., Rodríguez-Cáceres, M.I. & Mora-Díez, N. (2021) **Monitoring winemaking process using tyrosine influence in the excitation-emission matrices of wine.** Food Chemistry, vol. 344, 128721.

Patrignani, M., Fagúndez, G.A., Tananaki, C., Thrasyvoulou, A. & Lupano, C.E. (2018) **Volatile compounds of Argentinean honeys: Correlation with floral and geographical origin.** Food Chemistry, vol. 246, pp. 32-40.

Pereira, P.A.A.P., de Souza, V.R. & Carneiro, J. D. S. (2020) **Sensory perception of Brazilian petit suisse cheese by a consumer panel using three-way internal and external preference maps.** MOJ Food Processing & Technology, vol. 8, pp. 109-112.

Pérez, A.R., Iglesias, M.T., Pueyo, E., González, M & de Lorenzo, C. (2007) **Amino acid composition and antioxidant capacity of Spanish honeys.** Journal of Agricultural and Food Chemistry, vol. 55, pp. 360-365.

Perkowski. M.C. & Warpeha, K.M. (2019) **Phenylalanine roles in the seed-to-seedling stage: Not just an amino acid.** Plant Science, vol. 289, 110223.

Pita-Calvo, C. & Vázquez, M. (2018). **Honeydew Honeys: A Review on the Characterization and Authentication of Botanical and Geographical Origins.** Journal of Agricultural and Food Chemistry, vol. 66, pp. 2523-2537.

Privitera, M.L. & Lozano, V.A. (2017) **Development of a second-order standard addition fluorescence method for the direct determination of riboflavin in human urine samples without previous clean up and separation steps.** Microchemical Journal, vol. 133, pp. 60-66.

Rebane, R. & Herodes, K. (2008) **Evaluation of the botanical origin of Estonian uni- and polyfloral honeys by amino acid content.** Journal of Agricultural and Food Chemistry, vol. 56, pp. 10716-10720.

Rebane, R. & Herodes, K. (2010) **A sensitive method for free amino acids analysis by liquid chromatography with ultraviolet and mass spectrometric detection using precolumn derivatization with diethyl ethoxymethylenemalonate: Application to the honey analysis.** Analytica Chimica Acta, vol. 672, pp. 79-84.

Ribeiro, J.S., Ferreira, M.M.C. & Salva, T.J.G. (2011) **Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy.** Talanta, vol. 83, pp. 1352-1358.

Ríos-Reina, R., Elcoroaristizabal, S., Ocaña-González, J.A., García-González, D.L., Amigo, J.M. & Callejón, R.M. (2017) **Characterization and authentication of Spanish PDO wine vinegars using multidimensional fluorescence and chemometrics.** Food Chemistry, vol. 230, pp.108-116.

Ríos-Reina, R., Callejón, R.M., Savorani, F., Amigo, J.M. & Cocchi, M. (2019) **Data fusion approaches in spectroscopic characterization and classification of PDO wine vinegars**. Talanta, vol. 198, pp. 560-572.

Rodriguez Delgado, M. A., Malovana, S., Perez, J. P., Borges, T., & Garcia Montelongo, F. J. (2001) **Separation of phenolic compounds by high performance liquid chromatography with absorbance and fluorometric detection.** Journal of Chromatography A, vol. 912, pp. 249–257.

Roldán, A., van Muiswinkel, G.C.J., Lasanta, C., Palacios, V. & Caro, I. (2011) **Influence of pollen addition on mead elaboration: Physicochemical and sensory characteristics.** Food Chemistry, vol. 126, pp. 574-582.

Rubio, L., Sanllorente, S., Sarabia, L.A. & Ortiz, M.C. (2019) **Determination of cochineal and erythrosine in cherries in syrup in the presence of quenching effect by means of excitation-emission fluorescence data and three-way PARAFAC decomposition.** Talanta, vol. 196, pp. 153-162.

Ruoff, K., Karoui, R., Dufour, E., Luginbbühl, W., Bosset, J., Bogdanov, S. & Amadò, R. (2005) **Authentication of the botanical origin of honey by front-face fluorescence spectroscopy. A preliminary study.** Journal of Agricultural and Food Chemistry, vol. 53, pp.1343-1347.

Sádecká, J. & Tóthová, J. (2007) **Fluorescence spectroscopy and chemometrics in the food classification - A review.** Czech Journal of Food Sciences, vol. 25, pp. 159-173.

Specialty Coffee Association (SCA) (2003) Available in: <https://sca.coffee/research/protocols-best-practices>. Accessed in April 2021. Bicknacre, London, United Kingdom.

Schmidtke, L.M., Blackman, J.W., Clark, A.C. & Grant-Preece, P. (2013) **Wine metabolomics: Objective measures of sensory properties of semillon from GC-MS profiles.** Journal of Agricultural and Food Chemistry, vol. 61, pp. 11957-11967.

Schueuermann, C., Silcock, P. & Bremer, P. (2018) **Front-face fluorescence spectroscopy in combination with parallel factor analysis for profiling of clonal and vineyard site differences in commercially produced Pinot Noir grape juices and wines.** Journal of Food Composition and Analysis, vol. 66, pp. 30-38.

Sena, M.M., Trevisan, M.G. & Poppi, R.J. (2005) **PARAFAC: A chemometric tool for multi-dimensional data treatment. Applications in direct determination of drugs in human plasma by spectrofluorimetry.** Quimica Nova, vol. 28, pp. 910-920.

Sena, M.M., Trevisan, M.G. & Poppi, R.J. (2006) **Combining standard addition method and second-order advantage for direct determination of salicylate in undiluted human plasma by spectrofluorimetry.** Talanta, vol. 68, pp. 1707-1712.

Seopela, M.P., Powers, L.C., Clark, C., Heyes, A. & Gonsior, M. (2021) **Combined fluorescent measurements, parallel factor analysis and GC-mass spectrometry in evaluating the photodegradation of PAHS in freshwater systems.** Chemosphere, vol. 269, 129386.

Sergiel, I., Pohl, P., Biesaga, M. & Mironczyk, A. (2014) **Suitability of three-dimensional synchronous fluorescence spectroscopy for fingerprint analysis of honey samples with reference to their phenolic profiles.** Food Chemistry, vol. 145, pp.319-326.

Shang, F., Wang, Y., Wang, J., Zhang, L., Cheng, P. & Wang, S. (2019) **Determination of three polycyclic aromatic hydrocarbons in tea using four-way fluorescence data coupled with third-order calibration method.** Microchemical Journal, vol. 146, pp. 957-964.

Shen, S., Wang, J., Chen, X., Liu, T., Zhuo, Q. & Zhang, S.Q. (2019) **Evaluation of cellular antioxidant components of honeys using UPLC-MS/MS and HPLC-FLD based on the quantitative composition-activity relationship.** Food Chemistry, vol. 293, pp. 169-177.

Shiddiq, M., Zulkarnain, Z., Asyana, V. & Aliyah, H. (2019) **Identification of Pure and Adulterated Honey Using Two Spectroscopic Methods.** Journal of Physics: Conference Series, vol. 1351.

Siddiqui A.J., Musharraf S.G., Choudhary M.I. & Rahman A. (2017) **Application of analytical methods in authentication and adulteration of honey.** Food Chemistry, vol. 217, pp. 687-698.

Silva, A.C.D., Soares, S.F.C., Insausti, M., Galvão, R.K.H., Band, B.S.F. & Araújo, M.C.U.D. (2016) **Two-dimensional linear discriminant analysis for classification of three-way chemical data.** Analytica Chimica Acta, vol. 938, pp. 53-62.

Skoog, D.A., Holler., F.J. & Crouch, S.R. (2009) **Princípios de Análise Instrumental.** 6ª ed., Bookman, Porto Alegre, Brasil, 1055 p.

Smilde, A. K., Tauler, R. Saurina, J. & Bro, R. (1999) **Calibration methods for complex second-order data.** Analytica Chimica Acta, vol., 398, pp. 237-251.

Smilde, A., Bro, R. & Geladi, P. (2004) **Multi-way Analysis: Applications in the Chemical Sciences.** John Wiley & Sons, Ltd, Chichester, England, 396p.

Stanković, M., Bartolić, D., Šikoparij, B., Spasojević, D., Mutavdžić, D., Natić, M. & Radotić, K. (2019) **Variability Estimation of the Protein and Phenol Total Content in Honey Using Front Face Fluorescence Spectroscopy Coupled with MCR–ALS Analysis.** Journal of Applied Spectroscopy, vol. 86, pp. 256-263.

Strelec, I., Brodar, L., Flanjak, I., Kenjerić, F.Č., Kovač, T., Kenjerić, D.Č. & Primorac, L., (2018) **Characterization of Croatian honeys by right-angle fluorescence spectroscopy and chemometrics.** Food Analytical Methods, vol. 11, pp. 824-838.

Valli, E., Bendini, A., Popp, M. & Bongartz, A. (2014) **Sensory analysis and consumer acceptance of 140 high-quality extra virgin olive oils.** Journal of the Science of Food and Agriculture, vol. 94, pp. 2124-2132.

Vasić, V., Đurđić, S., Tosti, T., Radoičić, A., Lušić, D., Milojković-Opsenica, D., Tešić, Ž. & Trifković, J. (2020) **Two aspects of honeydew honey authenticity: Application of advance analytical methods and chemometrics.** Food Chemistry, vol. 305, 125457.

Vasilev, M., Taneva, I., Velikova, M. & Mihova, R. (2016) **Interpreting sensory data of cheese "Krema" by Principal Component Analysis.** Journal of Technics and Technologies, vol. 4, pp. 139-144.

Vera, M., Martín-Alonso, J., Mesa, J., Granados, M., Beltran, J.L., Casas, S., Gibert, O. & Cortina, J.L. (2017) **Monitoring UF membrane performance treating surface-groundwater blends: Limitations of FEEM-PARAFAC on the assessment of the organic matter role**. Chemical Engineering Journal, vol. 317, pp. 961-971.

Viana, F.R., do Carmo, L.S. & Bastos, E.M.A.F. (2018) **Antibacterial activity of Aroeira honeys produced in Minas-Gerais against bacteria of clinical importance.** Acta Scientiarum - Biological Sciences, vol. 40, 4p.

Wang, T., Wu, H., Long, W., Hu, Y., Cheng, L., Chen, A. & Yu, R. (2019) **Rapid identification and quantification of cheaper vegetable oil adulteration in camellia oil by using excitation-emission matrix fluorescence spectroscopy combined with chemometrics.** Food Chemistry, vol. 293, pp. 348-357.

Wei, J., Tu, C., Yuan, G., Zhou, Y., Wang, H. & Lu, J. (2020) **Limited Cu(II) binding to biochar DOM: Evidence from C K-edge NEXAFS and EEM-PARAFAC combined with two-dimensional correlation analysis.** Science of the Total Environment, vol. 701, 134919.

Wold, S, Geladi, P., Esbensen, K. & Öhman, J. (1987) **Multi-way Principal Components and PLS-Analysis.** Journal of Chemometrics, vol. 1, pp. 41-56.

Wold, S. & Sjöström, M. (1998) **Chemometrics, present and future success,** Chemometrics and Intelligent Laboratory Systems, vol. 44, pp. 3–14.

Zhi, R., Zhao, L. & Shi, J. (2016) **Improving the sensory quality of flavored liquid milk by engaging sensory analysis and consumer preference.** Journal of Dairy Science, vol. 99, pp. 5305-5317.

Zeng, J., Guo, Y., Han, Y., Li, Z., Yang, Z., Chai, Q., Wang, W., Zhang, Y. & Fu, C. (2021) **A Review of the discriminant analysis methods for food quality based on near-infrared spectroscopy and pattern recognition.** Molecules, vol. 26, 749.